

**Convergent transcription and nested gene models studied by
AFM**

Daniel Jeffrey Billingsley

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

The University of Leeds
School of Physics and Astronomy
Department of Oral Biology

April, 2012

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.

Chapters 1 and 2 are based on work published within;

Billingsley, D.J., Bonass, W.A., Crampton, N., Kirkham, J., and Thomson N.H. Single molecule studies of DNA transcription using atomic force microscopy in *Physical Biology*. 2012; 9(2):021001.

Chapter 4 is based on work published within;

Billingsley, D.J., Kirkham, J., Bonass, W.A., and Thomson N.H. Atomic force microscopy of DNA at high humidity: irreversible conformational switching of supercoiled molecules in *Physical Chemistry Chemical Physics*. 2010; 12(44):14727-34.

Chapter 6 is based on work published within;

Billingsley, D.J., Crampton, N., Kirkham, J., Thomson, N.H. and Bonass, W.A. Single-stranded loops as end-label polarity markers for double-stranded linear DNA templates in atomic force microscopy in *Nucleic Acids Research* 2012

In these publications, D.J. Billingsley was responsible for acquisition and analysis of all data and wrote the first draft of the manuscripts. N. Crampton is an independent research fellow who assisted the candidate in interpretation and analysis of data. The other three authors are the academic supervisors of D.J. Billingsley.

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.

© 2012 The University of Leeds and Daniel Jeffrey Billingsley

Abstract

Genomic DNA is organised in complex spatial arrangements, and a given stretch of DNA may encode more than one gene. In some cases one gene may be entirely contained within a region of the DNA already occupied by another larger gene. The presence of these nested genes, often situated in introns and in the opposite orientation, poses important implications with regards to gene expression, function and regulation.

A consequence of the nested gene arrangement is convergent transcription, occurring when two promoters on opposite DNA strands are active. Elucidating the mechanics of multiple interacting proteins on single DNA templates requires single molecule methods such as atomic force microscopy (AFM). AFM can accurately determine the relative positions of enzymes, such as RNA polymerase (RNAP), on individual DNA templates. The central aim of this thesis is to use AFM to study the outcomes of convergent transcription, using linear DNA templates that function as models for nested genes.

Fundamental aspects of imaging DNA on mica with AFM were investigated, with a view to optimising sample preparation. The main processes involved with preparing DNA samples, ready for scanning, were examined in turn. Effective binding was achieved by introducing divalent cations into a deposition buffer. Mica ion exchanged with Ni(II) usually gave rise to kinetically-trapped DNA molecules, however short linear fragments (< 800bp) were seen to deviate from the expected behaviour, indicating that ion-exchanged mica is heterogeneous, and contains patches or domains. These findings can be used to more readily control binding of DNA to substrates.

The outcomes of varying the relative humidity while imaging biomolecular systems are largely unexplored to date. Various DNA samples were imaged in conditions of varying humidity. In particular when supercoiled plasmids are scanned at very high relative humidity (> 90% RH), localised DNA backbone motions or conformational changes were observed. Humidity controlled AFM

will be a useful technique for probing DNA topology without some of the drawbacks of imaging under bulk solution.

Initial studies of transcription utilised templates containing two promoter sites and *E. Coli* RNAP. Two promoter arrangements were studied: a convergent template containing the promoter sites on opposite strands directed towards each other, and a tandem template containing the promoters in the same direction, on the same strand. It was shown that collisions between RNAPs led to similar outcomes in both cases: RNAPs are unable to pass each other and remain stalled against each other. In the convergent case, it was observed that after collision one RNAP could cause another to backtrack along the template.

By end-labelling double-stranded (ds) DNA templates with a single-stranded DNA loop, the polarity of the molecules can be established in the AFM. It allowed better discrimination between outcomes of collision events on single DNA molecules and importantly, it enabled a quantitative comparison of the relative frequencies of the outcomes. The most common outcome is a collision between an actively transcribing elongation complex (EC) and a “sitting duck” (SD), which is a stalled RNAP or open promoter complex (OPC). In collisions initiated from OPCs, the most likely outcome, a collision between an EC and an SD occurs ~74% of the time. This causes sizeable back-tracking of the inactive RNAP, on average 59 nm upstream of the promoter. A significant fraction of the collisions (~15%) are between actively transcribing RNAP while the remainder (~11%) are undetermined. End-labelling of dsDNA using nucleic acid structures did not interfere with AFM sample preparation and can be used as a generic approach to studying interactions of multiple proteins on DNA templates at the single molecule level.

Acknowledgments

My greatest thanks go to my supervisors Dr. Neil Thomson, Dr. Bill Bonass, and Professor Jennifer Kirkham whose expertise, knowledge, and guidance has been invaluable.

I am grateful to everyone in the Molecular and Nanoscale physics group and the Department of Oral Biology at the University of Leeds for their assistance over the past few years and for making the work place an enjoyable place to be. Similarly, I have enjoyed working with the White Rose DTC and value the training they have given me.

Thanks must also go to my friends for all their support, and for providing distractions from work when necessary. I would like to give special thanks to my girlfriend, Natalie for all her help, patience and words of encouragement. Thank you in particular to my family for their encouragement and belief in me throughout all of my studies.

Contents

1	Introduction to DNA and convergent transcription	1
1.1	Nucleic acid structure	1
1.1.1	The discovery of DNA as the genetic material	1
1.1.2	The constituents of nucleic acids.....	3
1.1.3	The DNA double helix	6
1.1.4	RNA structure	10
1.2	DNA Transcription.....	12
1.2.1	The central dogma of molecular biology	12
1.2.2	Introduction to DNA Transcription and RNA polymerase	13
1.2.3	Transcription Initiation	14
1.2.4	Elongation.....	18
1.2.5	Termination.....	21
1.3	Nested genes and convergent transcription	22
2	Introduction to AFM of single molecule DNA systems	29
2.1	Overview of AFM	29
2.1.1	Introduction to scanning probe microscopy	30
2.1.2	AFM Instrumentation.....	31
2.1.3	AFM Forces.....	33
2.1.4	Imaging Modes	36
2.2	Imaging DNA with AFM.....	40
2.2.1	Imaging substrates and practicalities	40
2.2.2	Preparation of DNA samples for AFM	43
2.2.3	Practicalities of studying transcription with AFM	47
2.3	Previous AFM studies of transcription	50
2.3.1	Initiation studies	50
2.3.2	Elongation studies	52
2.3.3	Termination.....	54
2.4	Outlook and developments.....	55
2.4.1	Fast scan AFM	55
2.4.2	Moving to more complex systems.....	57

3	AFM imaging of DNA on mica	60
3.1	Introduction.....	60
3.2	Preparation of DNA samples	61
3.2.1	The Polymerase Chain Reaction (PCR)	61
3.2.2	Gel Electrophoresis.....	64
3.2.3	Spectrophotometry.....	67
3.2.4	Restriction Enzymes.....	69
3.2.5	Column Purification.....	70
3.2.6	Circular Samples.....	72
3.2.7	AFM imaging of DNA on mica	75
3.3	Sample Preparation Investigations	76
3.3.1	Deposition kinetics.....	76
3.4	Comparison of different sample rinsing regimes	80
3.4.1	Effect of varying rinsing volume.....	80
3.4.2	Washing with alcohol.....	82
3.4.3	Drying Investigation	88
3.5	Effect of surface treatment	93
3.5.1	Dependence of equilibration on deposition time.....	94
3.5.2	Effect of fragment length on deposition	98
3.5.3	Discussion.....	106
3.6	Conclusions.....	111
4	Humidity controlled AFM of DNA	113
4.1	Introduction.....	113
4.2	Samples and preparation	114
4.3	Results	116
4.3.1	Linear fragments.....	116
4.3.2	Circular Molecules.....	124
4.3.3	Summary of results.....	127
4.4	Discussion	130
4.4.1	Mica water layer	130
4.4.2	Linear Fragments.....	133
4.4.3	Circular molecules.....	136
4.4.4	Differences between linear and circular fragments.....	139
4.4.5	Possible use of humidity-controlled AFM in transcription studies	139

4.5	Conclusions	142
5	Study of transcription interference in DNA: E. Coli RNAP complexes.....	144
5.1	Introduction.....	144
5.2	Sample preparation.....	145
5.3	Convergent pDSU template	149
5.3.1	Open Promoter complexes (OPCs)	149
5.3.2	RNAP size.....	154
5.3.3	Stalled Elongation Complexes (SEC).....	156
5.3.4	Collided complexes (CCs).....	159
5.3.5	Visualisation of RNA	162
5.3.6	Comparison of complexes.....	164
5.3.7	Inter-RNAP separation	166
5.3.8	Arm-lengths.....	167
5.4	Tandem pDSP template.....	168
5.4.1	Initial Imaging of pDSP complexes	169
5.4.2	pDSP results.....	169
5.4.3	pDSP arm-length measurements.....	172
5.4.4	pDSP Inter-RNAP separation.....	174
5.5	Discussion	176
5.5.1	Wrapping	176
5.5.2	Inter-RNAP separation	182
5.5.3	Arm-Lengths.....	185
5.5.4	Numerical Analysis.....	188
5.6	Conclusions	192
6	Discrimination of collision outcomes in convergent transcription using single-stranded DNA loop end-labelled DNA templates.....	194
6.1	Introduction to DNA labelling.....	194
6.2	Ligation-based labelling.....	196
6.2.1	Introduction to approach and structures used.....	196
6.2.2	AFM of small oligonucleotides	198
6.2.3	Strategy for ligation-based labelling.....	200
6.2.4	AFM of ligated fragments	204
6.3	Attaching stem-loop via a single-cycle PCR approach.....	206

6.3.1 Protocol for labelling.....	206
6.3.2 Initial imaging of complexes.....	209
6.3.3 Convergent Transcription	215
6.3.4 Analysis of contour lengths.....	218
6.4 Discussion.....	225
6.4.1 Labelling.....	225
6.4.2 RNAP separation on labelled template.....	226
6.4.3 Back-tracking.....	227
6.5 Conclusions.....	231
7 Future work.....	233
8 Conclusions.....	238

Figures list

Figure 1-1 Nucleic acid bases.....	4
Figure 1-2. The sugar units present in nucleic acids.	5
Figure 1-3. DNA nucleotides.....	6
Figure 1-4. DNA double helical structure consisting of two complementary DNA strands winding around each other.....	8
Figure 1-5. Comparison of the A, B and Z forms of DNA.....	9
Figure 1-6. Example of an RNA stem loop.	11
Figure 1-7. The central dogma describes the residue by residue transfer of sequence information between the information-carrying biopolymers.	12
Figure 1-8. (a) X-ray crystal structure of <i>Thermus aquaticus</i> core RNA polymerase. Adapted from Zhang <i>et al.</i> [219] (b) Schematic diagram of RNAP-nucleic acid complex.	13
Figure 1-9. Schematic diagram of the different methods of promoter location..	15
Figure 1-10. Scrunching model of abortive initiation.....	17
Figure 1-11. Method of rho-independent termination.....	21
Figure 1-12. A nested gene is entirely located within the boundaries of another host gene.....	23
Figure 1-13. Methods of transcriptional interference (TI).....	25
Figure 1-14. Different stages involved with convergent transcription.	28
Figure 2-1 Typical set-up of an AFM.....	32
Figure 2-2. A graph of potential ($U(r)$) versus distance for the 6-12 Lennard-Jones potential.....	34

Figure 2-3. Capillary neck formation between a sphere (tip) and surface in the presence of a bulk water layer.	35
Figure 2-4 . Main modes of AFM operation.....	37
Figure 2-5. Demonstration of AFM phase imaging.....	39
Figure 2-6. Schematic showing model for counter-ion correlations between DNA and mica.....	42
Figure 2-7 DNA sample preparation.....	44
Figure 2-8. Schematic diagram of the adsorption mechanisms of DNA onto mica substrate from solution.....	46
Figure 2-9. Method for studying transcription dynamically in liquid with AFM.	49
Figure 2-10. Model demonstrating how DNA can wrap around an RNAP core, and how the characteristic bend angle is defined.	51
Figure 2-11 Schematic of how dynamic studies into transcription elongation appeared under AFM.....	53
Figure 3-1. Schematic diagram of the PCR process.....	63
Figure 3-2. Diagram showing origins of PCR generated linear fragments.	64
Figure 3-3. Example of a gel electrophoresis experiment, with the main features indicated.	65
Figure 3-4. Schematic diagram of a spectrophotometer.	68
Figure 3-5. Diagram showing the action of two restriction enzymes.	70
Figure 3-6. The spin purification procedure.....	71
Figure 3-7. Comparison of different plasmid fragments of plasmid pBR322 imaged together by AFM.....	73

Figure 3-8. Gel electrophoresis of a plasmid preparation and its linear digestions.....	74
Figure 3-9. AFM images of linearised plasmid pBR322.....	75
Figure 3-10. Comparative AFM images of (a) 800 bp and (b) 4.2 kbp fragments incubated for increasing time.....	77
Figure 3-11. The deposition kinetics of (a) 800 bp and (b) 4.2 kbp DNA fragments on mica.....	79
Figure 3-12. Comparison of different rinsing volumes for 800 bp fragments....	81
Figure 3-13. Comparison of different rinsing volumes for the linear plasmid fragments.....	81
Figure 3-14. Comparison between a conventional rinse with milliQ water, and with isopropanol.	83
Figure 3-15. Effect of alcohol on a conventionally made sample.	85
Figure 3-16. Series of images of samples prepared by washing the mica surface with adsorbed DNA with different strengths of ethanol.	86
Figure 3-17. Occasionally, after washing the sample with highly concentrated ethanol solutions, complex structures were observed.....	87
Figure 3-18. Diagram showing each mode of drying.....	89
Figure 3-19. Software zooms of individual 800 bp molecules used to illustrate classification scheme used..	90
Figure 3-20. Graph showing the fraction of molecules adopting a particular orientation for each drying method.....	91
Figure 3-21. Graph showing number density of DNA molecules on the mica surface for the different drying directions.	92
Figure 3-22. The two different sample preparations. (a) MgCl ₂ prep. (b) Mg(II) with extra NiCl ₂ prep.....	94

Figure 3-23. Plot showing how the experimentally measured average square end-to-end distance varies with deposition time for two differently sized fragments: 800 bp (red points) and 4.2 kbp (black points).....	96
Figure 3-24. Typical AFM images of each fragment prepared with (a) Mg(II) only on the left-hand side, and (b) additional Ni(II) pre-treatment of mica on the right-hand side.....	101
Figure 3-25. Experimental values of square end-to-end distances (points) of samples prepared with Mg(II) alone, plotted with a curve of the theoretical values for surface equilibration.....	103
Figure 3-26. Experimental values of square end-to-end distances (points), of samples prepared with Ni(II) pre-treatment of mica, plotted with a curve of the theoretical values for kinetic trapping.....	105
Figure 3-27. Software zooms of two linear 4.2 kbp fragments prepared with (a) Mg(II) only and (b) additional pre-treatment of mica with Ni(II).....	107
Figure 3-28. (a) Schematic diagram of an idealised possible structure of the ion-exchanged mica surface, containing regions of K(I) and patches of Ni(II). (b) Representations of how different DNA fragments may appear.....	110
Figure 4-1. Experimental set-up used for controlled-humidity imaging. The AFM was placed inside a hood to isolate it from the laboratory environment.	115
Figure 4-2. Average DNA height versus relative humidity (RH) for individual double-stranded long linear (LL) DNA helices lying against the mica surface (n=50 in each point).....	117
Figure 4-3. Average DNA width versus relative humidity (RH) for individual double-stranded long linear (LL) DNA helices lying against the mica surface (n=50 for each point).....	118
Figure 4-4. Images of linearised plasmid (LL) at both low and high humidity of the same region prepared with (a) Mg(II) alone and (b) Ni(II).	120

Figure 4-5. Software zooms of individual LL molecules, depicting molecules imaged from low to high humidity on (a) Mg(II) only prepared samples, and (b) with additional Ni(II) pre-treatment..... 121

Figure 4-6. Images of 800 bp linear fragment (SL) at both low and high humidity prepared with (a) Mg(II) and (b) Ni(II) pre-treatment..... 122

Figure 4-7. Software zoom of an area depicted in Figure 4-6 taken at (a) low humidity , and one at (b) high humidity..... 123

Figure 4-8. Comparison of DNA imaging at low humidity (top) and high humidity (bottom) for closed circular (SC) plasmid prepared with (a) Mg(II) only and (b) with additional Ni(II) pretreatment. 124

Figure 4-9. Software zooms of a number of SC DNA molecules depicted in Figure 4-8 showing the molecules imaged at low humidity (left) and high humidity (right) and illustrating the molecular perturbations.. 125

Figure 4-10. Images of relaxed (OC) plasmids at both low and high humidity prepared with Mg(II) (a) on mica and (b) Ni-mica. 126

Figure 4-11. Software zooms of relaxed (OC) plasmids at both low and high humidity prepared with (a) Mg(II) on mica and (b) Ni(II)-mica..... 127

Figure 4-12. Series of AM AFM images of supercoiled (SC) plasmid prepared with Mg(II) on mica. 128

Figure 4-13. Series of AFM images showing the same region of a sample of closed circular (SC) plasmid pBR322 prepared with Mg(II) on mica.. 130

Figure 4-14. On some images a step in the background was observed (black arrows).. 132

Figure 4-15. Schematic diagram of the lattice model of mica proposed by Pastre *et al.* [153]..... 133

Figure 4-16. Schematic representation of the behavior observed when surface equilibrated linear fragments are imaged from low to high humidity. 135

Figure 4-17. Schematic representation of how SC molecules behave when imaged at high RH.	137
Figure 4-18. Models of supercoiling produced by elongating RNAP originally proposed by Liu and Wang [121].	141
Figure 4-19. Summary showing typical conformational changes in SC supercoiled DNA as the humidity is increased, with a corresponding increase in the thickness of the surface water layer.	142
Figure 5-1. pDSU and pDSP plasmid constructs. Both plasmids can be digested with Hind III to produce linear templates (1149 bp) containing two distinct promoters.	145
Figure 5-2. (a) Schematic representation of the 1149 bp fragment used in the convergent transcription experiments	147
Figure 5-3. Schematic diagram showing a summary of the different classes of complex formed for investigations into convergent transcription.	148
Figure 5-4. Montage of individual software zooms depicting typical open promoter complexes.	150
Figure 5-5. Compaction of DNA by RNAP. Contour length measurements for DNA with 0 (red), 1 (green), and 2 (blue) RNAP bound.	151
Figure 5-6. (a) Schematic diagram demonstrating how RNAP induced DNA bend angles were measured. (b) OPC bend angle distribution (n = 92)..	152
Figure 5-7. Protein-induced bend angles measured for (a) specific promoter-bound (n = 55) and (b) unspecific binding events (n = 30).	154
Figure 5-8. Equivalent disc radius for RNAPs that are (a) Bound to the mica substrate only, and (b) associated with DNA as an OPC.	155
Figure 5-9. Montage of individual software zooms depicting typical stalled elongation complexes (SECs).	156
Figure 5-10. SEC bend angle distribution.	157

Figure 5-11. Montage of individual software zooms depicting collided complexes prepared via (a) a 2-step process with a stalled intermediate (b) a 1-step process.....160

Figure 5-12. Inter-RNAP measurements from CCs prepared via (a) a 2-step process (n =67) and (b) a 1-step process (n =62).161

Figure 5-13. (a) Software zooms of DNA-RNAP complexes with putative RNA transcripts exiting the protein (indicated by arrows). (b) Model of the elongation complex.....163

Figure 5-14. Examples of complexes that have an additional RNAP associated with a template-bound one.....164

Figure 5-15. Evolution of RNAP position and separation in convergent transcription. (a) pDSU arm length contour distances for OPC (n = 54), SEC (n = 88) and CC (1-step reaction) (n = 62). (b) pDSU inter-RNAP distance contour measurements for the same complexes.....166

Figure 5-16. pDSP template.....168

Figure 5-17. Montage of single-molecule software zooms of RNAPs bound to pDSP template representing (a) Promoter bound RNAPs (OPCs), (b) Stalled RNAPs (SECs), and (c) Collided RNAPs (CCs).169

Figure 5-18. Evolution of RNAP position and separation in transcription originating from tandem promoters (a) pDSP arm length contour distances for OPC (n =48), SEC (n = 54) and CC (n =65) (1-step case). (b) pDSU inter-RNAP distance contour measurements for the same complexes.....173

Figure 5-19. Montage of OPCs.....177

Figure 5-20. Model of the E.coli RNAP OPC proposed by Rivetti *et al.* [168].....178

Figure 5-21.Examples of main conformations of OPC seen: S- and U-shaped.. 180

Figure 5-22. Summary of collisions between RNAPs firing from two convergent promoters.....185

Figure 5-23. Schematic diagram of how backtracked CCs appeared.....	186
Figure 5-24. After collision, three main classes of CCs were observed for (a) pDSU and (b) pDSP.	190
Figure 6-1. Example of how DNA stem-loops can form.	196
Figure 6-2. The three structures designed to end-label the template, as predicted by DNA secondary structure prediction tools.	198
Figure 6-3. AFM images of the three different structures proposed for use in end-labelling.	200
Figure 6-4. Schematic diagram showing how labelling was performed using an oligonucleotide and a PCR-generated fragment as starting material.	201
Figure 6-5. Gel showing aftermath of ligation reactions.	203
Figure 6-6. Montage showing individual software zooms of different ligation reactions.	205
Figure 6-7. Structure used for PCR-based end-labelling.	207
Figure 6-8. Summary of method used to end-label DNA, utilising a PCR based approach.	208
Figure 6-9. Schematic representation of the labelled template used in this study.	209
Figure 6-10. Montage showing software zooms of individual molecules with a globular feature at one end: the stem-loop attached through the PCR reaction.	210
Figure 6-11. Examples of the different morphologies of end-loop structures, together with the relative occurrence of each class.	210
Figure 6-12. Example of a typical scan of the end-labelled DNA.	212

Figure 6-13. Contour length distributions for (a) bare DNA (Average = 386 ± 3 nm, n = 70), and (b) DNA with an end feature (Average = 407 ± 3 nm, n = 74).213

Figure 6-14. Distribution of diameter of the end-feature.214

Figure 6-15. Montage of single molecule software zooms depicting OPCs..215

Figure 6-16. Montage of single molecule software zooms depicting CCs.....217

Figure 6-17. Inter-RNAP measurements from (a) OPCs (n = 58) and (b) CCs (n =91)..218

Figure 6-18. Distribution of arms lengths for (a) OPCs (n = 58) and (b) CCs (n =91).219

Figure 6-19. Numerical analysis of the positions of one RNAP relative to the other revealed three main classes of complex.....222

Figure 6-20. Plots depicting the data for each pair of arm-lengths sorted in numerical order of (a) the short-arm and (b) the loop arm.224

Figure 6-21. Collision outcomes observed in labelled complexes.229

Table list

Table 1-1 Structural features of the main forms of DNA.....	9
Table 3-1 Experimental values of the mean square end-to-end distance for samples prepared using Mg(II) only.	103
Table 3-2. Experimental values of the mean square end-to-end distance for samples prepared using an additional Ni(II) pre-treatment step.....	105
Table 4-1 Comparison of theoretical values of average square end-to-end distances with for each binding case, with experimental values for the two linear fragments.....	119
Table 4-2. Summary of the occurrence of local conformational changes or local motion of dsDNA molecules on mica at high RH (> 90%) from two sample preparation conditions..	129
Table 5-1 Summary of measurements taken for pDSU template.....	165
Table 5-2 Summary of measurements for pDSP template.	170
Table 6-1. Comparison of different classes of DNA molecules observed for each preparation mix and a control that had not undergone an end-labelling reaction.	211
Table 6-2. Comparison of length measurements for labelled and un-labelled OPCs.....	217

Abbreviations

A: adenine

AFM: atomic force microscopy

AM: amplitude modulation

ATP: adenosine triphosphate

bp: base pair

C: cytosine

CC: collided complex

CPC: closed promoter complex

CTP: cytidine triphosphate

DNA: deoxyribonucleic acid

dNTP: deoxynucleoside triphosphate

dsDNA: double stranded DNA

DTT: dithiothreitol

EC: elongation complex

FM: frequency modulation

G: guanine

GTP: guanosine triphosphate

HOPG: highly oriented pyrolytic graphite

LL: long linear

nt: nucleotide

NTP: nucleoside triphosphate

OC: open circular

OPC: open promoter complex

PCR: polymerase chain reaction

RH: relative humidity

RNA: ribonucleic acid

RNAP: RNA polymerase

RT PCR: reverse-transcription PCR

SC: supercoiled

SD: sitting duck

SEC: stalled elongation complex

SL: short linear

SPM: scanning probe microscopy

STM: scanning tunnelling microscopy

TAE: Tris-Acetate EDTA

TI: transcription interference

U: urasil

UHV: ultra-high vacuum

UTP: uridine triphosphate

Chapter 1

1 Introduction to DNA and convergent transcription

The molecule deoxyribonucleic acid (DNA) carries a crucial importance for life. Within its structure DNA contains the genetic information to construct proteins, molecules which participate in nearly every biological process in cells, and hence has a role in the development and effective functioning of every living organism. This chapter begins with a description of the components of nucleic acids, and how the double-helical structure of DNA lies at the heart of its ability to convey the genetic information. Transcription is the process whereby RNA polymerase catalyses the synthesis of an RNA chain using a complementary DNA sequence as a template. The main stages and processes in this process are described. This thesis focuses on the investigation of an interesting spatial-arrangement of genes. A nested gene is a functional gene that is completely embedded in a larger gene. In particular we look at the implications of simultaneous expression of both nested and host genes. Previous studies into this phenomena and situations where multiple RNAPs interact on the same sequence are also reviewed.

1.1 Nucleic acid structure

1.1.1 The discovery of DNA as the genetic material

The Austrian monk Gregor Mendel can possibly be thought of as the originator of the science of genetics. He chose the pea plant as his model system, and observed that various inheritable characteristics occurring in alternate identifiable forms, such as flower colour, persisted over many generations. More importantly he went on to study crosses between plants with differing traits, and concluded that these qualities were controlled by two factors (later known as genes), one from each parent. Around the time of Mendel's work it was discovered that the nucleus was responsible for carrying the heritable

characteristics of the cell, and that heredity in humans was transferred through the egg and sperm.

The development of special dyes allowed the interior of the nucleus to be probed, and led to the discovery of chromosomes using optical microscopy techniques. These were present in a constant number for a particular species, and observed to exactly double prior to cell division, with each daughter cell subsequently receiving one copy of each chromosome present in the parental cell. It was also found that sex cells had half the number of chromosomes expected, but that fertilization between the sperm and egg acted to restore the chromosomal number, with one set coming from the male parent and one from the female. It was subsequently proposed that factors such as those observed by Mendel were physical objects, or genes, located on the chromosomes. The first physical characteristic discovered to be governed by chromosomes was sex itself. Studies on the red-eyed fruit fly *Drosophila* found that heritable traits were dependent on the fly's sex [138]. One genetic cross was observed to have white eyes. The mutated gene responsible for eye colour was subsequently located on the X sex chromosome. Later it was discovered that x-rays could induce changes in the structure of chromosomes, and that these mutations would alter the heritable characteristics governed by the chromosomes [16, 191].

The realisation that genes were located on chromosomes was an important development, however as long as their molecular structure remained unknown it would be difficult to further understand genetics. With the development of the electron microscope in the 1940s, it was hoped that with the improved imaging resolution it would be possible to resolve individual genes situated on chromosomes. However, the early images were highly disordered and suggested an irregular gene arrangement. Later, it did become possible to purify chromosomes away from the other cellular components, providing new insights into their constituents. The two main components were found to be DNA and small positively charged proteins known as histones. DNA was discovered as early ago as 1869 by the Swiss scientist Frederick Miescher, when he discovered a curious phosphorus compound in white blood cells. He originally

named this compound *nuclein*, but it later became known as nucleic acid. DNA certainly had the location expected for the genetic material, but it was still believed that another small, poorly understood protein would be the true genetic carrier.

This view began to change after studies into the pathogenicity of bacteria. It was found that if an amount of pathogenic cells previously inactivated by heat were mixed with living non-pathogenic cells, then a small percentage of the living cells became pathogenic [11]. The transforming agent was later discovered to be DNA. The role of DNA in heredity was finally confirmed after studies into T2 bacteriophages [89]. The phages were radiolabelled and it was discovered that when the virus infected a bacterium only DNA entered and that their protein coats remained outside. Infection led to the appearance of mutated phages, and thus it was concluded that DNA had to be the true genetic material.

1.1.2 The constituents of nucleic acids

DNA (deoxyribonucleic acid) is a long chain-like molecule constructed from small nucleotide monomers. These nucleotides comprise of three key components: a sugar, a phosphate group, and one of four bases. There are two groups of bases; both classes are flat, ring-shaped molecules containing carbon and nitrogen, but differ in that the pyrimidines (cytosine and thymine) contain one ring, whilst the purines (adenine and guanine) have two (see Figure 1-1). The sugars in DNA are always 2' deoxyribose, which contain five carbon atoms, labeled 1' to 5', and adopts a ring structure via an esterification reaction between the carbonyl group on carbon 1' and the hydroxyl group on carbon 4'. (Figure 1-2). The phosphates in DNA are always monophosphates, whilst free nucleotides can contain di- and triphosphates. The acidic natures of nucleic acids arise from the presence of phosphate esters, which release hydrogen ions in solution.

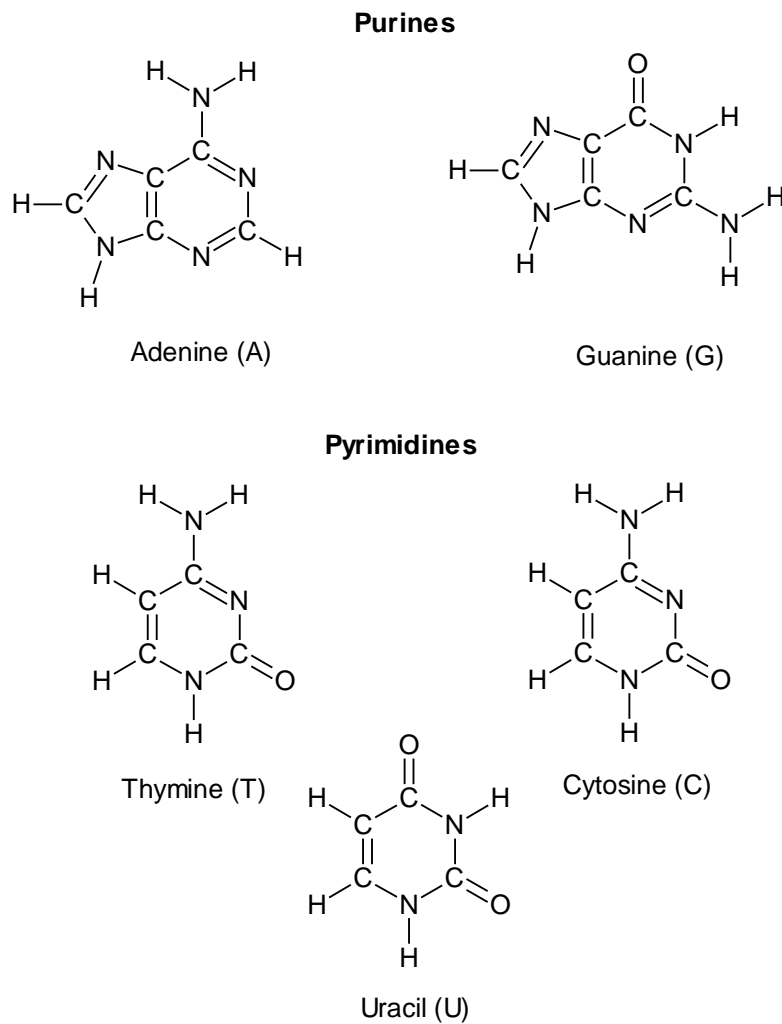


Figure 1-1 Nucleic acid bases

RNA (ribonucleic acid) shares some of the features of DNA; its structure consists of chains of nucleotides, each containing a phosphate group, sugar and either a purine or a pyrimidine base. However, the sugar group in RNA is different to that of DNA, containing an extra oxygen atom on carbon atom 2' to become ribose (see Figure 1-2). In addition, the base thymine is not present in RNA, but is replaced by the structurally similar pyrimidine uracil.

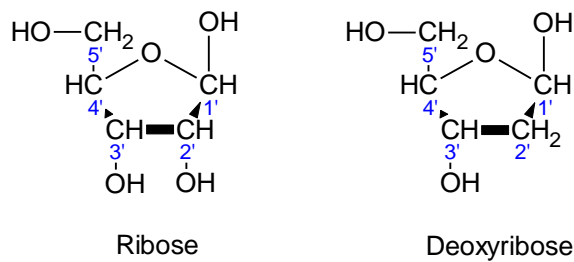


Figure 1-2. The sugar units present in nucleic acids. In the case of RNA this is a ribose, whereas DNA contains the sugar deoxyribose. Deoxyribose lacks an oxygen atom that is present in ribose.

Nucleotides are formed in the same way: the 1'-carbon of the pentose sugar is attached to nitrogen-9 of a purine base (Figure 1-3a) or nitrogen-1 of a pyrimidine base (Figure 1-3b), while a phosphate attaches to the 5'-carbon of the sugar. Nucleotide monomers are joined together to form a nucleic acid, and are linked together by chemical bonds between the phosphate group of one nucleotide, and the sugar (ribose or deoxyribose) unit on the next. These linkages take the form of phosphodiester bonds, with the phosphate group connecting the 5'-carbon on one deoxyribose residue to the 3'-carbon in the next sugar residue (Figure 1-3c). As a result, a nucleic acid chain will always have a terminal sugar group in which the 5'-carbon is not linked to another residue, whilst the other end contains a 3'-carbon that is not linked to another nucleotide. It is thus possible to define distinct ends of a nucleic acid chain; the 5'-end and the 3'-end. The phosphodiester bond is a group of strong covalent bonds between the phosphate group and a sugar, meaning that the bonding between groups in the molecules backbone is strong compared to the weaker hydrogen bonding between complementary base pairs. This keeps the molecule together during processes such as replication and transcription, when the base pairing is broken.

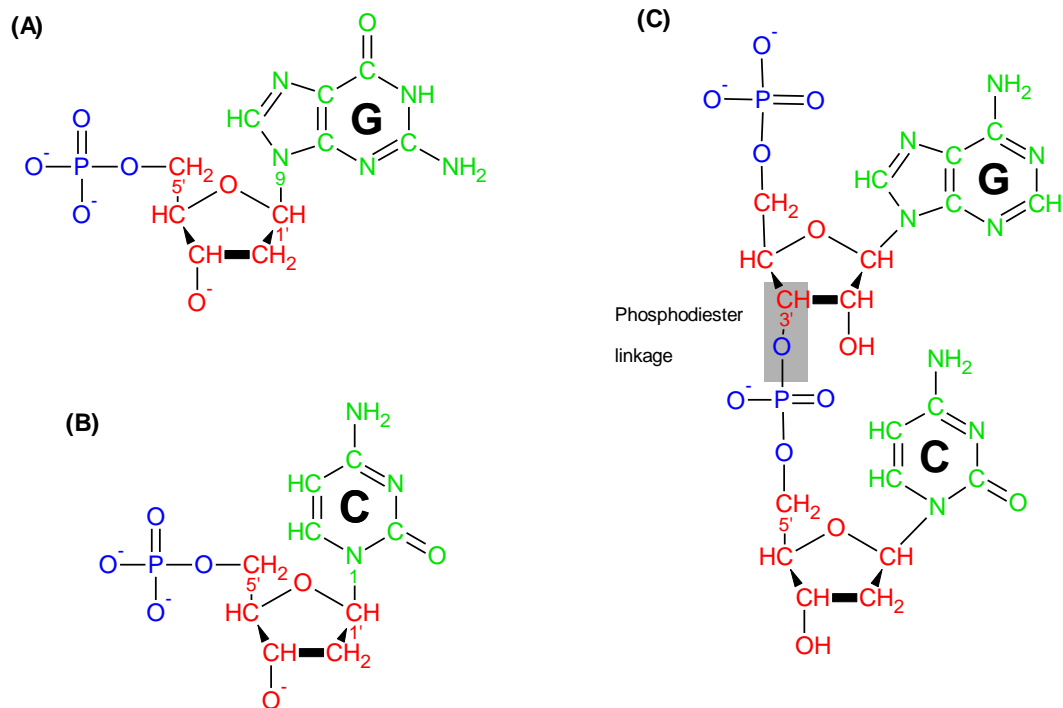


Figure 1-3. DNA nucleotides: (A) Deoxyguanosine monophosphate (contains the purine guanine), (B) Deoxycytosine monophosphate (contains the pyrimidine cytosine). The base (shown in green) is attached to a sugar pentose ring (red), which is in turn bonded to a phosphate group (blue). (C) Nucleotides link together via a phosphodiester bond to form nucleic acids. Many of these linkages come together to form a DNA chain.

1.1.3 The DNA double helix

It took some time from DNA first being discovered until its structure was finally elucidated. The most important early method used to probe DNA structure was x-ray diffraction. In an x-ray crystallography experiment, a crystal is mounted between an x-ray source and an x-ray detector. The crystal diffracts the source beam into many discrete beams, each of which produces a distinct reflection on x-ray film. A distinguished early x-ray worker was Rosalind Franklin. She passed a beam of x-rays through aligned DNA fibres, and observed a distinctive cross-like diffraction pattern [70]. This particular diffraction pattern is typical of helical structures, and showed that DNA has a preferred orientation.

Edwin Chargaff was the first person to study the bases of DNA quantitatively. He isolated DNA from the nuclei and broke it down to its constituent nucleic acids, before using paper chromatography to separate the purines from the pyrimidines. The different solubilities of the two classes means that they travel up the paper at different rates. He then exposed the two components to UV light, and by virtue of the fact that each base absorbs light of a characteristic wavelength he was able to determine how much of each base was present in DNA. He saw that the amount of each base varied massively between different organisms, but that the amount of the purine adenosine (A) was very similar to the pyrimidine thymine (T), and similarly that the amounts of cytosine (C) and guanine (G) were almost identical [33, 34]. As such, the number of purines always equaled that of pyrimidines.

These two findings led Watson and Crick to propose their now famous double helical structure of DNA [209]. In their model, the sugar and phosphate molecules are oriented on the outside of the molecule, making the backbone. Two chains of nucleotides coil around each other to form a right-handed helix (Figure 1-4A). The two strands in duplex DNA are anti-parallel; that is the polarity of the phosphodiester bond is 5'→3' in one strand and 3'→5' in the opposite strand. The four bases are situated in the interior of the molecule, and are oriented in such a way that they can form hydrogen bonds with bases on the opposite strand. These base pairings are not random, and follow so called Watson-Crick base pairing rules. Adenine (A) can pair only with thymine (T), while guanine (G) can form bonds with cytosine (C) alone. Cytosine is linked to guanine by three hydrogen bonds, whereas only two hydrogen bonds join adenine to thymine (Figure 1-4B). This means that a C≡G bond is more stable than an A=T link. In fact, DNA molecules with a high occurrence of G-C base pairs can withstand higher temperatures before the two strands unwind and come apart, a process known as denaturing. The structure of DNA produces a largely hydrophobic interior consisting of DNA bases, and a hydrophilic exterior comprising of the sugar-phosphate backbone.

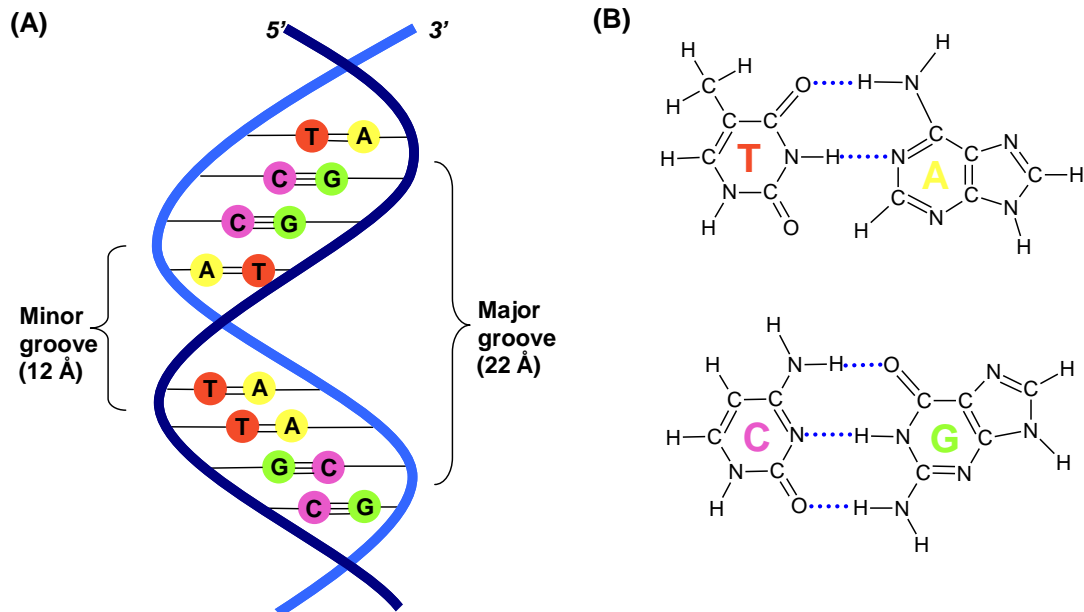


Figure 1-4. (A) DNA double helical structure consisting of two complementary DNA strands winding around each other. These consist of sugar-phosphate backbones and four different bases. The two strands are anti-parallel to each other, and their twisting forms spaces or grooves in the molecule. The major and minor grooves allow proteins to make contacts with exposed bases. (B) Watson-Crick base-pairs, showing hydrogen bonding as dotted lines. Thymine (T) bonds to adenine (A) by two hydrogen bonds. Cytosine (C) bonds to guanine (G) by three hydrogen bonds.

DNA can take various conformations and the kind described by Watson and Crick is known as the B-form (see Table 1-1 and Figure 1-5 for comparisons of main forms). B-DNA represents the conformation found most commonly *in vivo*, and consists of a right-handed helix with around 10 base pairs per helical turn, and a repeat distance between each turn of 3.4 nm. A right-handed helix observed along its helical axis spirals in a clockwise fashion away from the observer. The bases lie perpendicular to the helical axis, giving the molecule an overall diameter of 20 Å. In 1980, the structure of B-DNA was analysed to atomic resolution, and on the whole the findings agreed with the Watson-Crick model [211]. A dominant feature of B-form DNA is the presence of two distinct grooves running around the surface, a major and a minor groove (Figure 1-4A). These two grooves provide very distinct topographical features with which proteins can interact.

Table 1-1 Structural features of the main forms of DNA

DNA Conformation	Helix handedness	Bp/ Turn	Helix diameter (nm)	Major groove	Minor groove
B	Right	10.5	2	Wide and deep	Narrow and deep
A	Right	11	2.6	Narrow and deep	Wide and shallow
Z	Left	12	1.8	Flat	Narrow and deep

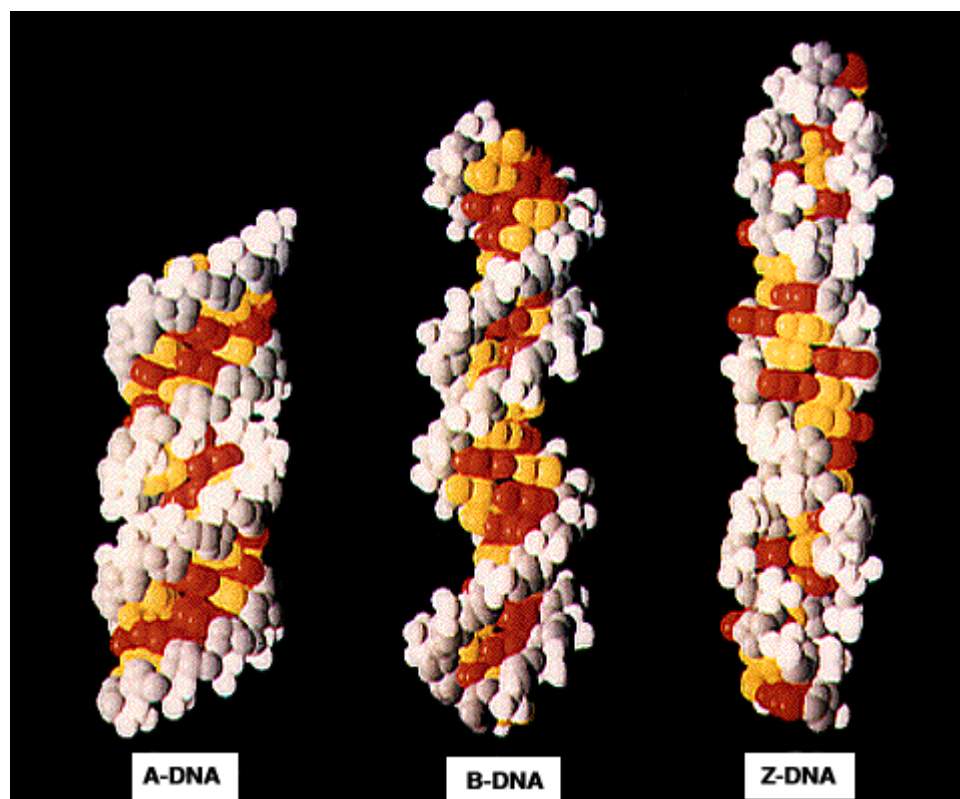


Figure 1-5. Comparison of the A, B and Z forms of DNA. B- and A-DNA are right handed with around 10 and 11 base pairs per helical turn. B-DNA is thinner than A-DNA. Z-DNA is left handed with 12 base pairs per turn. The different grooves at the sides of the helices can be seen. The major and minor grooves of B-DNA are especially prevalent [1].

In her experiments Franklin actually observed two distinct diffraction patterns for DNA, termed the 'A' and 'B' form. What form the diffraction pattern took was dependent on the humidity in which the experiment was performed, with the 'B' pattern being obtained at high humidity (e.g. 92 %) and the 'A' pattern being identified at a lower humidity of around 75 % [70]. The 'B' pattern was later

analysed to give rise to the structure of B-form DNA. Compared with B-form DNA, the A-form helix is broader (2.6 nm in diameter) and the bases are tilted and lie well off the helix axis. A-DNA contains a narrow and deep 'major' groove and a shallow and wide 'minor' groove. Hence it is more appropriate to refer to the A-DNA grooves as deep and shallow [141]. A-DNA is structurally similar to double-stranded RNA and may participate in RNA-DNA binding.

By 1980, chemists had learned how to synthesise DNA chemically in large amounts, and how to purify it so that it was possible to grow crystals of short base sequences. The sequences CGCG and CGCGCG, as analysed independently, each crystallised as left-handed double helices in the Z-form [57, 208]. A left-handed Z-form helix observed along its helix axis spirals in an anti-clockwise fashion away from the observer. For years it was assumed that DNA could only be right-handed. Earlier work by Fritz Pohl and Tom Jovin, had suggested that alternating C-G sequences such as CGCG might be either right-handed or left-handed depending on the salt concentration [155], but only a few specialists had taken them seriously.

The Z-form is also distinguished by the zigzag path of the sugar phosphate backbone. The Z conformation can be found in DNA molecules with alternating purine-pyrimidine sequences and is favoured by conditions such as high salt. The Z-DNA helix, at 1.8 nm in diameter is narrower than in B-DNA, and the helical repeat is 12 bp per turn compared with 10.5 bp per turn.

1.1.4 RNA structure

DNA carries the genetic information, but the reading of this code requires a different kind of nucleic acid, namely ribonucleic acid (RNA), which is subsequently translated by ribosomes to form a functional protein. RNA forms very different structures to DNA, a direct consequence of the additional hydroxyl group present on the ribose sugar of an RNA nucleotide (see riboses in Figure 1-2). This prevents the molecule from forming a stable double helix, meaning that RNA exists as single-stranded molecules. It is therefore possible for distant regions of the chain to interact via Watson-Crick base pairing.

Depending on its primary sequence, RNA can form a number of secondary structures including hairpins, loops and junctions [15, 35]. The structure of RNA perhaps has more in common with proteins than DNA, as an RNA chain will fold up to form a complex 3D shape.

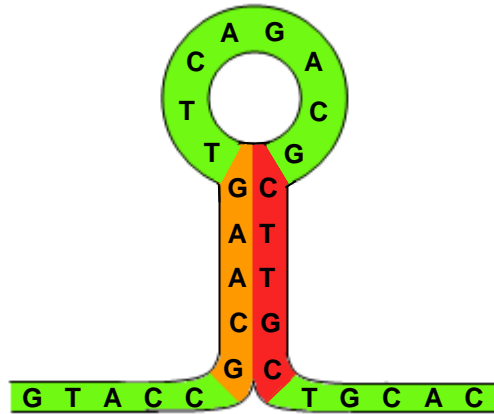


Figure 1-6. Example of an RNA stem loop. Due to its single stranded structure and inherent flexibility distant segments of RNA consisting of inverted bases can form base pairs. Here, two regions (red and orange) come together to form a stem, while the region in between forms a circular loop.

The variety in RNA structure and function is illustrated by comparing the three main RNA types found in the cell. Messenger RNA (mRNA) is a copy of a DNA region which encodes a specific amino acid sequence. Transfer RNAs (tRNA) are small adaptor molecules that carry amino acids to ribosomes, and have a structure comprised of stems and loops (see Figure 1-6), which can interact with distinct elements. Ribosomal RNA (rRNA) is the main constituent of the ribosome, and shows that RNA can also function as a catalyst and not just as a structural element. Ribosomes facilitate the interaction between mRNA and tRNA, allowing amino acids to be joined together, resulting in protein translation.

1.2 DNA Transcription

1.2.1 The central dogma of molecular biology

DNA's main functions are in the storage of genetic information used for encoding proteins, and the ability to pass-on this information to future generations. All cellular processes depend on proteins in some way, and as they carry such an importance for life they must be constantly produced in a regulated and reproducible manner. The fact that DNA is made of two strands of complementary shapes lies at the heart of its ability to convey the genetic information [53, 209]. The information flow from DNA to RNA to protein was first described by Crick, and became to be known as the central dogma of molecular biology (Figure 1-7) [52].

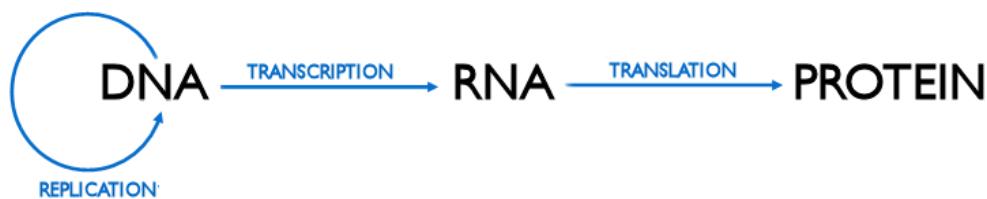


Figure 1-7. The central dogma describes the residue by residue transfer of sequence information between the information-carrying biopolymers. This ultimately leads to the expression of genes into proteins.

If the hydrogen bonds connecting the two chains of DNA are broken, then the helix will unwind and the strands will separate. The two resultant single-stranded DNA molecules can act as templates for the formation of new complementary chains, and thus two identical daughter strands can be formed from the original double helix. During cell division DNA replicates itself to form two DNA molecules with identical sequences, allowing the information needed to produce proteins to be passed on to future cell generations.

The sequence of bases in a long chain of DNA carries the information needed to ultimately determine specific amino acid sequences, and hence protein structure and function, and a segment of DNA that contains the information needed to make a particular protein is known as a gene. DNA is the starting point for protein expression and the information contained within its base

sequence is passed on during the process of transcription; a process whereby RNA polymerases (RNAPs) use DNA as a template for the synthesis of messenger RNA (mRNA). The mRNA is a direct copy of the coding DNA strand. The mRNA is then processed, before moving outside the cell nucleus where it is translated into chains of amino acids (proteins) by the ribosomes.

1.2.2 Introduction to DNA Transcription and RNA polymerase

Transcription is a central step in the process of gene expression and its study by AFM is the focal part of this thesis. DNA transcription is carried out by the enzyme RNA polymerase (RNAP), a protein first described in 1960 [96]. RNAPs can occur as both single and multiple subunit enzymes. Bacteriophages and mitochondria contain RNAPs of the single-subunit variety, whilst polymerases from bacterial, archaeal, and eukaryotic cells are representative of the multi-subunit class. Of the multi-subunit RNAPs, perhaps that of *E.coli* is the best-studied. These enzymes are constructed from four subunits: β containing 1407 amino acids, β' (1342 amino acids), and a dimer of α (329 amino acids) [55].

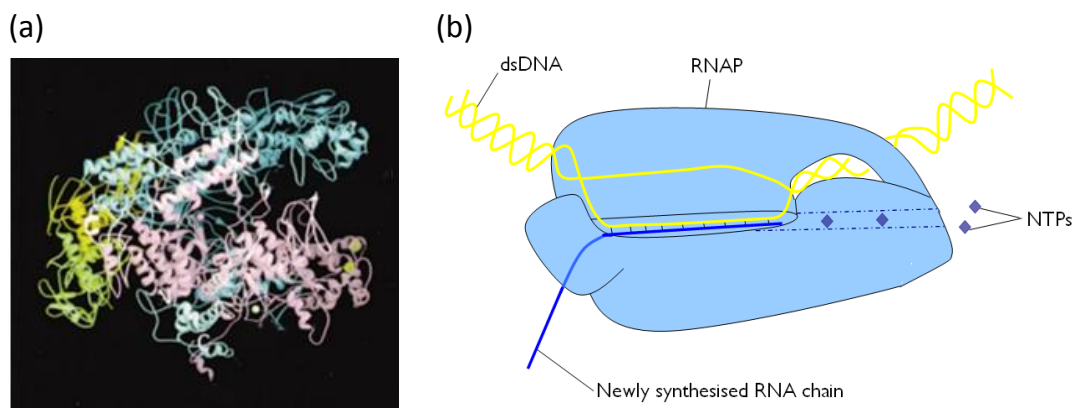


Figure 1-8. (a) X-ray crystal structure of *Thermus aquaticus* core RNA polymerase. Adapted from Zhang *et al.* [219] (b) Schematic diagram of RNAP-nucleic acid complex. The enzyme acts to unwind the DNA helix. The structure resembles that of a hand, with a region of single stranded DNA lying in its “palm”. NTPs can enter through a channel in the interior of the enzyme. RNAP catalyses the joining of these nucleotides, using DNA as a template to form a complementary chain of RNA, which exits the enzyme under its “thumb”.

Zhang *et al.* used X-ray crystallography to reveal the structure of RNAP from *Thermus aquaticus* (*Taq*), with a resolution of 3.3 Å [219] (Figure 1-8a). The molecule is characterised by its “crab-claw” shape (Figure 1-8b), housing a groove running the entire length of the inner surface. This is a distance of approximately 30 Å, allowing around 9 base pairs (bp) of DNA to be accommodated. The active site of the enzyme is located deep within the molecule, where there are also binding sites for catalytic Mg(II) ions. In addition to the main channel there is a secondary pore, which allows individual nucleotide triphosphate (NTP) molecules access to the active centre of the RNAP. The crystal structure of *Taq* RNAP corresponds well with the structure of *E.coli* RNAP obtained by electron crystallography [156], meaning that functional data obtained from experiments on *E.coli* RNAPs can be explained using the high resolution structure of *Taq* polymerase. The dimensions of *E.coli* RNAP as revealed by electron crystallography at ~25 Å resolution, have been found to be ~100 × 100 × 160 Å [55].

1.2.3 Transcription Initiation

The process of transcription can be separated into three main stages:

- Initiation
- Elongation
- Termination

In the initiation stage, the RNA polymerase is able to recognise a specific DNA sequence, termed the promoter region. This region lies just upstream of the main coding sequence of the gene. At this point the polymerase acts to melt, or separate, the two DNA strands to form a “transcription bubble”. Certain DNA sequences are highly conserved and are found in the promoter region of most genes. For example, the commonly occurring TATA box is an AT-rich sequence, occurring around 30 bp upstream from the transcription start site [43]. Promoter recognition is aided by specific proteins called σ -factors, which recognise distinct DNA sequences. A major protein involved in the *E. coli* transcription cycle is σ^{70} , with the superscript denoting the molecular weight in

kDa [130, 163]. DNA binding-proteins such as transcription factors are typically a pre-requisite for the cell to become transcriptionally-active.

For transcription to initiate, the polymerase must first associate with the small region of DNA known as the promoter. This is an extremely short area of DNA when compared with the full length of a typical gene. If the polymerase were to diffuse in three dimensions until it encountered the promoter site, then the time taken would be prohibitively long and transcription would not proceed at the required rates for adequate protein expression. It has been suggested that if a diffusion-controlled promoter search were to occur in a space of reduced dimensionality, then the efficiency could be increased by orders of magnitude [206].

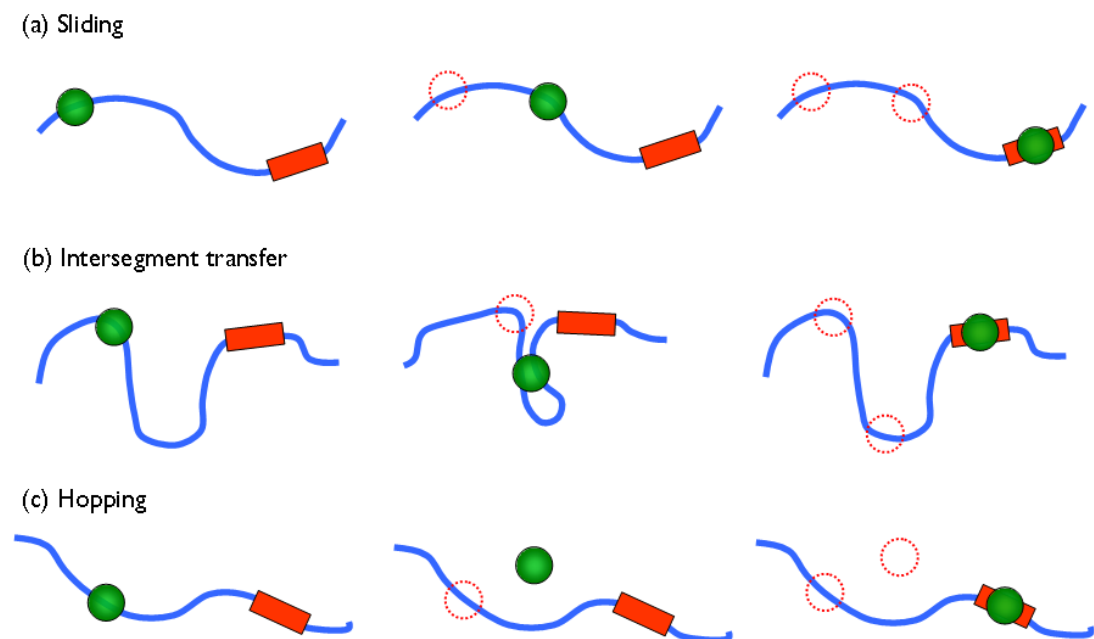


Figure 1-9. Schematic diagram of the different methods of promoter location. The DNA is shown as a blue line, the RNAP as a green sphere, and the promoter region as a red rectangle. The RNAPs progress is indicated by dashed circles. (a) In sliding, the RNAP diffuses one-dimensionally along the DNA contour. (b) In intersegment transfer the RNAP is able to reach a distant site on the DNA via the formation of an intermediate complex. (c) During hopping the RNAP loses contact with the DNA, before re-attaching further along the chain.

A number of mechanisms have been suggested to explain the increased binding rates associated with promoter regions (Figure 1-9). If the RNAP were to have an affinity for non-promoter DNA, then the likelihood of non-specific binding to the DNA would be high. The polymerase could then undergo one-dimensional diffusion along the chain, until it reached the target sequence at the promoter [20]. This method is known as *sliding* (Figure 1-9a). Another mechanism, known as *intersegment transfer* (Figure 1-9b), involves the transfer of the polymerase from one position on the template to a more distant segment by means of an intermediate state where the DNA is looped, such that the protein is associated both to the initial and distant regions at the same time [206]. Multiple transfer events occur until the promoter region is reached. A final process, called *hopping* (Figure 1-9c) has also been put forward, in which the polymerase bounces along the template DNA until it finds the target site [206]. All three of these mechanisms serve to reduce the dimensionality of the promoter search process.

Once the RNAP, either alone, or in conjunction with a σ -factor, becomes bound to its promoter region, the complex undergoes a conformational change. The RNAP acts to orientate itself so as to trap the DNA into its “claw”. During this process the system moves from what is termed a closed promoter complex (CPC), to an open promoter complex (OPC), with the nomenclature denoting the state of the “transcription bubble”. The resulting complex is relatively stable. The initial binding is to double-stranded or closed DNA but the reorientation of the RNAP is accompanied simultaneously with the unwinding of a 10-15 bp segment of DNA, to form the single-stranded “transcription bubble” or open complex [56]. The specific binding of the RNAP at the promoter site alone is not always enough for the formation of open complexes. Other DNA-binding proteins (σ -factors) help to enhance the reaction. A major σ -factor in *E. coli* is σ^{70} , which can act solely to unwind the template without the need for any additional protein factors [54].

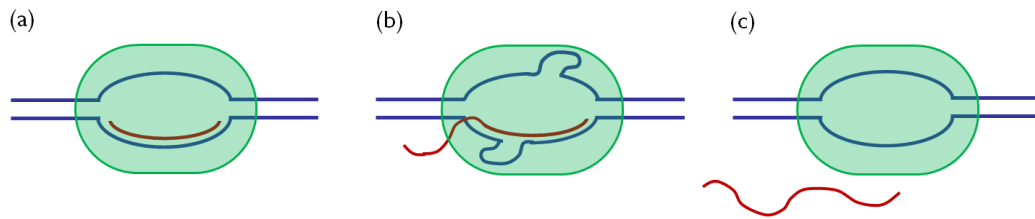


Figure 1-10. Scrunching model of abortive initiation. (a) NTPs are added end-to-end on to form an RNA chain using single-stranded DNA contained within the transcription bubble as a template. (b) The RNAP acts to pull in downstream DNA to its core, such that it becomes scrunched after successive nucleotide additions to the RNA chain. (c) The abortion of the transcripts leads to the expulsion of the nascent RNA chain and the release of scrunched DNA.

Next the complex undergoes a process known as abortive initiation, whereby competition exists between short RNA oligo release and continuous NTP incorporation ultimately leading to a completed messenger RNA chain [31]. The σ -factor remains bound to impart stability to the RNAP and prevent disengagement from the template. Upon addition of NTPs, the RNAP catalyses the polymerisation reaction that forms an RNA transcript from the single-stranded DNA template within the transcription bubble. The necessity of the RNAP to produce many short transcripts coupled with the ability to quickly reinitiate when one synthesis has been aborted, stipulates that the active site of the enzyme must move stepwise along the DNA template whilst simultaneously remaining in close contact with the promoter region. This is a paradoxical problem in which a number of models have been proposed to come to a solution, including transient excursions [32] and “inchworming” [193]. Abortive initiation arises from an inability of an RNAP to break free of its stable contacts with the promoter. Straney and Crothers suggested that the energy required to break free might be stored in a “stressed intermediate”, and that abortive initiation was a direct consequence of this energy being used inefficiently [193].

The most likely solution is known as the scrunching model (Figure 1-10). Scrunching as a model for abortive initiation has been demonstrated through experiments utilising the single-molecule technique fluorescence resonance energy transfer (FRET) to study the separation of the RNAP core and downstream regions on DNA [100], and magnetic tweezers to detect the double

helix unwinding [161]. In the scrunching model, the RNAP functions as a rigid body, increasing its effective footprint by pulling DNA downstream of the promoter into its core such that it becomes “scrunched” after successive nucleotide additions. The abortion of nascent RNA transcripts leads to the downstream expulsion of the scrunched DNA, before it is reeled back in subsequent cycles. The scrunching model suggests that the energy cost of breaking base-pairs to form abortive transcripts can be transferred into a stressed intermediate, and that this free energy is high compared to the free energy of the RNAP-promoter interaction. It is proposed that this energy disrupts the interactions of the RNAP with the promoter, and drives the reaction towards the elongation stage [161].

Once the DNA transcript reaches ~12 nucleotides, the RNA-DNA complex becomes stable and the RNAP loses contact with its σ -factor, escaping the promoter. This point marks the end of the initiation stage and the beginning of chain elongation [92].

1.2.4 Elongation

During the elongation stage the RNAP proceeds along the DNA template, synthesising a chain of RNA in a processive manner as it travels down the strand. An elongating complex is an extremely stable structure, consisting of RNAP, DNA and a nascent chain of RNA, that can grow to be thousands of nucleotides in length. The RNA chain is constructed by a polymerisation reaction in which nucleotide triphosphate (NTP) monomers are hydrolysed, and added to the 3' end of the RNA chain. The hydrolysis of NTPs provides the energy to drive the reaction, and when a nucleotide is incorporated into the growing chain pyrophosphate PP_i is released. The RNA chain is a direct copy of the coding DNA strand, with the notable exception that the base thymine present in DNA is replaced by uracil (U) in RNA.

The elongation stage is highly complex, and consists of a series of "on pathway" events leading to successful incorporation of the correct NTP and translocation along the chain in the 5'→3' direction, and other "off-pathway" states, which are

a hindrance to transcriptional elongation, such as pausing [88]. Each cycle of NTP incorporation consists of several steps, including movement of the RNAP from one base to the next, NTP binding, and hydrolysis. Translocation of the RNAP along the chain must be directly coupled to the nucleotide incorporation condensation reaction, and the catalytic abilities of the enzyme. Different models have been put forward to explain transcription elongation. It has been suggested that the release of PP_i can promote the RNAP to undergo a conformational change, which induces forward translocation of the enzyme by a single base pair, and thus for each nucleotide addition there is a single movement event of the RNAP along the chain [215]. As such, this powerstroke mechanism states that forward translocation is directly linked to substrate hydrolysis.

There is other evidence that RNAP can slide back and forth along a DNA chain via Brownian motion, prior to nucleotide addition. Uni-directional motion is possible because the RNAP functions as a Brownian ratchet [80]. This model incorporates the concepts of a ratchet and pawl, in which certain movements of a cogged wheel (ratchet) are restricted by a wedge (pawl), such that any motion is uni-directional. A pawl can either be stationary, with the wedge inserted in the teeth of the cog preventing reverse motion, or reciprocal, allowing uni-directional motion of the wheel. The Brownian ratchet mechanism for transcription elongation proposed by Bar-Nahum *et al.* considers the elongating complex as consisting of two ratchets (one stationary and one reciprocating situated within the same catalytic site), and that no energy other than thermal fluctuations are responsible for translocation [13].

The stationary pawl represents the incoming complementary substrate. The addition of each complementary NTP prevents the 3' terminus of the nascent RNA chain from occupying this site, suppressing back-tracking. From this point, the elongating complex catalyses the phosphodiester bond formation, extending its transcript by one nucleotide. Structural data of the elongating complex provides evidence for a flexible unit existing within the interior of the enzyme. The so called F bridge helix was observed to be bent in the crystallographic structure of bacterial RNAP [219] and straightened in that of yeast RNAP II [46].

Gnatt *et al.* subsequently proposed that bending of the F bridge induces translocation of the RNAP by one nucleotide [78]. In the Brownian ratchet model, the oscillating F bridge domain represents a reciprocating pawl. Upon bending, the F bridge acts to eject the 3' end of the RNA chain from the catalytic center. Subsequent straightening of the F bridge allows the next corresponding substrate to enter the now empty site. In short, the incoming substrate acts to restrict backward tracking of the RNAP, whilst the oscillating F bridge allows the RNAP to achieve uni-directional motion along the chain and initiates the next cycle of nucleotide addition. Hence, RNAP motion is controlled by the thermal oscillation frequency of the F-bridge. After binding of the correct NTP the centre of the Brownian oscillation is moved 1 nt downstream.

At each step of elongation there is competition between addition of the correct nucleotide, and other "off-pathway" events not conducive to RNA chain formation. These include pausing (a transient conformation incapable of elongation) and arrest (a conformational state incapable of nucleotide addition without the factor-assisted isomerisation back to an active complex). The enzyme RNAP moves along a varying DNA sequence, and one consequence of this changing substrate is that transcription does not proceed at a constant rate, but rather becomes sequence dependent [12]. RNAP actively engaging in transcription elongation tends to dwell at certain regions, known as pause sites. It is believed that *in vivo* these pause sites function to regulate gene expression, by allowing transcription factors to bind to stationary RNAP and modify its activity [120]. However, other pause sites are observed to have no biological function. Pausing represents an alternate branch in the reaction pathway, and it has been suggested that the phenomena arises from misalignment of the RNA 3' end with the enzyme active site [109, 110]. Pausing is frequently observed in experiments under conditions of low NTP concentration, where the slow rate of the on-pathway reaction may drive the reaction into the branched pathway [207].

Experiments with single molecule optical tweezer techniques demonstrated numerous short pauses (< 10s) that were weakly sequence dependent [3, 145]. These short or ubiquitous pauses occurred roughly every 100 bps on regions

devoid of the much stronger regulatory pauses. It has been suggested that long pauses may be preceded by numerous short pauses occurring in close proximity, that eventually lead to the stabilisation of the paused state [9].

RNAP must have efficient proof-reading capabilities to ensure accurate expression of a particular gene. The model for proof-reading by RNAP suggests that incorporation of the incorrect nucleotide causes the molecule to back-track along the chain [110]. This event is followed by endonucleolytic cleavage of the 3' end of the RNA chain carrying the error. This duty is performed by the transcription factors GreA and GreB in prokaryotes [195].

1.2.5 Termination

During termination transcription is halted, the RNAP dissociates from the DNA template, and the RNA transcript is released. This can be caused by the RNAP encountering a particular DNA sequence, known as a terminator. Bacteria possess two types of termination strategies, termed rho-dependent and rho-independent (also called intrinsic termination) [210].

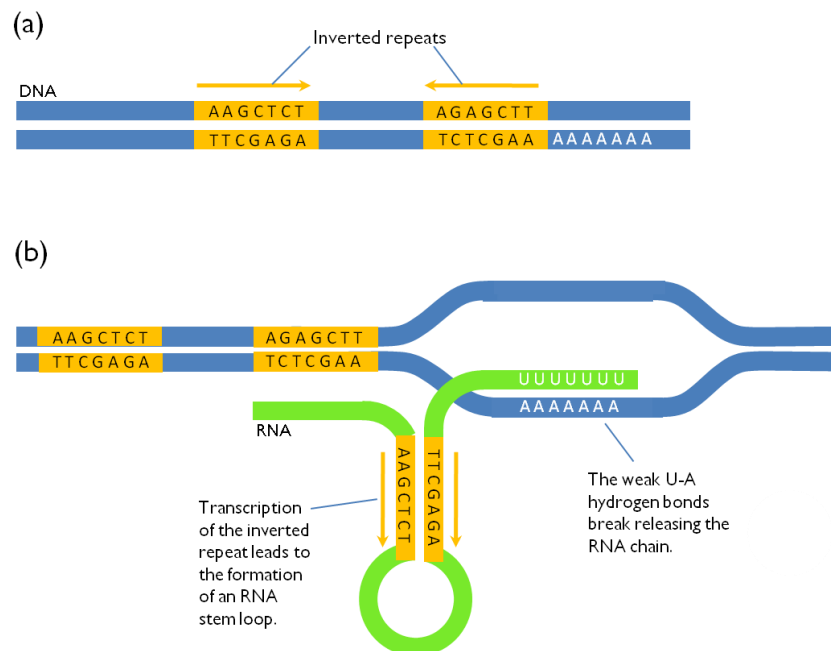


Figure 1-11. Method of rho-independent termination. (a) This arises due to the presence of inverted sequence repeats, followed by a string of A bases. (b) Transcription of this sequence leads to a formation of an RNA stem loop, and weak U-A bonds within the catalytic center of the enzyme. The complex is destabilised and loses contact with the template strand, releasing the chain of RNA in the process.

During rho-independent termination, transcribing RNAP encounters a sequence (terminator) that is then followed shortly after by an inverted repeat of the same sequence (see Figure 1-11). The repeat patch is then followed by a short string of adenine nucleotides. Transcription of the inverted repeat leads to the formation of an RNA stem loop and the pausing of the transcription complex. The string of adenines are transcribed into uracil monomers in the RNA, meaning that there is a region of weak hydrogen bonding in the RNAP catalytic center. The formation of the hairpin structure has a destabilising effect on the elongation complex, and the RNA chain is easily released from the DNA template, thus terminating transcription [214].

Termination can also be controlled by protein factors [162]. Rho-dependent termination is so named as it is dependent on a small factor called rho [62]. Rho acts to unwind the DNA-RNA hybrids formed during transcription, thus facilitating the release of the RNA transcript.

Following termination, and whilst still in the nucleus, the RNA undergoes post-transcriptional processing including capping, poly-adenylation and splicing, before it is recognised by nuclear membrane receptors and transported to the cytoplasm. Here translation of the RNA transcript is performed, which ultimately leads to the construction of protein.

1.3 Nested genes and convergent transcription

The arrangement of genes in DNA is complex. One stretch of DNA consists of non-coding regions (introns) that are removed and not translated into proteins, and coding regions (exons) present in the mature form of mRNA. The organisation of genes follows a complex spatial arrangement, and a particular stretch of DNA can encode more than one product. For example, genes can overlap on opposite strands. One extreme example of overlapping genes is that of nested genes (Figure 1-12).

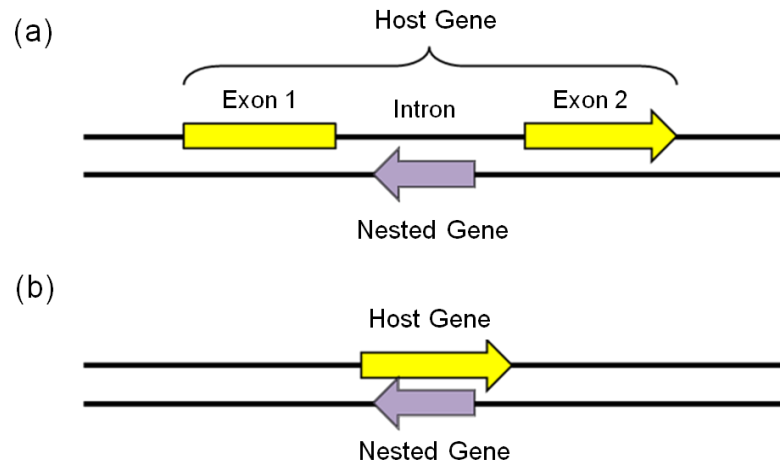


Figure 1-12. A nested gene is entirely located within the boundaries of another host gene. (a) Example of an intronic located nested gene (shown in purple). The host gene is shown as yellow arrow. (b) A less common case is the nested gene located within an actual coding region of host DNA.

The nested gene refers to a gene that is located within the boundaries of another gene, often within an intron and in the opposite orientation. Nested genes were first discovered in *Drosophila* flies, when it was shown that the gene *Pcp* encoding pupal cuticle protein was found situated within an intron of adenosine 3 (*ade3*), lying on the opposite DNA strand [87]. *Ade3* acts as a housekeeping gene and is involved in basic functions needed for the sustenance of the cell while the cuticle protein gene is associated with the development of tissue. Since then, other nested genes have been discovered, including a number in the human genome [218].

There are two main classes of nested genes: (i) Genes nested within the introns of larger genes (Figure 1-12a); (ii) Genes which lie directly opposite or within the coding region of the host gene (Figure 1-12b). The first class of nested gene is much more common. Yu *et al.* identified 158 genes nested within introns predicted to encode protein. Over half of the human nested genes were observed on the strand opposite the host gene, in an antiparallel arrangement. The remaining genes were situated on the same strand in the same direction to the host gene. Human nested genes are typically situated on large introns [218].

It has been shown by Asis *et al.* that the majority of nested genes arise from the insertion of a DNA coding sequence into a pre-existing intron [10]. Further

analysis of nested protein coding genes in vertebrates, fruit flies and nematodes revealed substantially higher rates of evolutionary gains than losses. The increase in nested genes could not be attributed to any functional relationship between the genes involved, and thus it was concluded that the evolution of animal genomes is accompanied by an increase in organisational complexity [10].

There may be a number of explanations for the existence of nested genes. Gibson *et al.* have reviewed the biological implications of nested genes [75]. They suggested that natural selection may favour the presence of nested genes which encode proteins of similar function, in order to enhance a biological response. While nested genes with similar function have been reported, work by Yu *et al.* suggests that this is the exception rather than the norm [218]. In their study of the human genome using Gene Ontology annotations, only five anti-parallel and one parallel nested gene exhibited similar functions.

The presence of nested genes in introns raises the question that they may be co-regulated and transcribed simultaneously. For example, the nested genes originally discovered in *Drosophila* are both induced during development. It is also possible that oppositely aligned nested genes may be down-regulated. A RNA polymerase moving along a host gene may experience steric hindrance from other RNAPs or transcription factors associated with the nested gene.

The functional benefits of nested genes are still unclear. In prokaryotes a compact genome is advantageous, allowing replication to be carried out at a faster rate. It is possible that nested gene expression could be co-ordinated with their host genes, both positively and negatively. The nested gene provides a mechanism for the co-ordinated expression of proteins with similar functions, which could act to enhance a biological response [75]. However, many nested genes exhibit entirely different functions and gene expression patterns from their hosts. Kumar suggested that the acquisition of nested genes is an evolutionary neutral process aided by the presence of numerous introns that provide a pathway for gene insertion [111].

Simultaneous expression of both nested and host genes containing two oppositely aligned promoter sites would require a process of convergent transcription [135]. This involves separate polymerases binding to two different promoter sites, and then beginning to transcribe their respective genes. If the two promoter sites are aligned on opposite DNA strands, the transcribing enzymes will move ever closer to each other, until they become so close that they will begin to affect each other, a situation of transcriptional interference (TI). This is a consequence of the two strands in DNA being anti-parallel. If the nested gene was on the same DNA strand as the host then the simultaneous expression of both both genes would lead to both RNAPs transcribing in the same direction. In which case, a collision will only occur if the front RNAP pauses or stalls, before the other RNAP runs into the back of it.

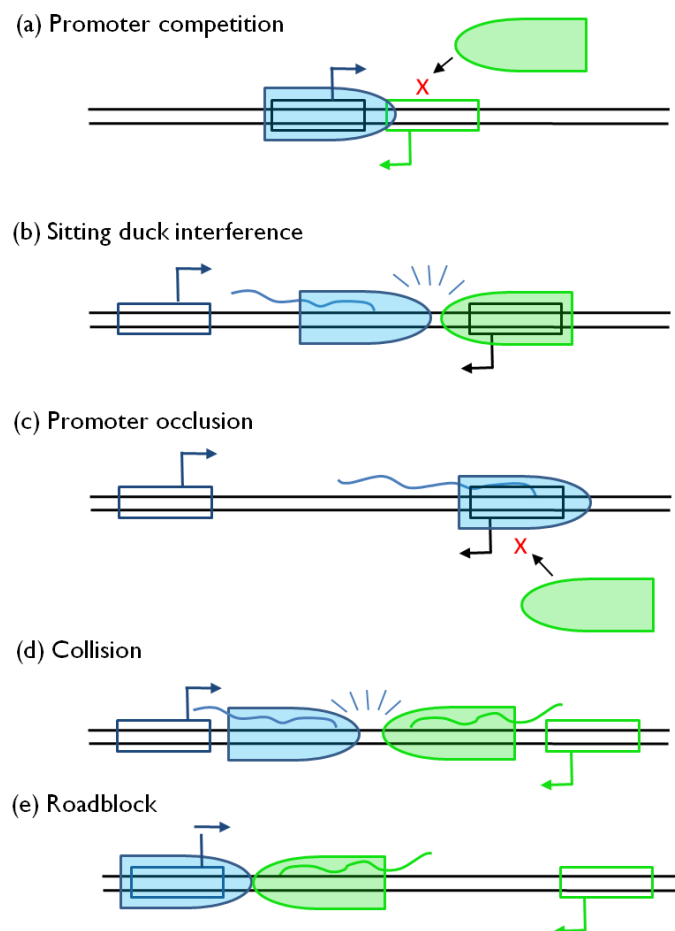


Figure 1-13. Methods of transcriptional interference (TI). See below for further explanation.

Transcriptional interference (TI) refers to the direct negative impact of one transcriptional activity by a second activity. TI is often asymmetric and results from the existence of two promoters (the strong reducing the weak), and has been demonstrated in both eukaryotes and prokaryotes. Shearvin *et al.* outlined a number of potential mechanisms of TI [183], three implemented at the initiation phase and two at the elongation phase (see Figure 1-13 also):

- (a) Promoter competition: TI arises from the occupation of the promoter region by one RNAP preventing the occupation of a second promoter by another RNAP. This is valid for over-lapping or closely spaced promoters.
- (b) Sitting duck: A sitting duck (SD) is defined as an RNAP that is situated at a weak promoter that is slow to transit to an elongation complex. A sitting duck is prone to being hit and dislodged by the arrival of an elongation complex "fired" from a second promoter.
- (c) Occlusion: TI from occlusion arises as a result of the inability of one RNAP to bind to the chain due to a second actively elongating complex (EC) traversing across its promoter region.
- (d) Collisions: Collision between converging elongation complexes could lead to the halting of transcription elongation for one or more of the complexes.
- (e) Roadblock: An open complex bound at one promoter could act as a roadblock to an elongation complex originating from a second promoter.

Callen *et al.* used a LacZ reporter assay to study methods of transcription interference between promoters of unequal strength [30]. They observed that the activity of the weaker promoter was very sensitive to TI. Mechanisms involving RNA-RNA hybridisation and promoter competition were excluded. However, the addition of a terminator sequence, such that transcription from the strong promoter was halted before the weak promoter was reached, led to a reduction in interference. This indicated a TI requirement of elongation across the promoter. Due to the speed of elongation, effects due to promoter occlusion were negated, and as a result a "sitting duck" model of TI was proposed.

Sneppen *et al.* developed a mathematical model of convergent TI incorporating occlusion, collisions between elongating RNAPs, and SD collisions. The models in the prediction agreed well with the *in vivo* results of Callen *et al.* [30]. It was assumed that both RNAPs disengaged during collisions between ECs, whilst only the SD was knocked off in SD collisions. In their study, SD collisions were observed to be the strongest mechanism of TI. This was dependent on two conditions, (i) One promoter was stronger than the other, meaning it fired preferentially and (ii) the aspect ratio (relative rates of RNAP binding and firing) of the sensitive promoter was sufficiently large enough, such that there is a high likelihood of an RNAP being situated at the promoter when the other aggressive promoter fires. Collisions between elongating RNAPs persist if the promoters are of equal strength and aspect ratio, whilst also increasing in occurrence as the distance between promoters increases. Occlusion was only observed if the interfering promoter was very strong, and the separation was small.

Gibson *et al.* suggested that an RNAP transcribing a host gene would experience steric hindrance from either RNAPs or regulatory proteins, relating to a nested gene, in which a number of outcomes could be considered [75]:

- (i) After the approach and “collision” event, one RNAP disengages. The successful by-passing of one RNAP could lead to expression on one gene.
- (ii) Both RNAPs approach and stop, before finally disengaging.
- (iii) The RNAPs are able to by-pass each other, perhaps leading to expression of two fully formed genes.

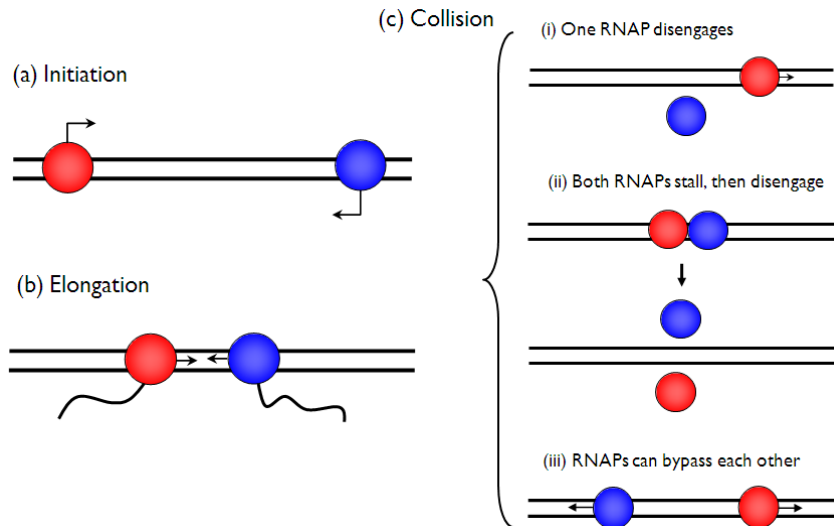


Figure 1-14. Different stages involved with convergent transcription. (a) Convergent transcription begins with two promoters, situated on opposite strands, recruiting RNAPs. (b) After firing, the two RNAPs enter the elongation stage and begin to approach each other. (c) Eventually there will be a head-on collision. Collision outcomes suggested by Gibson *et al.*: (i) One RNAP is knocked off the template, leading to the possibility of expression of one gene only. (ii) The RNAPs stall against each other, disengaging after a period of time. (iii) The RNAPs are able to bypass each other, perhaps allowing both genes to be expressed unhindered.

These outcomes are summarised in Figure 1-14. Gibson *et al.* also put forward an approach, during which AFM could be used to characterise interactions between DNA and RNAP during convergent transcription [75]. They proposed using nested gene models consisting of two promoters in opposite orientation with a cDNA cloned between them. The ability to observe events occurring at a single genetic locus by AFM was suggested as an improvement over studies utilising biochemical techniques such as the reverse transcription polymerase chain reaction (RT-PCR). The central aim of this thesis is to use AFM techniques to study mechanisms of transcription interference, as would arise from convergent transcription of nested genes. Complex processes such as these benefit from the use of single-molecule techniques, where distinct members of an inhomogeneous population can be observed.

Chapter 2

2 Introduction to AFM of single molecule DNA systems

2.1 Overview of AFM

Atomic force microscopy (AFM) is an extremely versatile technique that is particularly suitable for studying biological material. Since the instrument's inception perhaps the most widely studied biological sample has been DNA. AFM is one of few techniques that allows direct visualisation of the molecule, and has the added advantage that sample preparation is relatively quick and simple. DNA can be imaged directly on its own, without the need for complex preparation methods which can alter the molecule's biological function. It is possible to use transmission electron microscopy to visualise DNA but the molecule must typically be labelled with heavy atoms to achieve sufficient resolution [159]. The instrument's versatility can be highlighted by the wide range of studies with regard to DNA that the technique has permitted. For example DNA structure, condensation, and its interactions with a number of different proteins have all been studied extensively [5, 12, 14, 103, 196]. This thesis focuses on a particular area of study, namely the study of DNA transcription by AFM. DNA transcription, the process by which RNA polymerase travels along a DNA chain unwinding the helix, and synthesising an RNA chain using single stranded DNA as a template, is a vitally important step in gene expression before translation into proteins. Transcription is not a simple one-step process but is composed of a number of distinct sub-stages, and it is a testament to the instrument that many of the steps in the transcription pathway can be studied by AFM. Being a single-molecule technique, AFM allows new insights into transcription to be gathered, results that would not be possible with some of the more traditional biochemical techniques.

2.1.1 Introduction to scanning probe microscopy

The AFM belongs to a family of instruments known as scanning probe microscopes (SPMs). These devices all consist of a probe specifically designed to measure a particular physical quantity. Images are obtained by mechanically moving the probe with respect to the sample surface. Probe-surface interactions are recorded at distinct points, allowing a picture of the surface to be constructed. The first instrument of this type was the scanning tunnelling microscope (STM), developed in the early 1980s [22]. The principle of this technique is the detection of tunnelling currents between a sharp tip (probe) and the sample surface. The tunnelling current is exponentially dependent on the tip-surface separation making the device very sensitive to topographical features. STM has been used to image DNA placed on a conducting highly oriented pyrolytic graphite (HOPG) surface [58], allowing atomic-scale information to be gathered. However it has been suggested that the use of graphite as a substrate gives rise to structures that are periodic, and that could be misidentified as the pitch of DNA [41]. The principle drawback of this technique is the necessity that the sample under investigation be electrically conductive. This limitation led to the developments which gave rise to the invention of AFM by Gerd Binnig, Calvin Quate and Christoph Gerber in 1986 [21]. AFM works by measuring the deflection of a cantilever with response to surface features, events arising from ubiquitous inter-molecular forces, and hence any sample surface can be examined. Molecular and even sub-molecular imaging [172] have been shown to be possible. The main advantages over other techniques are the ability to investigate biological samples in their native states, without the need for complex sample preparations (e.g. fluorescent staining) that may damage or hinder the samples' biological activity. It is possible to perform imaging in both air and liquid [86, 157], with the latter allowing measurements to be taken in an environment similar to physiological conditions.

2.1.2 AFM Instrumentation

Central to AFM is the measuring probe, which consists of a cantilever with an integrated tip at its end that interacts with the sample surface. These cantilevers are micro-fabricated and vary in design depending on the substrate and imaging mode used. Common models are the “diving-board” shapes used for tapping mode and the triangular ones used for contact mode AFM. They are usually constructed out of silicon and silicon nitride, and the tips have a radius of curvature of the order of nanometres. Typical spring constants range from 0.01 to 100 Nm^{-1} , enabling a force resolution of the order of piconewtons to be obtained [5].

AFM works by observing small deflections of the cantilever caused by small surface forces (see Figure 2-1 for schematic diagram depicting the typical AFM setup), and as such it is important to have a detection method that is able to resolve the small changes to the probe. The first AFMs worked on detecting the tunneling current between an additional tip on the back side of the lever and the main body [21]. The exponential distance dependence of the tunneling current allows for high sensitivity to be obtained, however the tunneling tip can exert extraneous forces on the cantilever which can distort the actual force measurements. Optical interferometry has also been used to measure the cantilever deflection [127, 128]. This method involves a laser being reflected off the cantilever and interfering with part of the beam reflected from a flat reference. The cantilever deflection has also been measured through changes in capacitance [144] and by using strain gauges based on piezoresistivity [201].

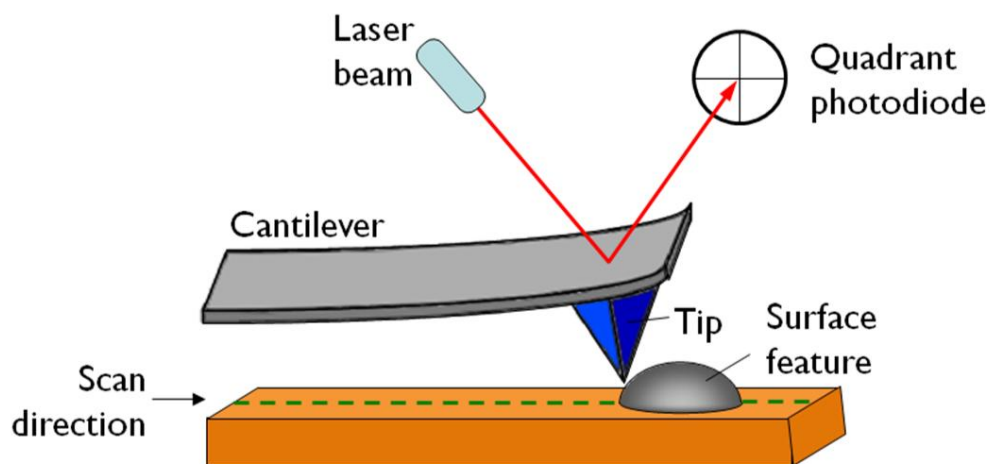


Figure 2-1 Typical set-up of an AFM. Laser light is reflected off the back of the tip onto the photodiode. The deflection of the cantilever is measured as it tracks the surface, allowing an image of the sample topography to be obtained. The tip-sample separation is controlled by feedback electronics.

However, these detection techniques have largely been superseded by the optical lever method [133]. The movements of the tip are measured using a quadrant photodiode. In the first instance, a laser is focused on the cantilever and the reflected beam hits the photodiode in its centre. Any deflection of the cantilever causes an imbalance in the amount of laser light hitting each quadrant in the photodiode and will thus provide a measurement of the magnitude of the force.

As the intermolecular forces that cause the cantilever deflection, and ultimately give rise to imaging contrast, are very short ranged, the cantilever must be positioned vertically in very close proximity to the sample surface. In addition to this, very precise movements in the X and Y directions are necessary during scanning. Most AFM systems achieve such accuracy by using tube piezoceramics, which allow the positioning of the probe with high resolution. They utilise the piezoelectric effect, whereby materials expand or contract when a voltage is applied to them. Four electrodes cover the outer surface of the tube, while a single electrode occupies the inner surface. Applying a voltage to one of the four outer quadrants causes the quadrant to expand and the scanner to tilt away from it (XY movement). Applying a voltage to the inner electrode causes the entire tube to expand or contract (Z movement). The cantilever is

moved with respect to the surface in a raster pattern, measuring the tip-sample interaction as it goes, and building up an image of surface topography line by line.

2.1.3 AFM Forces

The force the tip experiences varies due to surface changes as the scanning of the sample is performed. It is these changes in force that are transformed into an image of surface topography. The forces manifest themselves as the deflection of the cantilever, the extent of which is dependent on a number of different interatomic forces. The relationship between deflection and force can be summarised by Hooke's law:

$$F = -kx \quad 2-1$$

where F is the force, k is the spring constant, and x is the deflection. The origin of the force arises from the interaction of the tip with the surface, and can be modeled by the "Lenard-Jones" or 6-12 potential:

$$U(r) = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right] \quad 2-2$$

This relationship models the interaction between two atoms or molecules, in this case one at the tip apex, and one at the sample surface. As their separation (r) changes, so does the potential energy $U(r)$ between them. The relationship consists of both attractive and repulsive components (see Figure 2-2 for graphical representation).

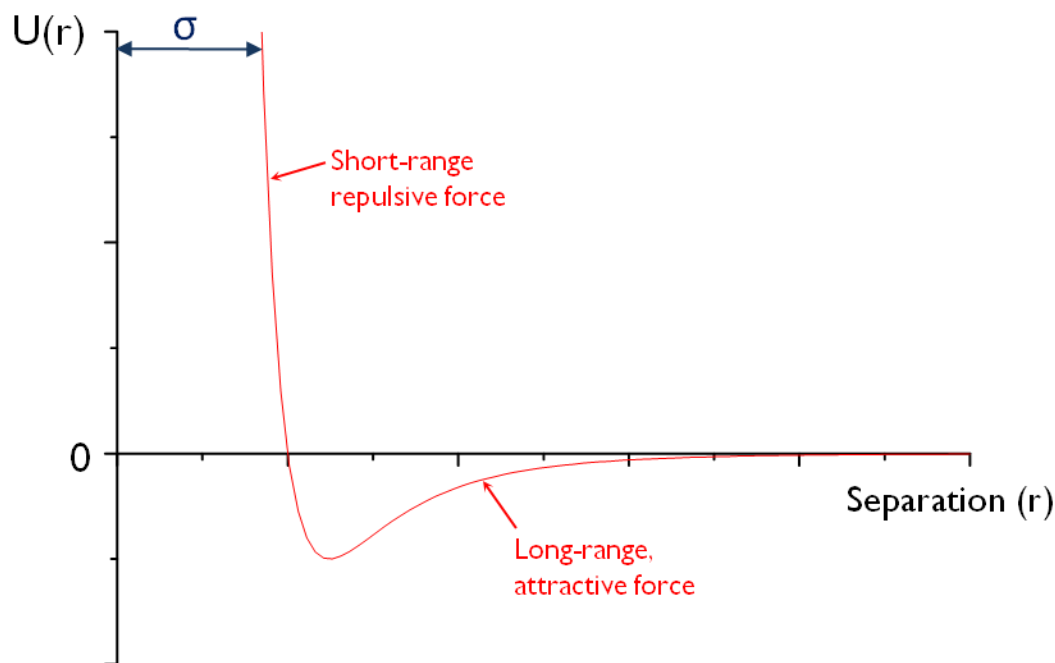


Figure 2-2. A graph of potential ($U(r)$) versus distance for the 6-12 Lennard-Jones potential. The potential is dominated by both attractive and repulsive components depending on the separation (r). σ is the hard sphere radius and describes a distance of closest approach.

The $1/r^{12}$ term accounts for the steep increase in potential at small separations, when the molecules strongly repel each other due to the Pauli exclusion principle. This states that two fermions, such as electrons cannot have the same quantum numbers. Two approaching particles will begin to experience a repulsive force as their separation is decreased. Eventually a point known as the hard sphere limit is reached, which describes a distance of closest approach. Any further approach would lead to the overlapping of electron orbitals, and the sharing of a quantum state, and so this repulsive force acts to separate the electron clouds. This force is very strong but short-ranged.

The $1/r^6$ term accounts for the more gradual decrease (attractions) at larger separations, and is down to attractive van der Waals forces. The geometry of the AFM tip on a surface resembles a small sphere above a plane [158]. Assuming this geometry, the attractive van der Waals force is given by,

$$F_{vdw}(d) = -\frac{HR}{6d^2}$$

where R is the tip radius, H is the Hamaker constant and d is the tip-surface separation. This force arises from the polarisation of the electron cloud surrounding the atomic nucleus. Fluctuations in instantaneous dipoles can become correlated so as to provide an attraction between two neighbouring bodies. These forces are relatively weak but long-ranged.

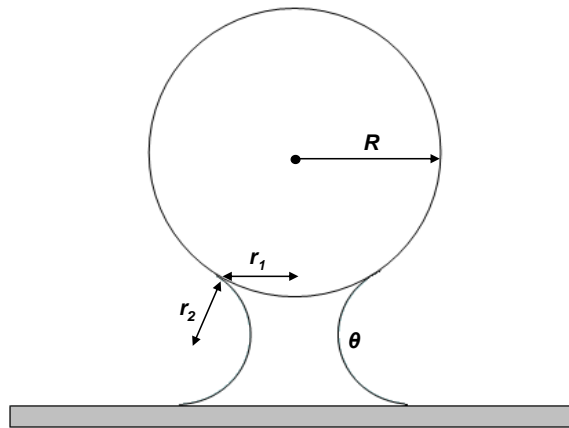


Figure 2-3. Capillary neck formation between a sphere (tip) and surface in the presence of a bulk water layer. R is the radius of the sphere, r_1 and r_2 are the principle radii of the curved surface, and θ is the liquid-vapour contact angle.

Adhesion is the force needed to separate the tip from the sample. When imaging in ambient conditions the adhesive interaction is often dominated by attractive capillary forces [23, 188]. In air it is possible for water to condense onto the surface and even form a bulk layer. A point contact with a small radius of curvature (such as an AFM tip), and resting on a surface, provides an ideal nucleation site for water present in the air. When the tip dips into this water layer a capillary neck forms, which provides an additional adhesive force in the range of nN [42]. The origin of this force is the formation of a meniscus between the AFM tip and the sample, caused by the condensation of water vapour onto a hydrophilic surface. Capillary condensation is driven by a pressure difference across the interface of a curved surface (see Figure 2-3 for a geometric

interpretation and definition of symbols). This pressure can be given by the Young-Laplace equation,

$$p = \gamma \left(\frac{1}{r_1} + \frac{1}{r_2} \right) \quad 2-4$$

where γ is the surface tension and r_1 and r_2 are the principle radii of the curved surface. The Laplace pressure acts on an area between the two surfaces, pulling them into contact. The capillary force works to increase the tip's adhesion to the surface, and this attractive contribution can be calculated as,

$$F = -p\pi x^2 = \gamma \left(\frac{1}{r_1} + \frac{1}{r_2} \right) \pi x^2 \quad 2-5$$

The capillary force can be eliminated by operating the instrument in liquid. If the tip is modelled as a sphere of radius R and the sample is assumed to be a flat plane, the capillary force is commonly approximated to be equal to [97]:

$$F = -4\pi R\gamma \cos\theta \quad 2-6$$

Pakarinen *et al.* estimated that the lower limit on sphere radius, where the above approximation was still valid, as being 1 μm [149]. Typical AFM tips have a radius of curvature of 10 nm, and show considerable deviation from the standard approximation. A number of attempts have been made at improving the description of the capillary force at the nanoscale [126, 149, 192]. These studies have centered on quantifying the influence of various parameters, such as tip geometry, tip-sample separation, contact angle, and surface tension, with the aim of obtaining a numerical solution for the capillary force.

2.1.4 Imaging Modes

The original method of AFM operation is known as contact mode; in which the tip remains in close-contact with the sample during its scan. There are different ways of obtaining image contrast. In constant-force mode (Figure 2-4a), the deflection of the tip is kept constant by use of feedback electronics that can alter

the tip height position, to keep the overall tip-sample distance constant. The deflection of the cantilever remains constant and the vertical position of the tip produces a good representation of the surface topography. When the AFM is operated in constant-height mode the feedback is turned off (Figure 2-4b), and the tip moves in a constant horizontal plane throughout. The deflection is used to monitor the sample surface and this method of operation allows for greater vertical sensitivity [103].

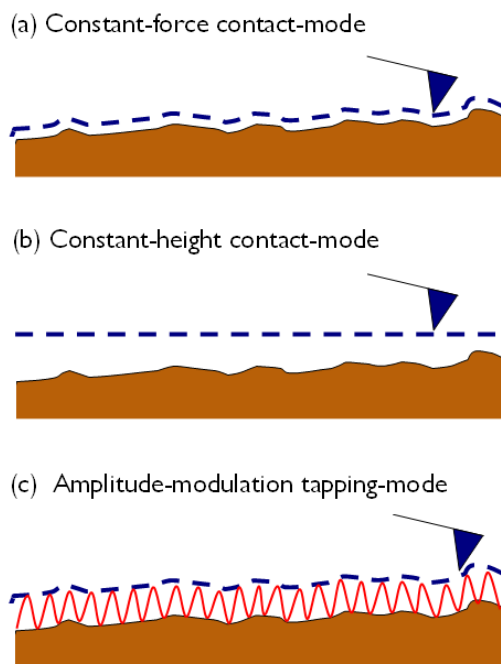


Figure 2-4 . Main modes of AFM operation. (a) In constant-force mode the tip deflection is kept constant and the tip is made to track the sample surface. (b) During constant-height mode the tip moves in a constant vertical plane. Both forms of contact mode exert large shear forces on the sample surface. (c) In tapping-mode, the tip oscillates at a constant amplitude above the surface. The shearing forces are greatly reduced, allowing soft samples to be imaged.

Early AFM studies of DNA utilised contact mode [85, 118], however it took some time before routine imaging of the molecule could be performed in ambient conditions [28]. Problems were caused by the type of cantilever available at the time and the substrates used, but mainly from the mode of imaging. The tracking of the tip across the surface in contact mode leads to the exertion of large shear forces, which can displace weakly bound particles or damage soft samples such as DNA. This problem is made worse by performing imaging in

air, where the capillary neck provides a large contribution to the shearing force. This can be reduced by using sharper tips to minimise the dimensions of air-liquid phase [28]. It had been shown that it is possible to image DNA repeatedly using contact-mode, as long as the humidity is kept below a critical value [199], such that sucking of the AFM tip down onto the surface by capillary action is avoided. Reliable imaging of DNA was found not to be possible at a humidity above 40 % RH [199, 205], which corresponds with the humidity where water condenses to form bulk monolayer films on the surface [93].

A huge advance came with the development of AFMs that operated with intermittent-contact between the tip and sample, otherwise known as tapping-mode AFM [220] (Figure 2-4c). In this mode, the cantilever is made to oscillate near its resonant frequency, striking the surface at the end of each oscillation. The energy associated with the oscillation must be such that adhesion of the surface is able to be overcome. For this reason, relatively stiff cantilevers, having a relatively high spring constant and a high quality factor Q are used to ensure that sufficient energy is stored in the cantilever to overcome adhesion. Intermittent contact with the surface at the end of each oscillation leads to the dissipation of energy from the system and a decrease in oscillation amplitude. The amplitude can be used as a set-point parameter, so that feedback electronics can act to adjust the tip height so as to maintain the set-point amplitude at its pre-set value. The amount the amplitude must be modulated to be kept constant can be used to gather topographic information of the surface, thus tapping mode AFM is known more generally as amplitude modulation AFM (AM AFM).

The contact force the tip exerts at the end of each cycle is considerably smaller than in contact mode [220]. Since the tip is no longer being dragged across the surface, shear forces are almost completely eliminated. The small interaction forces exerted mean that tip-induced sample deformation is minimised and soft samples can be imaged repeatedly. The use of tapping mode for DNA imaging has almost entirely replaced contact mode.

AM AFM is an example of a dynamic atomic force microscopy method; because dynamic properties, in this case amplitude, of a vibrating tip interacting with a surface is used to obtain a view of the sample surface [73]. Another dynamic method of operation is frequency modulation AFM (FM AFM), and is usually used under ultra high vacuum (UHV) conditions with the aim of obtaining atomic resolution on solid state systems [4, 76]. The technique has also been adapted for use in a liquid environment [72]. Frequency modulation works by detecting small changes in the cantilever resonance frequency, arising from tip-sample interactions, and using them as a feedback parameter.

AM is typically preferred in air, as FM requires two feedback loops which makes its electronics complex and slow for operation in ambient conditions [73]. AM AFM utilises relatively stiff cantilevers (10 to 50 N/m) in air to enable the tip to “punch through” any water layer on the surface without irreversibly sticking due to capillary neck adhesion. Almost all biomolecular imaging in air has been performed under ambient conditions of laboratory humidity, usually between 30 and 40 % relative humidity (RH). Following the introduction of intermittent contact modes AFM imaging of DNA in contact mode was largely abandoned.

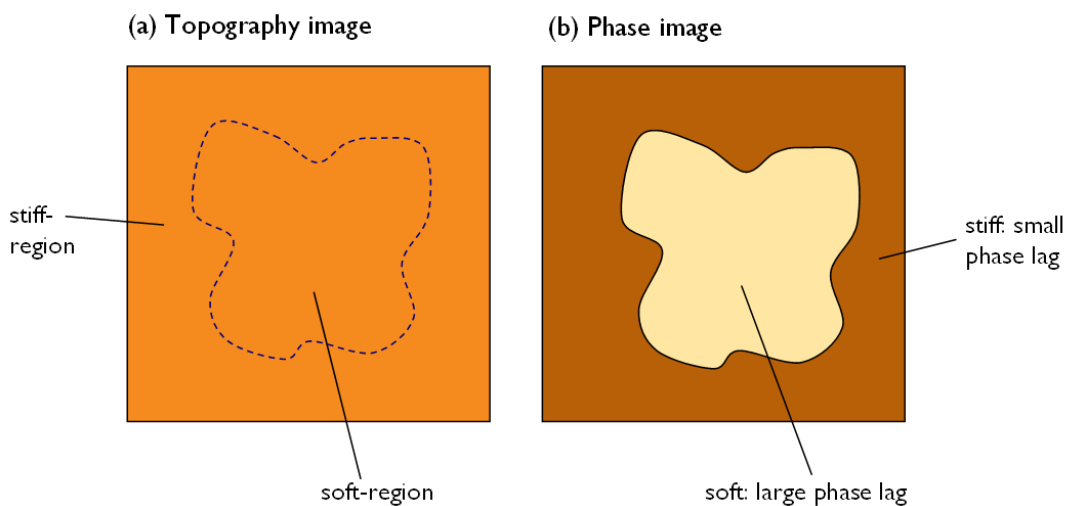


Figure 2-5. Demonstration of AFM phase imaging. (a) Example of the imaging of two distinct areas of different material properties. If the areas are of a similar height, then they become indistinguishable in a topography image. (b) Representative phase image. The different mechanical properties of the distinct regions lead to phase differences between the vibrating cantilever and the driving piezo motor. This will show up clearly as different contrasts in a phase image.

An interesting dynamic technique is to measure the phase difference between the driving signal of the cantilever piezo and the resulting motion of the tip, as measured by the detection system [178]. The phase lag is closely related to surface properties, allowing investigation of material properties such as friction and adhesion to be undertaken (see Figure 2-5). For example, hard surfaces tend to lead to a low phase shift, whereas soft areas tend to damp the tip oscillation, causing a greater lag in phase. The phase signal can be recorded simultaneously with the topographic image, allowing features on either image to be correlated with each other. Phase imaging of DNA can allow for faster scan rates, lower forces and improved resolution of mobile DNA molecules, in addition to the visualisation of features previously hidden in conventional height imaging [8, 105].

2.2 Imaging DNA with AFM

2.2.1 Imaging substrates and practicalities

By its very nature, AFM is a surface scanning technique and as such, any molecule that is to be investigated must be deposited on to a suitable substrate prior to imaging. There are therefore inevitably a number of criteria that the support must satisfy for it to be usable. The substrate in question must have a low surface roughness, such that molecules of interest can easily be distinguished from any spurious features of the support surface. It is crucial that there is a strong enough interaction between the surface and molecule, so that the sample will move from being in solution to being adsorbed onto the substrate. Strong binding to the substrate surface is also required so that the sample will resist the motion of the moving tip, without being damaged or detached from the surface.

Mica has been found to be a reliable substrate for imaging of DNA, and is now the most commonly used support surface [205]. Most biomolecules such as DNA are hydrophilic in nature and thus a hydrophilic material such as mica provides an ideal support. Mica is the collective name for a number of different silicate based minerals. Muscovite mica is used for AFM studies and has the structure

$\text{KAl}_2(\text{AlSi}_3\text{O}_{10})(\text{OH})_2$. The main characteristic of mica materials is the perfect cleavage of the basal layers. This is ideal for AFM studies where the upper layer can be removed by adhesive tape to provide a clean surface, free from impurities each time an experiment is to be performed.

The use of mica does however pose a problem to effective DNA binding as a direct consequence of its layered structure. The mica layers are 1 nm thick and are bonded together ionically with potassium ions ($\text{K}(\text{I})$). When freshly cleaved mica is immersed in solution the potassium ions dissociate from the lattice sites into solution, forming a counter-ion cloud above the surface [151]. This leaves the mica surface with a net negative charge, meaning that DNA, which also has a negative charge as a result of its phosphate backbone, will no longer experience an attraction to the surface, leading to an unacceptably low number of surface-bound DNA.

However, this problem can be negated by pre-treating the mica surface with a solution containing cations, or under liquid by introducing positively charged species into the aqueous medium [84, 197, 205]. These replace the dissociated $\text{K}(\text{I})$ ions, and leave the mica surface positively charged, such that DNA molecules experience an attractive force from the surface. The cations are able to form a diffuse atmosphere of counter-ions close to a surface, which act to partially neutralise the negatively charged phosphate backbone of the DNA and the mica surface. Correlations between the counter-ion clouds can lead to the situation where a net attraction exists between the DNA and mica, promoting the adsorption of the molecule [152, 153].

Pastre *et al.* modelled the DNA and mica surface as charged planes and hypothesised that if their respective counter-ions were to adopt a staggered configuration, they can be shared between the two surfaces [153] (Figure 2-6). On the mica surface lateral diffusion of DNA molecules is inhibited by large frictional forces, related to electrostatic interactions between the two surfaces. It is possible to consider the mica surface as a lattice of charges [147], and for DNA segments to move laterally, divalent counter-ions condensed on the DNA

must “jump” from one staggered position to the next [154]. For this to occur a large energy barrier must be overcome.

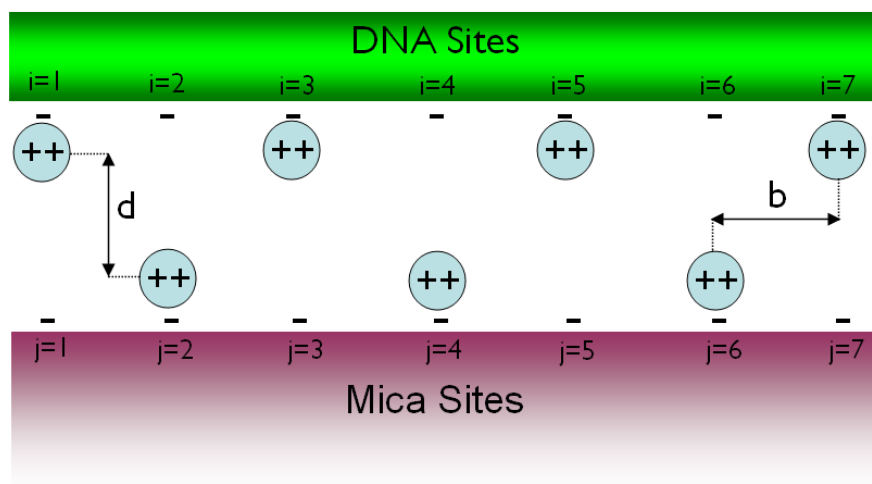


Figure 2-6. Schematic showing model for counter-ion correlations between DNA and mica. The labels i and j represent charged sites on the DNA and mica respectively. If the counter-ions adopt a staggered orientation then a situation would exist whereby there was a net attractive interaction between DNA and mica, and the molecule would adhere to the surface.

It has been shown that different cations allow for varying degrees of DNA surface binding, with transition metal ions being the most effective [84]. In particular, it appears that Ni(II), Co(II), and Zn(II) are the best candidates among the transition metal ions, and bind DNA tightly enough for imaging under liquid. Mg(II) was also discovered to adhere DNA to mica sufficiently for imaging in ambient conditions.

These three transition metal ions have ionic radii from 0.69 to 0.74 Å, so are small enough to fit into vacant binding sites above the hydroxyl groups in mica; vacancies which were previously occupied by K(I) ions. However, the Group 2 cation Mg(II), which has a similarly small ionic radius, does not bind DNA well enough for it to be imaged under liquid. This can be explained by considering the enthalpies of hydration of the different species: Ni(II), Co(II), and Zn(II) have anomalously high enthalpies of hydration in relation to their ionic radii. The enthalpy of hydration is related to the strength of bonds that the ion makes, and as such the transition metal ions are able to form strong complexes not only

with ligands such as DNA but also with mica itself [84]. Additionally, Mg(II) is a Group 2 metal with p electrons in its outer orbital, whereas Ni(II) is a transition metal with d electrons in the outer orbital, meaning that divalent transition metals can form a wider range of complexes in aqueous solutions. However, after drying it is possible to image DNA in buffers containing Mg(II) ions under ambient conditions.

It has also been shown that it is possible for Ni(II) to bind to the major or minor grooves of DNA [2, 189], making it very effective at balancing the surface charge of the molecule. Pastre *et al.* showed that pre-treatment of the mica surface with Ni(II) gives a stronger DNA binding [153]. These factors act to increase the correlation force and bring the two surfaces together. On the other hand Mg(II) ions tend to bind via non-specific electrostatic interactions with the phosphate groups in the DNA backbone [59], leading to weaker interactions.

2.2.2 Preparation of DNA samples for AFM

Preparing a DNA sample for AFM imaging typically follows a process containing three main steps (see Figure 2-7): first a small amount of DNA in buffer is deposited onto a mica disc, before being left for a period of time to bind. After incubation the sample is rinsed with a small amount of pure water, before finally being dried in a weak flux of argon or nitrogen gas. The deposition of DNA onto the substrate is a two-step process; beginning with the transport of molecules to the surface and ending with binding to the mica surface. An important question to ask is: how does the population of DNA seen on a flat surface by AFM compare with that of the solution? It may also be possible for DNA to be flushed away from the surface, or displaced by impurities.

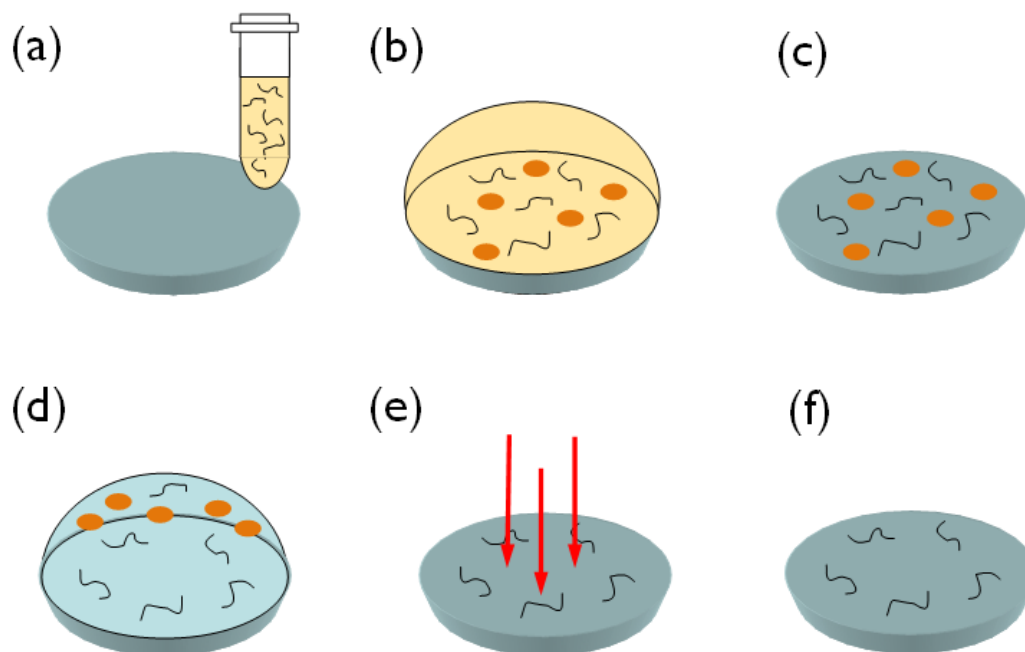


Figure 2-7 DNA sample preparation. (a) DNA is to be transferred from buffer onto a cleaved mica disc. (b) A small amount of DNA is deposited onto mica, and left for a period of time to bind. (c) After incubation the mica will contain bound DNA, but also other impurities present in the buffer (shown as brown spots). (d) Washing with water helps to remove partially bound DNA and impurities such as buffer salts. (e) The sample is then dried in a weak flux of gas to remove the liquid drop. (f) The mica disc then contains tightly bound DNA, and is ready for imaging.

Lang and Coates used electron microscopy to study the adsorption of DNA molecules from an ammonium acetate solution onto a protein monolayer [113]. By fitting their data to a theoretical model they were able to show that if the DNA was of a sufficiently low concentration, the transfer process was governed solely by diffusion and the fraction of molecules bound to the surface at any time t is given by,

$$\frac{n_F(t)}{n_0} = \sqrt{\frac{4D}{\pi}} \sqrt{t} \quad 2-7$$

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 $t = 0$, and D is the diffusion constant.

The deposition process was also studied using AFM by Rivetti *et al.*, in which small aliquots of DNA solution were transferred onto freshly cleaved mica, and left to incubate for times of up to 30 minutes [167]. They also observed that the surface density of molecules obeyed a $t^{1/2}$ power law, and concluded that the process is solely diffusion controlled, and that molecules are irreversibly bound to the surface.

The conformation that DNA adopts on the mica surface is a direct consequence of the various forces present during the binding process. Rivetti *et al.* were able to experimentally show that there existed two extremes of surface-bound DNA conformations, where the molecules can be said to be either surface equilibrated or kinetically trapped [167] (see Figure 2-8). During surface equilibration the molecules adhere but once touching the surface they are free to diffuse laterally and take up a minimum energy conformation (Figure 2-8a). These conformations are mediated by long-range, weak interactions, whereas short-ranged and stronger interactions will cause the molecule to be kinetically trapped. During kinetic-trapping, if any part of the chain touches the surface it is irreversibly pinned at that location on the surface and the rest of the molecule collapses onto that point (Figure 2-8b). It is no longer possible for the molecule to move sideways and as such, the conformation describes a 2D projection of the DNA's 3D structure.

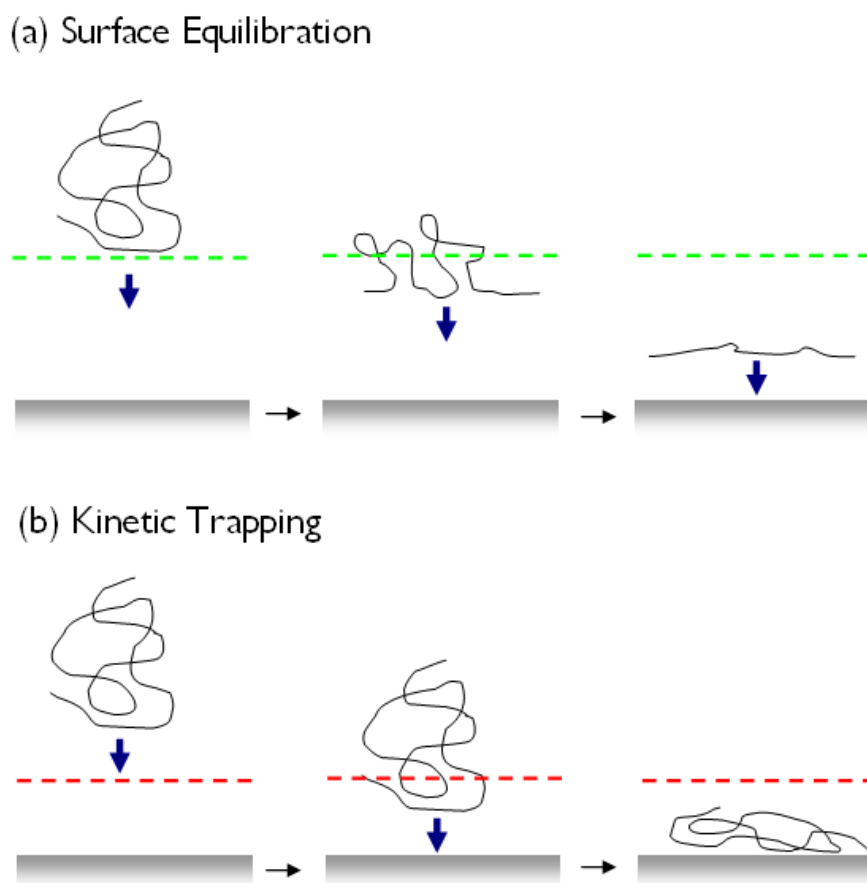


Figure 2-8. Schematic diagram of the adsorption mechanisms of DNA onto mica substrate from solution. (a) Surface equilibration is governed by long-ranged interactions (shown by green dashed line). As the molecule approaches the surface its constituent segments are able to adjust their positions, and the molecule takes up a flat minimum energy conformation. (b) Short-range forces (indicated in red) are responsible for kinetic trapping. The interaction is much stronger, with the molecule being pinned at one location and its remaining segments collapsing around it.

Surface equilibration occurs when DNA, containing Mg(II) in the buffer, is transferred onto mica, because Mg(II) is bound only weakly into the mica cavities. As a result, the DNA retains a degree of mobility, facilitating lateral diffusion [196]. Strong electrostatic interactions generally lead to molecules which are kinetically trapped, and this scenario is observed when the surface is pre-treated with a solution containing Ni(II) ions [153]. Analysis of the end-to-end distance of the molecules can elucidate which mechanism is responsible for binding. On average kinetically trapped molecules will have a shorter end-to-end distance than surface equilibrated DNA. Whether a DNA chain becomes kinetically trapped has been found to be dependent on the ionic species used to

promote binding, and also the relative concentration, but ultimately depends on the strength of the correlation force [152, 167]. For any given study, one mechanism may be more desirable than the other: equilibration provides a clearer view of a molecule, which can make examinations such as determining whether a protein is specifically bound easier to perform, whereas trapping provides a more complete observation of the solution conformations of the population.

2.2.3 Practicalities of studying transcription with AFM

AFM is a single-molecule technique and by virtue of this fact offers advantages over more traditional biochemical techniques for the study of DNA transcription. Earlier studies have typically been performed on large populations of molecules, whereby measurements of properties represent ensemble averages. This can lead to a situation by which events occurring at a single genetic locus are obscured. With AFM it is possible to observe each member of the population under study individually, and as such to obtain an overall distribution of results. AFM allows molecular structures to be observed directly, and can provide new information on the spatial arrangements of DNA, RNA and RNAP during the various stages of the transcription cycle.

Studies of transcription via AFM typically follow very similar sample preparations [101]. This usually involves the completion of the transcription reaction in a test-tube, in which the RNAP, specially designed DNA templates containing promoter sequences, and three of the four nucleotide triphosphates (NTPs), are mixed together. The RNAP associates with the promoter region, melting the target sequence and forming an open promoter complex (OPC). The presence of the NTPs allows transcription to be performed *in vitro*, and the OPC will go on to form an elongation complex. A chain of nascent RNA will begin to be constructed out of NTPs. This period of elongation continues until the RNAP encounters a base on the template DNA which codes for the NTP that is missing from the solution. This point is known as the stall site, and the structure

consisting of the protein, small chain of RNA and template DNA is known as a stalled elongation complex (SEC).

Static studies of properties such as general complex structure can be performed easily in ambient conditions, in a way that is identical to conventional studies of DNA alone. Divalent cations, which help to facilitate DNA binding, are added to the buffer containing the DNA·RNAP complex, before a small amount is transferred onto mica and allowed to incubate for a period of time. This method provides a “snap-shot” of the complex and allows the further exploration of OPC and SEC structure [160, 166, 168].

Dynamic studies of transcriptional elongation can also be performed using AFM but these measurements require the AFM to be operated under liquid. Imaging the process of transcription *in situ* directly presents a problem that must be overcome. Under liquid, molecules are less well bound to the substrate. Lateral frictional forces between the molecule and surface are reduced, leading to a situation where molecules can diffuse laterally across the surface and even detach entirely. In addition to this, the tracking of the tip over the surface takes a relatively long time. If these diffusional motions are faster than the time resolution of the instrument, then the tip becomes unable to resolve the molecule properly and the image obtained will be blurred. However, a strong anchoring of the molecule to the substrate is unfavourable for the study of dynamical processes such as transcription. For the RNAP to properly transcribe the template, the DNA must be allowed some freedom to move and rotate.

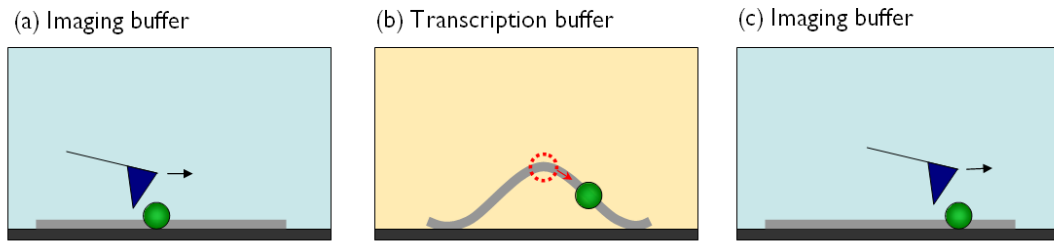


Figure 2-9. Method for studying transcription dynamically in liquid with AFM. This is achieved by changing the buffer in the AFM fluid cell. The exchange of solution allows imaging and transcription to be performed intermittently. DNA is shown as a grey line, and RNAP as a green sphere. (a) Use of an imaging buffer allows effective binding of DNA to the substrate, and leads to reliable imaging. (b) Substitution with a transcription buffer enables the DNA to detach from the surface and allows for RNAP translocation. (c) Subsequent imaging in the original buffer shows that the RNAP has moved along the chain. This method produces time-lapse images of transcription elongation.

This problem can be circumvented by the exchanging of buffers in the fluid cell of the AFM that either promote or inhibit DNA adhesion to the substrate. (see Figure 2-9). Using this method it is possible to achieve reversible binding of DNA to the mica surface, essentially by controlling the concentration of divalent cations in the environment around the DNA [197]. The apparatus used consists of containers suspended above the AFM, containing regulators to adjust the flow rate. Buffer solution is constantly being driven into the system by the influence of gravity, and the regulator is used to switch from one fluid to another. In the study of transcription, two solutions are used: an “imaging” buffer and a “transcription” buffer. The former allows enough adhesion for complexes to be imaged properly, whereas the latter promotes detachment, allowing the RNAP to move with respect to the DNA and transcribe it.

2.3 Previous AFM studies of transcription

2.3.1 Initiation studies

For transcription to initiate, the polymerase must associate with the small region of DNA known as the promoter. Guthold *et al.* have been able to observe the one dimensional diffusion of *E. coli* RNAPs along a promoterless DNA fragment with AFM [82]. Sequential images of these complexes were taken, in which the DNA was seen to slide back and forth beneath the enzyme. The mean diffusion distance, obtained from measurements on a large number of molecules, was found to be approximately proportional to the square root of time. This time-dependence is typical of diffusion-controlled processes. In conjunction with sliding, AFM studies of non-specific complexes have also observed intersegment transfer and hopping [26]. These events were preceded and followed by sliding, so as such, it is possible that a combination of sliding, intersegment transfer, and hopping are responsible for the most efficient promoter search.

However, as a cautionary note, the lifetime of the non-specific complexes studied in the AFM is much longer than bulk studies performed in solution [88, 187]. The fact that the DNA-protein complex is surface-bound in the AFM would account for this discrepancy. The mica surface imposes a conformational constraint on the complex, thus hampering its manoeuvrability, slowing the reaction and preventing the enzyme from dissociating.

The *E. coli* RNAP· σ^{70} holoenzyme provides a relatively simple system with which to study the structure of OPCs with AFM. Observations of RNAP· σ^{70} OPCs with AFM have shown that the DNA template always appears bent in the region bound to the polymerase [160, 168]. No significant curvature is observed in DNA alone and it can be concluded that these bends are caused by the conformational changes associated with the open complex formation. Rees *et al.* examined complexes formed at the λ_{PL} promoter and obtained a mean bend angle of 54° [160], while Rivetti *et al.* looked at the λ_{PR} promoter and observed a

wide distribution of bend angles, with the DNA being bent between 55° and 88° [168].

Rivetti *et al.* also performed a detailed analysis of the contour lengths of many DNA molecules and found that in the case where RNAPs were bound to the promoter in an open conformation, the length of the fragment was reduced by ~ 30 nm (~ 90 bp) relative to free DNA. This, together with the bend angle measurements, enabled a model of the open complex to be put forward. The reduction of contour length is consistent with a model in which the promoter DNA is wrapped around the polymerase (Figure 2-10).

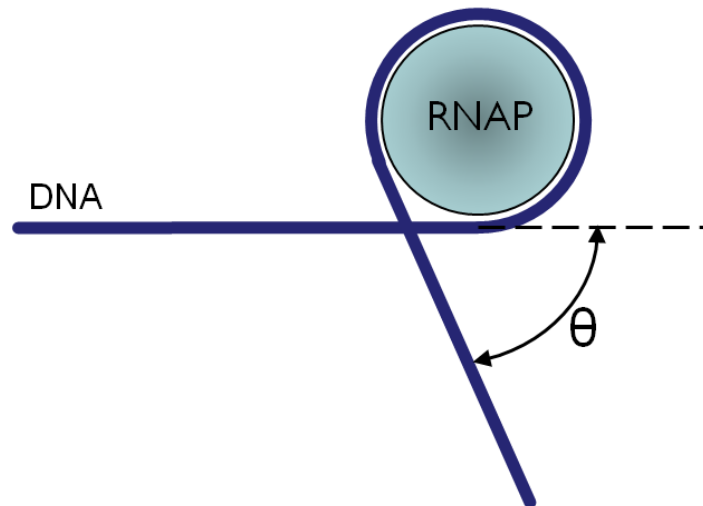


Figure 2-10. Model demonstrating how DNA can wrap around an RNAP core, and how the characteristic bend angle is defined.

AFM studies have also been completed in which transcription-factor σ^{54} was used instead of σ^{70} during transcription initiation [164, 179]. This method is slightly more complicated, as the $\text{RNAP}\cdot\sigma^{54}$ requires additional activators (e.g. nitrogen regulatory protein C (NtrC)) to unwind the DNA to initiate transcription. Upon ATP hydrolysis, the activator protein catalyses the transition of the holoenzyme from a closed complex to an open complex. Both closed and open complexes were imaged, along with an intermediate stage. The intermediate complex consisted of contacts between $\text{RNAP}\cdot\sigma^{54}$ at the promoter and the NtrC, with the intervening DNA forming a loop. Looping provides a way for the distal activator protein to access the promoter region containing its

associated complex, with the resulting interaction allowing transcription activation to take place. Both classes of complexes caused the bending of DNA in the region around the promoter, with closed complexes exhibiting an apparent bend angle of $49 \pm 24^\circ$, and open complexes yielding an angle of $114 \pm 18^\circ$ [164]. These differences were interpreted by explaining that the transition from a closed complex to an open complex during transcription activation was accompanied by structural changes in the complex.

2.3.2 Elongation studies

Rees *et al.* performed studies on both OPCs and SECs, with measurements of the protein induced bends on the template providing a comparison of the two different structures [160]. The mean bend angle in elongating complexes was found to be 92° , compared to 54° in open complexes. Measurements of the DNA contour length of stalled elongation complexes were undertaken, where it was found that the template exhibited a ~ 22 nm reduction in length [166]. This reduction was lower than earlier measurements performed upon open complexes [168]. The reduced wrapping and changes in bend angle were attributed to the loss of contacts between the DNA and polymerase during the transition from initiation to elongation [168].

More dynamic aspects of transcription elongation can be followed by performing the imaging under aqueous fluid, in the manner described previously in this chapter (Section 2.2.3). Kasas *et al.* performed tapping-mode AFM on a DNA fragment containing the λ_{PR} promoter site [102]. The concentration of each NTP was kept deliberately low (0.5 - 5 μM), in order to circumvent one of the inherent drawbacks of AFM. Data acquisition in AFM, as with other scanning probe microscopy techniques, is relatively slow with an image typically tens of seconds to be obtained. By using a relatively low concentration of reactants, NTP binding becomes the rate-limiting step in elongation. The slower elongation rate meant that various intermediates of the process could be observed.

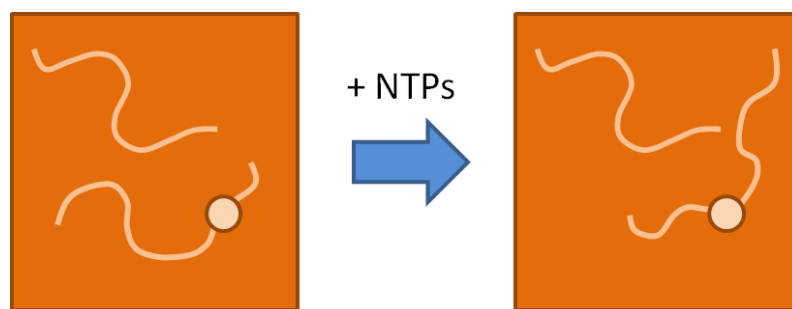


Figure 2-11 Schematic of how dynamic studies into transcription elongation appeared under AFM. One DNA strand is depicted with an RNAP bound, whilst the other strand can be used as a reference. After NTPs are added into the reaction mix the RNAP remained stationary, but was observed to thread the DNA through its core in a direction corresponding to active transcription elongation.

Kasas *et al.* observed RNAPs firmly bound to the mica surface with associated DNA that was free to diffuse laterally on the surface [102]. On addition of NTPs, DNA molecules were observed being pulled through the enzyme core (See Figure 2-11). The associated DNA arms were observed to concomitantly increase and decrease in size, with the template eventually being released from the polymerase. A sequence of images showing intermediates in transcription elongation was taken, and subsequent measurements of the variation of the relative position of the RNAP on the template with time, allowed the transcription rate to be estimated as being $0.5 - 2 \text{ nt s}^{-1}$. Similar measurements were later performed by Guthold *et al.* with a fragment containing the λ_{PR} promoter and t_{R2} terminator [82]. Before addition of NTPs the two arms of DNA either side of the polymerase were observed to diffuse laterally on the surface but never translocated through the enzyme. Upon nucleotide injection, RNAP began to thread the template in a unidirectional manner, consistent with the orientation of the promoter in the fragment. Measurements yielded a transcription rate of $1.5 \pm 0.8 \text{ nt s}^{-1}$.

In both these two dynamic studies of transcription it was not possible to observe the RNA transcripts. RNA is able to form base pairs with itself, allowing it to fold up into compact secondary structures. These should remain in contact with RNAP and perhaps become obscured by the broadening effect of the tip. However, it has been possible to confirm transcription activity with AFM, by use of rolling circle complexes [83, 102]. These are small, circular fragments of

single-stranded DNA, which mimic the transcription bubble formed during transcription initiation. An RNAP is able to transcribe a rolling circle for tens of minutes, or until NTPs are no longer available in solution. Kasas et al. deposited rolling circle complexes on mica and injected NTPs into the fluid cell continuously to initiate transcription [102]. The cell was then flushed out, after which the sample was dried. Long pieces of RNA, synthesised from rolling circle templates, were readily imaged by AFM in air.

It is much easier to observe the RNA chain when the sample is dried and imaged in air. Rivetti *et al.* studied transcription ternary complexes of *E. coli* RNAP and yeast RNAP III, using the method of nucleotide omission to form stalled elongation complexes [166]. They discovered that it was possible to observe extended RNA transcripts clearly, but only when the stall site was over 370 nucleotides away from the promoter site. In general, the RNA chain was located exiting from the RNAP, on the opposite side of the protein relative to the smaller angle subtended by the DNA arms. For the bacterial RNAP the distribution was centred about 140°. The authors believed that the location of the RNA exit site may help in keeping the transcript far away from the DNA arms, allowing the unhindered growth of the RNA chain [166].

2.3.3 Termination

Transcriptional elongation proceeds until the polymerase encounters a specific base sequence, known as a terminator. Limanskaya and Limanskii used AFM to visualise complexes of bacteriophage T7 RNAP with a DNA template during transcription [117]. The 1414 bp fragment used contained the T7 promoter and terminator asymmetrically located at the template ends. In this work, Limanskaya and Limanskii were able to control the reaction conditions to image complexes at the terminator site [117]. The transcription temperature was reduced from 37 °C to 31 °C, resulting in a decrease in the dissociation rate of the elongation complexes. Large globular features could be observed at the terminator site, with their size suggesting that a number of RNAPs had stopped

one after another on completion of elongation. Images were also obtained showing three RNAP molecules bound to the template: one at the promoter, another at the terminator site and a third located in between the two. This result shows that as soon as one RNAP begins elongation, another can bind to the promoter and initiate another round of transcription, and could contribute to the high rate of transcription observed *in vivo*.

2.4 Outlook and developments

2.4.1 Fast scan AFM

AFM has allowed both static and dynamic studies of DNA transcription to be undertaken. The different stages of the transcription process have been studied, both in air and liquid, illustrating the versatility of the technique. These studies have answered some of the questions surrounding transcription and can be used in conjunction with other single-molecule and biochemical techniques to obtain a much clearer picture of the mechanisms involved in transcription.

AFM is especially useful for studying the spatial arrangements of DNA and protein, such as RNAP during different stages of transcription, in order to reveal aspects of biological function. However, a number of technical limitations must be overcome for the technique to be reliably used to visualise dynamic DNA-protein interactions. To obtain any meaningful statistical conclusions about an event, a relatively large population must be studied. Data acquisition is relatively slow with AFM. The imaging time is especially important in dynamic studies, such as transcriptional elongation, where if the tracking is too slow, an event may be entirely missed. AFM scans over a relatively small area, providing a high resolution image. Scanning a small area does, however, limit the number of molecules visible in a single scan and as such, a large number of scans must be taken of different areas to derive a meaningful conclusion: a process which can be time-consuming.

Attempts have been made at improving the scan-speed, but these require radical re-design of the AFM instrumentation. For example, in the high speed

VideoAFM a micro-resonant scanner is used. [91, 95]. There is no electronic feedback loop, but by controlling the mechanics of the cantilever and applying a force directly to the tip, it is made to continuously track the sample surface. An image on a conventional AFM takes tens of seconds to obtain, whereas the VideoAFM can deliver videos at a rate of around 20 frames per second. However, this approach is so far limited to imaging in contact mode and therefore for DNA is restricted to ambient conditions utilising samples that are tightly bound to the surface.

Another strategy for high-speed AFM utilises small cantilevers [7]. Miniaturising AFM components means that their resonant frequencies increase, allowing them to be driven at higher speeds before mechanical resonances are excited. A number of studies utilising the small cantilever approach have centred on the dynamics of restriction nucleases and DNA complexes at a single-molecule level, during which it was possible to study the interplay between the two molecules, such as the one-dimensional diffusion of a restriction enzyme along DNA, and the eventual cleavage of the strand in real-time [51, 77, 216]. One such study by Crampton *et al.* studied the action of the type III restriction nuclease EcoP151 [51], which must interact with two recognition sites separated by 3500 bp before it is able to cleave the DNA. The fast-scanning method allowed the imaging of EcoP151 translocation in real time at a rate of 79 bp/s, and a scan rate of 1-3 frames per second. This is a huge breakthrough, as illustrated by comparing this translocation rate to the speed of transcription elongation observed by Guthold *et al.* using conventional tapping AFM (0.5-2 bp/s) [82]. They also observed the dynamic formation of DNA loops and transient DNA supercoiling. They concluded that the enzyme uses both translocation and diffusive looping to form contacts with its recognition sites.

Small cantilever AFMs using tapping mode have been used by Kobayashi *et al.* to visualise short DNA strands in motion in aqueous liquid [106]. Moreover, they demonstrated the ability of the technique to observe real-time images of biotinylated DNA binding to and dissociating from streptavidin protein. It has already been demonstrated that it is possible to form DNA-RNAP complexes in bulk fluid using conventional tapping mode AFM by injecting a solution of RNAP

holoenzyme into a fluid cell containing DNA [81]. As such, it should be possible to observe the real-time formation of OPCs using fast scan AFM techniques. Fast scan AFM still holds the promise of being able to study DNA transcription *in situ*, with a view to observing the entire process in its entirety from promoter recognition to termination.

Whilst it has already been shown that it is possible to study DNA-protein dynamics with fast-scan AFMs [51], the technique is in its infancy, and more work is needed to make the system reliable. Even at these enhanced scan rates, resolution is compromised due to molecular motion, and routine imaging of DNA is difficult to achieve because tuning of the DNA–mica interaction is difficult to achieve consistently. However, work on understanding how divalent cations effectively bind DNA to a mica surface through counter-ion correlations has allowed a greater degree of control over DNA conformations in buffer [152, 153].

2.4.2 Moving to more complex systems

To date, studies of DNA transcription with AFM have generally used fairly simple model systems, typically consisting of short DNA templates containing the required promoter regions, and RNAP holoenzymes. Transcription *in vivo* is much more complex; genomic DNA can be thousands of base pairs long, and a RNAP may require a number of different transcription factors to assist activation of transcription. For further insights into transcription to be gained through AFM techniques, it will become necessary to switch to systems of greater complexity, which better imitate the conditions found *in vivo*.

A major problem when imaging long strands of DNA with AFM is their tendency to cross-over numerous times on the surface during deposition. One method for aligning long DNA molecules on a surface uses the flow force generated by meniscus motion, otherwise known as molecular combing. Li *et al.* used gas flow to drive forward a drop of DNA solution on a bare mica surface [115]. By carefully controlling the direction of the flow, and the speed of the moving interface they were able to align long λ -DNA fragments (48,502 bp) on the

surface. Maaloum pursued a different strategy using lipid monolayers as a substrate for DNA [125]. Local demixing of the different lipid constituents caused λ -DNA to bind to the surface in a preferred orientation. The favourable orientation of the DNA relative to the fast scan direction allowed high resolution images, showing the right-handed DNA helix, to be collected under aqueous buffer. Molecular combing can, however, cause DNA over-stretching.

Under AFM, proteins appear as globular structures, and any molecules forming a complex with DNA can be easily identified from the template they are bound to. However, when studying systems of greater complexity, where two or more proteins could be interacting with DNA recognition sites close to each other, proteins of similar size become indistinguishable by their morphology alone. This poses a barrier to the study of more biologically relevant systems for transcription. Genes are often flanked by a number of different binding sites for different transcription factors, and efficient gene expression can require the cooperative behaviour of several different proteins. The inability of current AFM techniques to unequivocally identify different proteins by morphology alone makes it difficult to draw conclusions about how they interact with each other during transcription.

The work of Crampton *et al.* on convergent transcription represents one the first studies where more than one protein has been observed on a single DNA template [47]. To study systems of greater and greater complexity to further understand transcription *in vivo*, new methods complementary to AFM will be necessary, for example, to identify different protein components within a larger macromolecular complex. Combining AFM with fluorescence microscopy is an attractive solution, since specific molecules can be identified by attaching fluorescent dyes or objects. However, the spatial resolution of the technique is limited by the wavelength of light used during excitation. By combining both techniques together it becomes possible to overcome their own individual limitations. Ebenstein *et al.* used this strategy to examine the interaction of two distinct RNAPs (*E. Coli* RNAP and T7 RNAP) with T7 genomic DNA [60]. The DNA possessed promoter sites for both RNAPs. Three *E. Coli* RNAPs were observed with AFM in close proximity to each other. The *E. Coli* RNAPs were

labeled with a fluorescent quantum dot, and whilst individual molecules could not be distinguished, the fluorescence image identified the particular type of protein observed under AFM. The T7 RNAPs were labeled with four different colored quantum dots, which could be seen dotted along the DNA backbone with fluorescence imaging, demonstrating the ability to detect multiple interacting species.

Sanchez *et al.* used fluorescent polystyrene nano-spheres to align the optical image with the AFM topographic image, allowing both signals to be overlaid for easy analysis [174]. They studied the interaction of the human RAD51 protein labeled with a single fluorophore with λ -DNA. Under fluorescence imaging they observed green filamentous structures, which were interpreted as DNA covered with RAD51. In this case they were able to detect the presence of the protein by its fluorescence, and further define the DNA structure from the AFM topography images. Although single-fluorophore sensitivity has not yet been demonstrated by this technique, it is nonetheless an intriguing development in imaging, combining the single-molecular resolution of AFM with molecular recognition capabilities from optical signals.

Over the last couple of decades, AFM has been used to study every major step in transcription, from the initial formation of DNA-RNAP complexes to dynamic studies of transcription elongation, and eventual termination of complexes. The microscopes ability to provide direct structural information has proved extremely useful in ascertaining the spatial arrangements of DNA-RNAP complexes actively involved in transcription. The potential future developments in AFM imaging described here demonstrate that the technique still has much to offer, and will allow the possibility to move away from predominantly simple studies on static systems to dynamic systems containing a number of interacting proteins, more akin to the situations found *in vivo*.

Chapter 3

3 AFM imaging of DNA on mica

3.1 Introduction

AFM is a single-molecule technique and by virtue of this fact offers advantages over more traditional ensemble biochemical methodologies for the study of DNA transcription. With AFM it is possible to observe each member of the population under study individually, and thereby obtain an overall distribution of results, allowing any heterogeneities or different intermediates to be spotted. AFM allows molecular structures to be observed directly, and can provide new information on the spatial arrangements of DNA, RNA and protein during the various stages of the transcription cycle. As AFM detects every feature on the surface, having a similar sensitivity to all objects as it detects forces, it is crucial to have high-quality DNA samples that are free from contamination. It is important that the surface-bound complexes imaged under the AFM tip represent the *in vivo* structures of the molecules as accurately as possible. Clean images are important for studies of DNA;RNAP complexes as any extraneous factors, such as protein or buffer contamination, could lead to misinterpretation of data.

This chapter deals with the production and purification of DNA samples for imaging under AFM. A range of different fragments were used: both linear and circular, which were produced in a number of ways. This chapter begins with a description of some of the molecular biology techniques used, and the steps involved in sample production. Both restriction enzyme digestion of DNA plasmids, and amplification by the polymerase chain reaction (PCR) were used to generate the necessary linear fragments.

The purified DNA must then be deposited onto a mica substrate prior to imaging. Samples were prepared using the sessile drop method, in which a small drop of solution is deposited onto the substrate and allowed to incubate

for a period of time, before it is then washed and dried. The different stages of this process were investigated separately, with the aim of studying how they could potentially influence the surface population.

To achieve efficient binding of the molecule onto a substrate there must be an attractive interaction between the mica surface and DNA. We look at different ways of achieving this by introducing divalent cations into the system, and how the method used affects the conformation of the molecules seen on the surface. With these studies it becomes possible to understand how a population of surface-bound transcription complexes might represent those in solution.

August Johansson is acknowledged with collecting the data in Section 3.5.2.

3.2 Preparation of DNA samples

3.2.1 The Polymerase Chain Reaction (PCR)

A number of linear DNA templates used in this thesis were generated by PCR of linear DNA. PCR was discovered by Kary Mullis in the 1980s, and for this concept he was awarded the Nobel Prize. PCR is an incredibly powerful molecular biology technique which allows the generation of many copies of a single DNA template, using a relatively simple procedure whereby certain processes involved with DNA replication are simulated in a test tube using reagents similar to those found *in vivo*. Basically, the DNA double helix is denatured, leaving two individual strands that can act as templates for the manufacture of daughter strands. The starting material for the process is double-stranded template DNA that contains a certain sequence to be amplified. The region to be amplified is specified by the use of specially designed oligonucleotide primers. These provide a starting point for DNA polymerase to catalyse chain elongation, as in DNA replication *in vivo*. The reaction mix contains a small amount of template DNA, DNA polymerase, and the four deoxynucleotide triphosphate (dNTP) monomers. Each cycle in the PCR is essentially a three step process (see Figure 3-1 for a schematic diagram):

1. Denaturation
2. Annealing of primers
3. DNA synthesis

The denaturation step involves the heating of double-stranded template DNA, usually at 95°C for five minutes. The elevated temperatures break the hydrogen bonds making up the double helix, and the DNA is said to melt. The temperatures used are not sufficient to break the much stronger covalent phosphodiester bonds of the backbone and so the molecule exists as two single-stranded chains, which can act as templates.

Annealing of the sequence specific primers begins with the rapid cooling of the solution to a temperature usually 5-10 °C lower than the melting temperature T_m (defined as the temperature at which half of the DNA strands exist as double-helices, and the other half are still single-stranded) of the primer template complex. DNA polymerase is unable to begin a new chain alone (it can only add a nucleotide onto a pre-existing 3'-OH group), but needs short lengths of nucleic acid primers to begin the reaction. The temperatures used allowed primers to readily anneal to their complementary sequences on the single stranded DNA. The starting point for DNA synthesis can be specified by designing the primer so that it anneals to the template at a particular sequence along the chain.

The third stage involves heating the reaction to a higher temperature, typically 72 °C (the optimum working temperature of the heat stable DNA polymerase). PCR uses heat stable *Taq* DNA polymerases that add deoxynucleotides to the 3' end of the primers, using the other strand as a template, to produce new double duplex daughter DNA strands.

The original and newly synthesised strands then go on to become templates for another cycle of heating, annealing and synthesis. In the first cycle of PCR, each DNA template gives rise to two new daughter strands, doubling the number of strands in solution. Repeating the cycle again doubles the number of strands once more, and many cycles lead to the exponential growth and massive

amplification of the number of molecules. The original template becomes only a tiny fraction of the total population

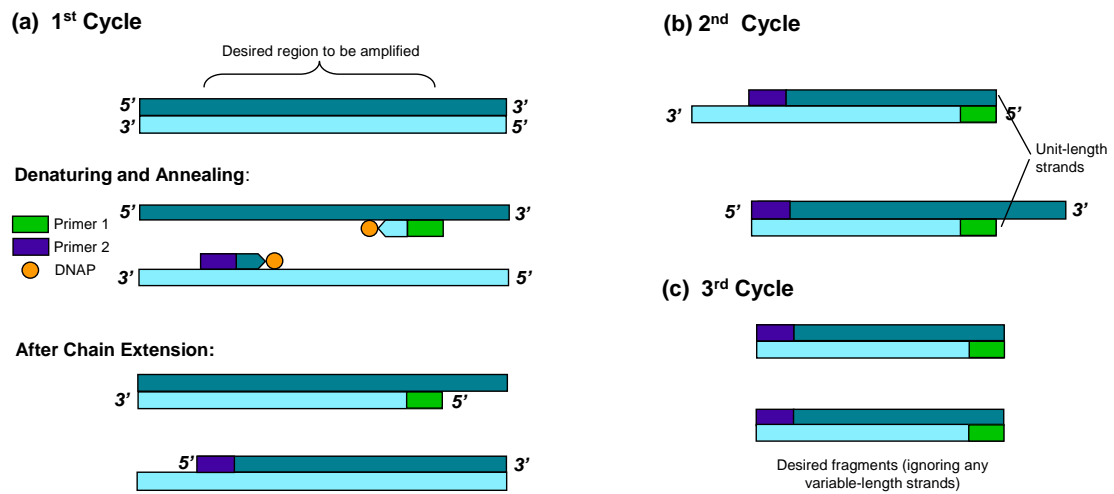


Figure 3-1. Schematic diagram of the PCR process. (a) The complementary DNA strands are represented as different shades of blue. The double helix is split by heating to produce two complementary strands. Primers are able to anneal to the exposed strands at specific target locations. Polymerases then move along the chain, building the chain as they go. (b) The process begins again, but this time there are four binding sites available to the primers. For simplicity, only the daughter strands that contain one strand of the same unit-length as the desired region are shown. (c) These strands subsequently act as templates for the third cycle. The result is two identical copies of the desired fragment. Subsequent cycles then produce many more of these fragments.

Important to PCR was the replacement of normal DNA polymerases with thermally stable *Taq* DNA polymerases, originally discovered in hot springs. These have the important characteristic that they are not inactivated at high temperatures, and increase the specificity of the reaction [38, 61, 171]. Originally PCR utilised *E. coli* DNA polymerases that become denatured at the temperatures needed to separate the DNA helix, meaning that fresh enzyme had to be added at the beginning of each cycle; a time-consuming process. *Taq* polymerase has an optimum working temperature of 72 °C, but remains stable up to 95 °C. *Taq* polymerase needs only be added at the start of the reaction and remains active throughout amplification, and allows the entire process to be automated by the use of thermal cyclers.

The initial studies on sample preparation described in this chapter required a number of DNA test molecules. These samples were obtained by amplifying different portions of a GAPDH cDNA template. The original template is 1310 bp long, but by using specifically designed forward and reverse primers it is possible to pick out specific regions of the original template (see Figure 3-2 for schematic). Primers must be carefully designed so that annealing to multiple sites on the template is avoided. Using these methods fragments of sizes 200, 400, 600 and 800 bp were produced. Practically the PCR reaction was performed by making up a 50 μ l solution containing the heat resistant *Taq* polymerase, template DNA, dNTP mix, and MgCl₂. Reactions were performed using the GoTaq Hot Start Polymerase (Promega) kit as per the manufacturers instructions. This involved the reaction mix being placed in a thermal cycler and 25 PCR cycles were performed. In an ideal (100 % efficient case) this would lead to a 2²⁵ fold increase in DNA number density.

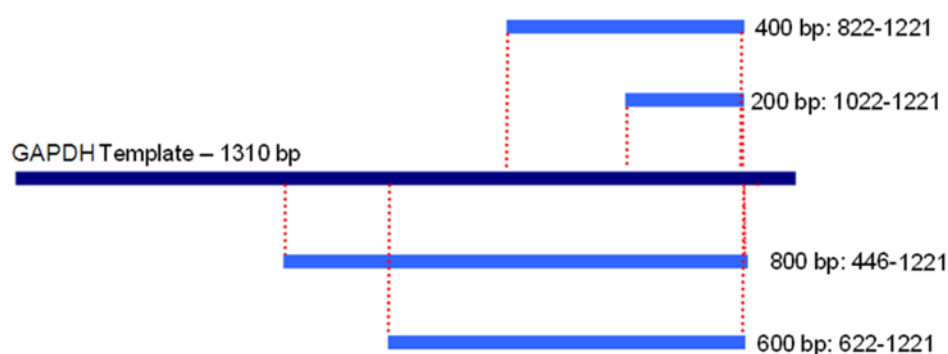


Figure 3-2. Diagram showing origins of PCR generated linear fragments. The original GAPDH template is shown in dark blue, and the fragments produced in light blue along with the specific locations of the product regions.

3.2.2 Gel Electrophoresis

Gel electrophoresis is an important analytical technique for separating DNA fragments according to size, and can provide both qualitative and quantitative information. For example, after PCR has been performed, it is important to detect whether the process has worked correctly and to analyse the amount of product. Gel electrophoresis separates molecules by size; the gel acts as a mesh

and passes small molecules quicker than larger fragments. DNA is negatively charged and naturally takes up a rod-like shape when placed in an electric field. By virtue of its charge, DNA molecules will move through the gel when a current is applied, at a speed governed by the molecules size. The gel is placed in a buffer to resist changes in pH, and to provide a suitable ionic field so as to allow a current to pass through the gel.

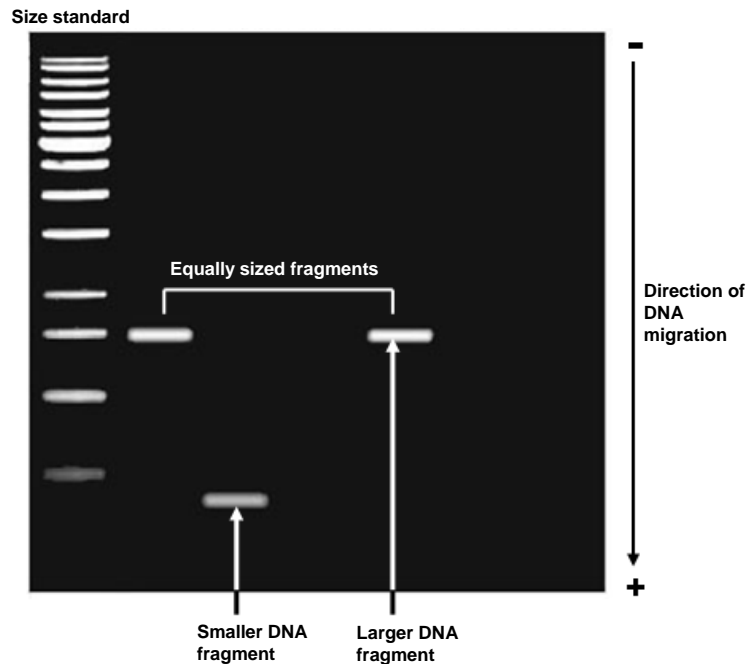


Figure 3-3. Example of a gel electrophoresis experiment, with the main features indicated. DNA fragments show up as fluorescent bands under UV light. These can be referenced against a size standard (e.g. 1 kbp ladder), allowing an estimation of the size of the DNA to be made.

The different sized fragments separate into discrete 'bands' by virtue of their different migration speeds through the gel (see Figure 3-3). DNA bands are detected in the gel by staining with ethidium bromide (EtBr) and through illumination by UV light. EtBr is an example of an intercalator: a molecule that inserts itself between two base pairs of double-stranded DNA. The EtBr molecule strongly fluoresces under UV illumination. This is a phenomenon whereby a molecule is excited by absorbing light at one wavelength and relaxes by emitting light at a longer wavelength. In this case, visible light is emitted allowing the DNA to be visualised as a bright band on the otherwise dark gel.

The amount of DNA and the size of the fragment determines the amount of EtBr bound and hence the brightness of the band.

Gel electrophoresis can be used not only simply to detect the presence of DNA in a solution, but also to determine the sizes of unknown fragments. A solution containing a number of different sized DNA fragments can be separated on the gel, by virtue of their different migration speeds through the mesh. An example of this is size standards or ladders; DNA samples which contain a variety of different fragments of known sizes. These separate on the gel, and can act as references to determine the size of unknown fragments.

Practically a 1-1.5 % agarose gel was produced by melting agarose powder (molecular biology grade, Sigma, UK) in 1 × Tris-acetate-EDTA (TAE) buffer, and adding 2 µl per gel of 10 mg/ml EtBr for nucleic acid visualisation. This mixture was poured into a horizontal slab gel rig, before a plastic sample comb was added to form wells for the addition of DNA samples, and allowed to set. Once set, the same 1 × TAE buffer was poured into the assembly such that the gel was submerged. Samples were mixed with 1 × Blue/Orange loading dye (Promega) and then loaded into the wells, together with a 100 bp ladder (Promega) in a separate well to allow effective size determination. Samples were then electrophoresed at 100 V until the Orange G component of the dye had migrated to the base of the gel, typically around 1 hour. Gels were then illuminated on a UV transilluminator and images were captured on a CCD camera. The DNA was run in non-denaturing conditions.

Agarose gels were also used in a quantitative manner to determine DNA concentration. In this case, an additional well containing λ DNA HindIII molecular weight marker was added. This contains λ DNA that has been digested with the restriction nuclease HindIII (see section 3.2.4), to produce a ladder of bands of known size. Each band is equimolar in concentrations, but smaller bands will appear fainter than larger ones, as there is less actual DNA material in that particular component. The intensity of each band can be compared to the known mass of the DNA in the band to determine the relationship between intensity and mass. As such, the total amount of DNA in

the sample bands can then be determined, together with the concentration. DNA concentrations were typically in the range 10-100 nM.

3.2.3 Spectrophotometry

DNA concentration was also measured using spectrophotometry. The spectrophotometer is a device that measures the intensity of light as function of wavelength, and is used to determine both the concentration and purity of a solution of DNA. The device consists of a light source, a sample holder and a detector (Figure 3-4). If a beam of light passes through a substance, a portion of the light will be absorbed by the substance, and the intensity of light exiting the sample will be diminished. Nucleotides in solution are able to absorb light in the UV region of the spectrum, with a maximum absorption at a wavelength of 260 nm. Proteins also absorb UV light, but closer to 280 nm wavelength. Therefore the purity of the sample can be determined by measuring the absorption at both 260 and 280 nm, and comparing the results to those obtained from pure solutions. The ratio of absorbencies (A_{260}/A_{280}) should be equal to 1.8 for a sample of pure DNA. If protein contaminants are present in the solution the absorbance at 280 nm increases more than at 260 nm, as proteins absorb light at 280 nm better, so the ratio A_{260}/A_{280} decreases.

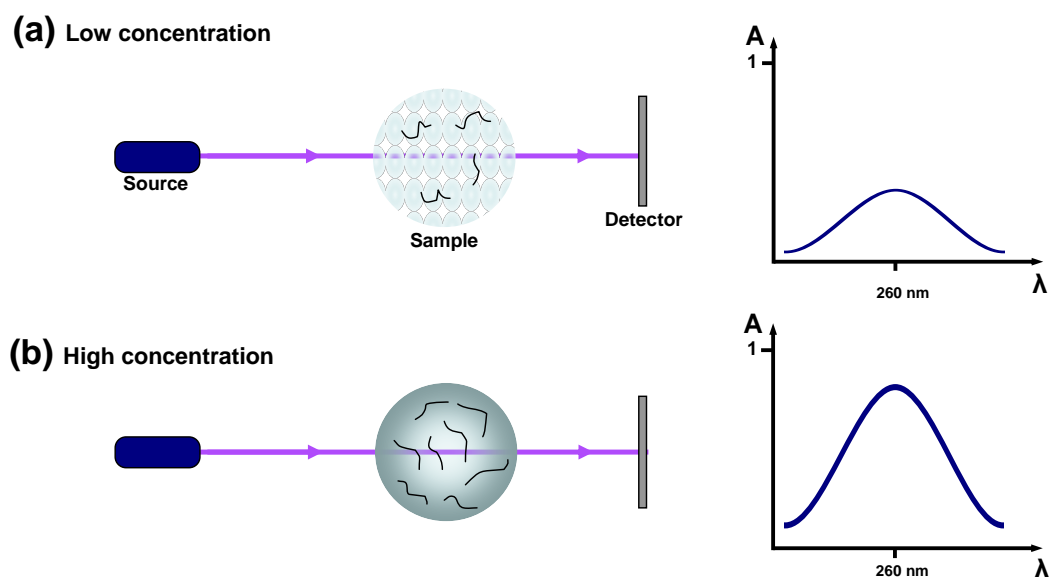


Figure 3-4. Schematic diagram of a spectrophotometer. A light source emits a beam of UV radiation, which passes through solution of DNA, before being received by a detector. The DNA solution absorbs some of the UV light, and by measuring the amount of light absorbed the DNA can be quantified. Solutions with low DNA concentration (a) absorb less light than those with a high DNA concentration (b).

The spectrophotometer can also be used to quantify solutions of DNA, as highly concentrated samples absorb more light than samples of low concentration (compare schematics in Figure 3-4). The device is usually calibrated before use. This involves measuring the absorbency of a reference sample, such as water, which is then set to a baseline value of zero. The spectrophotometer then measures the percentage of light absorbed relative to the original reference. The absorbance of a number of DNA solutions of known concentration can be measured, and plotted to produce a reference curve, showing absorbance as a function of concentration. The absorbance of the DNA sample to be tested is then measured, before being compared with the reference curve to find the accurate concentration. Although both the spectrophotometer and gel electrophoresis methods of DNA concentration quantification gave similar values of concentration, it was felt that the optical density readings were more reliable and quicker to obtain, and so this technique was used preferentially.

3.2.4 Restriction Enzymes

Long fragments of linear DNA (length of the order of kbp) were obtained following enzymatic digestion of plasmid DNA. Restriction endonucleases act as molecular scissors, allowing the fragmentation of large DNA molecules. Originally, it was observed that when DNA molecules from one strain of *E. Coli* was introduced into another strain, the foreign DNA was almost always broken down into smaller pieces. Occasionally the infecting DNA was not broken down but was somehow modified. Subsequent studies showed the presence of methylated bases in the surviving molecules [131]. Extracts of cells of *E. Coli* revealed the presence of a specific modification enzyme that methylated DNA and a restriction enzyme that broke down unmethylated DNA [119]. Further investigation led to the discovery of restriction nucleases inside other bacteria, with the important feature that they only cleave DNA at specific nucleotide sequences [104]. A nuclease is a type of enzyme that is capable of cleaving the phosphodiester bonds between nucleotide units. In this case, the enzymes restrict the transfer of DNA between certain strains of bacteria so the name restriction nuclease was chosen for them.

The restriction endonucleases used for research are mainly purified from bacteria, where they protect the prokaryotic cell from bacteriophage invasion. *In vitro* they sever the backbone of both strands at specific target sites on the DNA chain. The target areas are small (typically 4-8 nucleotide pairs), and so in a long DNA molecule there will be several possible cleavage sites. The target sites are often palindromic; the nucleotide sequence reads the same as the complementary strand in reverse. There are over 100 commonly used restriction enzymes, with names related to the bacteria they were isolated from [169]. They make their cuts in one of two ways: either straight through both strands at the same point to produce fragments with *blunt ends*, or displaced equally in opposite directions from the line of symmetry to produce fragments with single-stranded regions known as *sticky ends* (see Figure 3-5). They are named in this manner as they are free to base pair (stick) to complementary sequences from DNA cut from the same restriction enzyme.

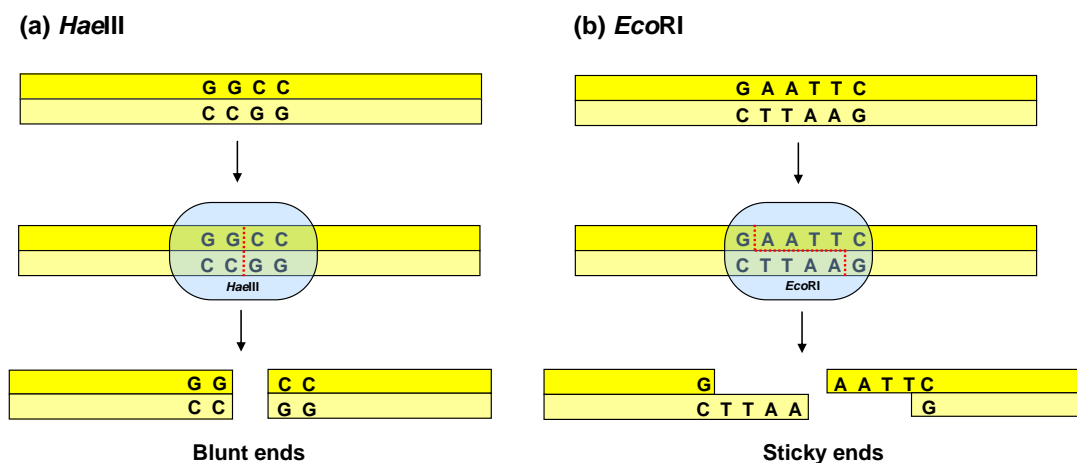


Figure 3-5. Diagram showing the action of two restriction enzymes: HaeIII (a) and EcoRI (b). The enzyme is shown as a blue oval. The enzymes can recognise the target sequences shown, and begin to cut both strands at specific points in the sequence (red dashed line). This produces two separate fragments, with blunt ends in the case of HaeIII and sticky ends with EcoRI.

To cut a DNA molecule with a restriction enzyme, a solution containing the DNA is put into a small test tube and a solution containing the nuclease is added. The reaction mixture is incubated at the enzymes optimum working temperature (often 37 °C). Time is allowed for the enzyme to find its recognition sequences and cleave the DNA backbone. Subsequent to this the digested fragments can be separated. The method generally used is gel electrophoresis. Staining of the DNA allows the fragments to be observed, and the fragment of interest can be extracted and purified.

3.2.5 Column Purification

As AFM is able to provide a complete picture of the molecule under study, it is important to have a clean solution of DNA. Impurities in AFM samples can be visualised as large globular features on the surface, which can obscure important features of the DNA. Any purification method must remove extraneous material left over from the PCR reaction, such as primers, nucleotides, enzymes, and buffer salts.

Two techniques of purification were used: QIAquick PCR Purification (Qiagen, Valencia, CA) and QIAquick Gel Purification (Qiagen, Valencia, CA). Both purification methods utilise similar protocols (see Figure 3-6 for main principles). The DNA-containing solution is transferred into a spin column, which contains a silica membrane. DNA is able to adsorb to the membrane, while contaminants remain in solution. The kits use special buffers that enable the efficient recovery of DNA and the removal of contaminants. During the washing step, buffer is added before the column is centrifuged. During this step unwanted impurities pass through the column and can be disposed. Finally, the DNA is eluted by centrifuging the column with a Tris buffer.

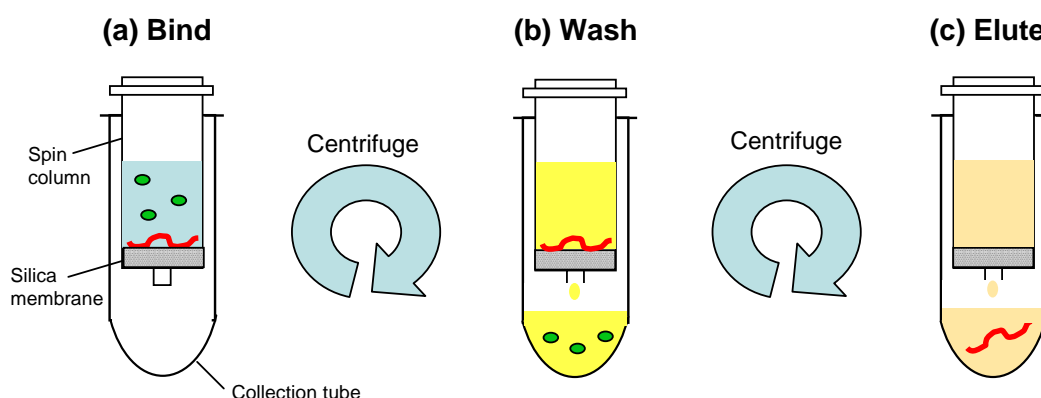


Figure 3-6. The spin purification procedure. (a) The DNA-solution is transferred into the spin column. DNA (shown as a red line) adsorbs to a silica membrane. Impurities (shown in green) remain in solution. (b) The solution is then washed with a buffer. Impurities pass through the column into the collection tube, and can then be disposed of. (c) Another spin with an elution buffer causes the purified DNA to be passed through the column.

The PCR purification kit was used when a solution contained only a single DNA component e.g. PCR-generated linear fragments. The gel extraction kit was used to separate samples when a range of fragment sizes or components (e.g. supercoiled and relaxed plasmids) were present in a single volume. This involved an additional step, whereby the inhomogeneous sample was run on a gel to separate the fragments into discrete bands. Different fragment sizes were identified by comparison with a DNA ladder. The fragments to be purified were

excised from the bulk gel using a clean scalpel before being melted and re-suspended into solution, and were subsequently purified. Both purification techniques were found to provide samples that could be effectively imaged under AFM with a minimum of contaminant particles.

3.2.6 Circular Samples

A number of experiments used small, circular molecules of DNA known as plasmids. These molecules are isolated from bacteria, and are located away from the main chromosome of the species. As early as the 1960s it was discovered that circular DNA molecules exist in two main structural forms. These different forms were observed using electron microscopy, during which it was seen that one component contained mainly open rings, whilst the other form was more compact, containing more crossings of the DNA backbone. It was also found that the twisted circular form of the DNA can be converted into the open form by a single break in one of the strands of the phosphodiester backbone.

The compact form arises from the relative twisting of both ends of the DNA chain, prior to being joined to form a circular molecule. The coiling of the DNA helix upon itself is known as supercoiling. This process introduces large amounts of strain into the molecule, and can cause the molecule to contort in space, leading to the crossing over of the helix and a more compact conformation. There are two opposite forms of supercoiling; a direct consequence of the DNA helix being right-handed. Right-handed twisting causes the helix to twist up further, and is known as positive supercoiling. Conversely, left-handed twisting causes the helix to untwist, and is known as negative supercoiling. A single cleavage of one phosphodiester bond (nicking) will mean the broken strand can rotate about the intact strand to dissipate the torsional stress created by supercoiling. The supercoiled form is often known as closed-circular DNA, and the nicked form as open-circular DNA (see Figure 3-7 for examples).

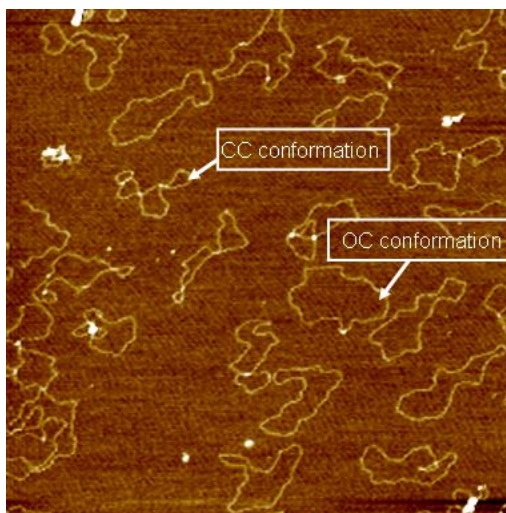


Figure 3-7. Comparison of different plasmid fragments of plasmid pBR322 imaged together by AFM. Here the spread-out open circular (OC) conformation can be seen alongside the more condensed, supercoiled closed circular (CC) plasmid, where -self-crossing of the strand is observed. Image is 5 μm \times 5 μm .

Plasmid pBR322 (4.2 kbp) was used for studies into circular DNA molecules. A sample of circular DNA will consist of two components: the naturally occurring closed circular form, in addition to the open-circular, formed through stress causing the nicking of one strand, leading to the release of the topological strain. The two forms were separated by virtue of their different migration speeds through agarose gel electrophoresis (Figure 3-8). The more compact closed-circular form moves faster than the spread-out open-circular conformation. The two fragments were then visualised by EtBr staining, and purified using the gel extraction method.

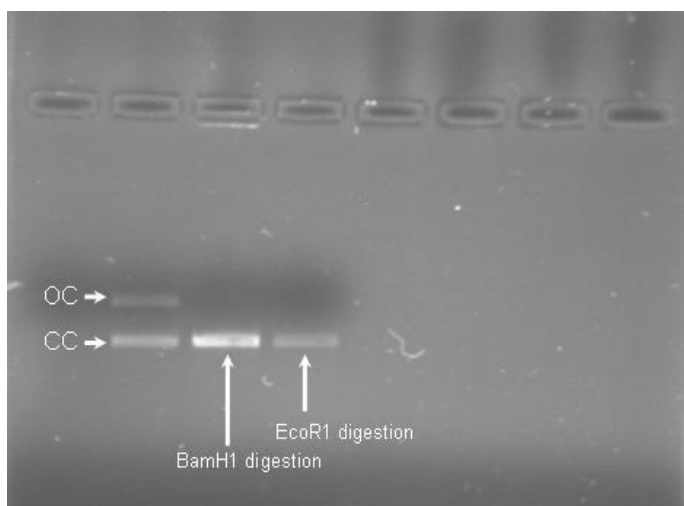


Figure 3-8. Gel electrophoresis of a plasmid preparation and its linear digestions. The two components of the plasmid can be seen: The closed circle (CC) moves through the gel faster than the open circle (OC) conformation because of its more compact shape. Only one fragment is visible after restriction enzyme digestion (In this case EcoR1 and BamH1), corresponding to the linearised plasmid. All the molecules exist as linear fragments when seen under AFM.

In order to create longer linear molecules, the pBR322 plasmid was digested using restriction enzymes. The restriction enzymes used were either BamH1(Promega) or EcoR1 (Promega), which are specific to different sequences. BamH1 recognises the sequence 5'-GGATCC-3' while EcoR1 recognises 5'-GAATTC-3'. The digestion of the plasmids was performed in the same way for both enzymes. A solution was made up containing the plasmid, restriction enzymes and buffers specific to that enzyme, which was incubated for two hours at 37 °C allowing digestion to take place. Plasmid pBR322 contains just a single recognition site for both BamH1 and EcoR1, meaning that digestion results in the linearisation of the circular molecule, rather than the production of fragments of different sizes. When imaged under AFM, the two different digestions are indistinguishable (Figure 3-9).

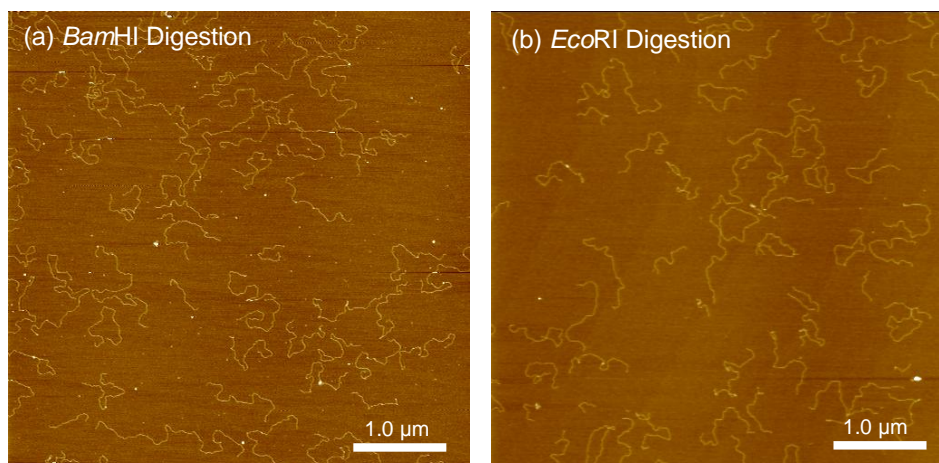


Figure 3-9. AFM images of linearised plasmid pBR322. The end result of both digestions is the same: linear fragments of 4.2 kbp size.

3.2.7 AFM imaging of DNA on mica

Muscovite mica (Agar Scientific) ($\text{KAl}_2(\text{AlSi}_3\text{O}_{10})(\text{OH})_2$) was used as a substrate for all DNA samples. Prior to deposition, 30 fmol of purified DNA samples were diluted tenfold into an imaging buffer consisting of Tris-HCl (4 mM, pH 7.5) and 4 mM MgCl_2 . Samples were typically prepared for imaging by depositing 10 μl of DNA solution in buffer onto freshly cleaved muscovite mica and incubating for 3 minutes at room temperature. After this, the samples were rinsed in milliQ water, before being dried in a weak flux of nitrogen gas. Variations on this protocol are explained in the text.

AFM images were collected in air with a Multimode Nanoscope IIIa AFM (Veeco, Santa Barbara, CA) operating in Tapping Mode using silicon cantilevers (OMCL-AC160TS, Olympus, Tokyo, Japan) of quoted spring constant 42 Nm^{-1} and typical resonant frequency 300 kHz. There may be some variation in spring constant between different tips of the same type, and if the value is too high damage to the molecule may occur. If during imaging the resolution appeared poor, or scanning led to deformation of the molecule then the tip was replaced and a new area on the sample imaged. Scans were collected at a line frequency of 2 Hz at 512×512 pixel resolution.

3.3 Sample Preparation Investigations

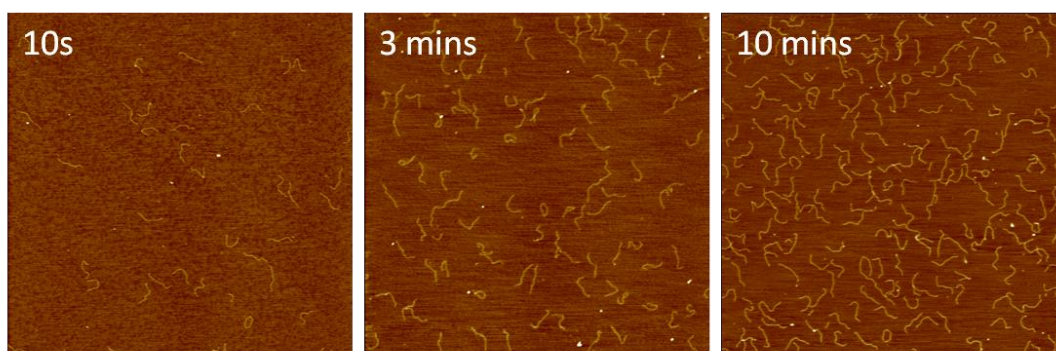
DNA was one of the first biological molecules imaged using AFM and is still one of the most widely studied biomolecules with researchers. Before imaging, the molecules must be immobilised on a surface; this typically involves the addition of a DNA solution onto a surface, before incubation, rinsing, and finally the drying of the sample in a flow of compressed gas. Preparation of mono-dispersed DNA molecules on a mica substrate depends on a number of different factors including DNA concentration, incubation time, buffer conditions, drying procedure and mica pre-treatment. An initial aim of this research was to perform a systematic investigation into how these variables affect imaging, with a view to optimising surfaces for the study of DNA-protein interactions. We focus on improving dispersion of DNA on mica for studies at the single molecule level. Samples were prepared using the sessile drop method.

3.3.1 Deposition kinetics

The first step of the sample preparation process involves a small drop of purified DNA in buffer being placed onto the sample surface using a pipette, in this case a mica substrate, before being left to incubate over a period of time. Providing that the DNA experiences an attractive interaction towards the mica the molecule will be able to bind to mica. DNA deposition consists of two parts: DNA transport from solution towards the surface, and DNA binding once at the surface. An experiment was performed to study the deposition process further, whereby a 10 μ l drop of DNA (30 fmol) in deposition buffer containing Mg(II) was transferred onto a cleaved mica disc. The DNA was incubated on the surface from various time periods from 5 s to 10 minutes, before the disc was rinsed and then dried in a weak flux of argon gas. As the deposition time became longer increasing amounts of impurities were observed on the surface, and could be a result of the crystallisation of buffer salts. The behaviour of two different linear fragments was investigated; namely a short 800 bp fragment and a significantly longer 4.2 kbp linearised plasmid. Figure 3-10a shows a

series of images of 800 bp DNA at increasing deposition times. These images appear to demonstrate that as the incubation time is increased, the number density of DNA on the surface appears to increase accordingly. The same experiment was performed using the larger 4.2 kbp linearised fragment. A similar result was obtained, with the number of molecules increasing with the time the drop is left on the surface before drying (Figure 3-10b).

(a) 800 bp



(b) 4.2 kbp

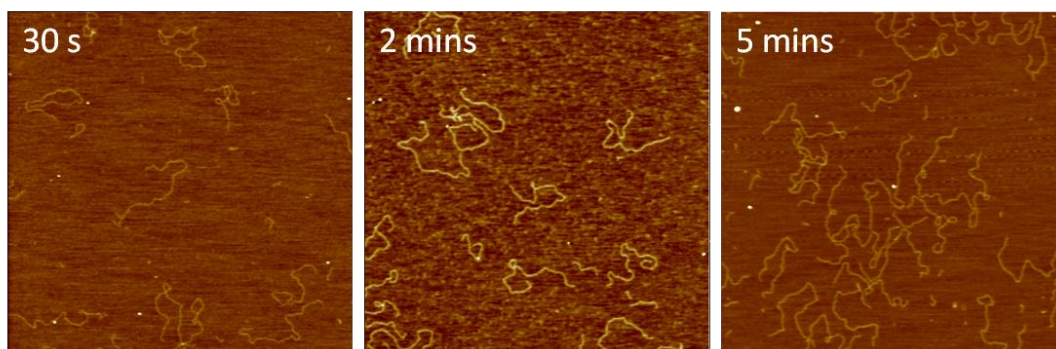


Figure 3-10. Comparative AFM images of (a) 800 bp and (b) 4.2 kbp fragments incubated for increasing time. With the increase in deposition time the number density of DNA on the surface increases. Each scan is $3 \mu\text{m} \times 3 \mu\text{m}$.

In order to get a more quantitative view of how DNA behaves during deposition, a number of different areas were imaged for each time point. The number of molecules in the imaging area was then counted manually, and normalised to calculate the average number of DNA molecules per μm^2 . Only molecules where more than half of the backbone was in the imaged frame were scored.

The results of such analysis are shown in Figure 3-11. Here it can clearly be seen that with increased incubation time, the surface density of DNA increases concurrently. The points for both samples were fitted to curves of the form $y=At^{1/2}$. The experimental data line up approximately with the curves, with R^2 values of 0.91 and 0.87 being obtained for the 800 bp and 4.2 kbp fragments respectively, indicating that such a relationship is valid. A $t^{1/2}$ relationship is typical of a diffusion controlled process [167]. The multiplication factor A is directly related to the diffusion coefficient D and gives a measure of how fast a species diffuses from an area of high concentration to an area of low concentration. Values of the constant A were calculated as 9.30 for the 800 bp fragment and 1.53 for the larger fragment. The ratio of these numbers is roughly similar to the ratio of the actual sizes of the DNA fragments. The curves show that the smaller fragments can move through the solution quicker and bind more readily. Smaller masses moving quicker than larger masses would also be typical of diffusive behaviour.

The experimental points were also fitted to exponential curves, with R^2 values of 0.94 (800 bp) and 0.82 (4.2 kbp) being obtained. Exponential dependences are typical of chemical reactions where an activation barrier controls the reaction rate. Although the points line up approximately with a curve of such form there is not expected to be any chemical reaction occurring between DNA and mica. Surface binding can be achieved through a chemical strategy where the molecule is covalently linked to the surface. This can be achieved by chemically functionalising the substrate with positively charged groups, such as amines (NH_2). Aminosilanes, particularly aminopropyltriethoxysilane (APTES), have been used to create positively charged surfaces, meaning that negatively charged DNA will interact strongly with the substrate [48, 123, 184, 186]. Due to the absence of any mica surface treatment in the experiments described here, the most straightforward explanation of the deposition of DNA onto mica uses physisorption processes, where the environment and the surface are controlled to immobilise the molecule via the natural forces present. In such a case a $t^{1/2}$ curve is expected to be more valid.

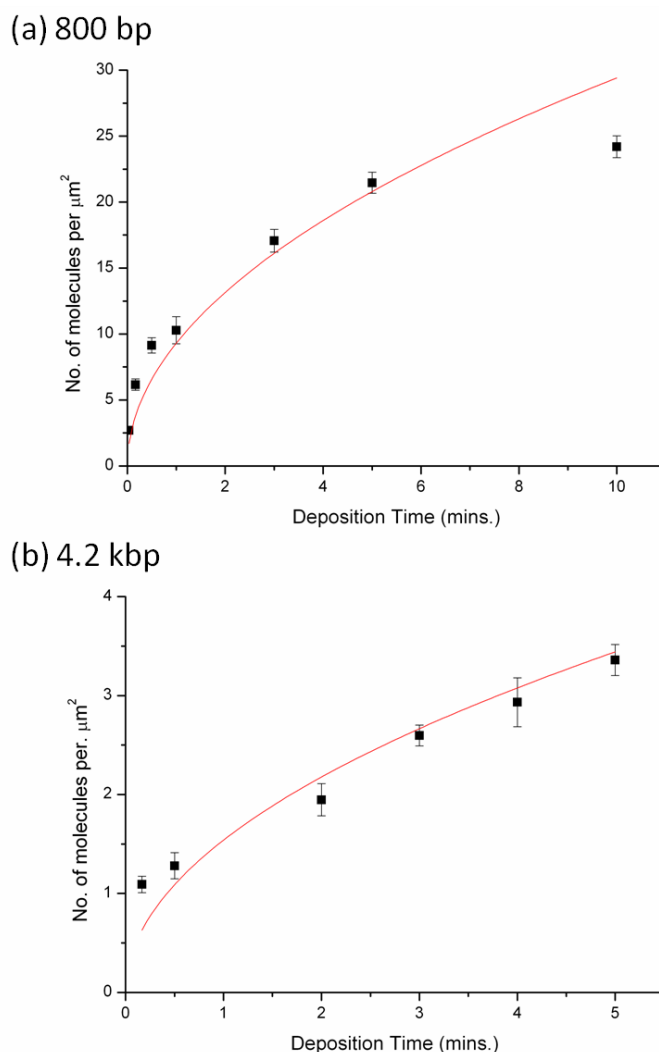


Figure 3-11. The deposition kinetics of (a) 800 bp and (b) 4.2 kbp DNA fragments on mica. The number density of DNA molecules bound to the surface with varying deposition times from seconds to minutes was calculated and plotted.

Several studies have also shown that the adsorption of DNA from a solution onto a surface is a diffusion-governed process [113, 167]. During their studies of DNA adsorption onto a protein membrane, Lang and Coates stated that random thermal motions (diffusion) was responsible for binding only if a number of conditions were satisfied [113]:

1. The DNA is irreversibly bound to the surface.
2. The DNA solution is sufficiently dilute to prevent molecules interacting with each other.
3. DNA binding sites do not become saturated.

4. The concentration of DNA in solution should be large compared to that on the surface, such that the concentration near to the substrate does not diminish significantly.
5. Diffusion is the sole transport mechanism. Other effects such as convection and sedimentation are neglected.

Experiments have shown that the effect of sedimentation is small, compared to the average displacement by diffusion. Additionally, if DNA is transported by convection currents, from for example temperature gradients, the number density of bound DNA would be directly proportional to time t [113]. Similarly the time dependence will be higher than 0.5 if the DNA is not irreversibly bound to the surface. At the length scales studied here, bare mica with no bound DNA could still be observed when imaging, making it likely that DNA binding sites were not yet saturated.

Rivetti also studied the deposition process of DNA onto mica with AFM, and found that the transport of DNA molecules from a sufficiently dilute solution was solely governed by diffusion and that the molecules were irreversibly bound to the surface [167]. Our AFM studies into the kinetics of DNA adsorption with two different fragment sizes also agree with this observation.

3.4 Comparison of different sample rinsing regimes

3.4.1 Effect of varying rinsing volume

After the incubation period, the mica surface is usually rinsed with pure water to remove any remaining buffer salts or weakly bound DNA. Ideally this process should leave a clean surface with homogeneously dispersed molecules. The rinsing process was investigated in further detail to observe the effect of the volume of rinse used on the DNA surface concentration seen under AFM. In the beginning samples were prepared in the same way: 10 μl of DNA solution was added and allowed to incubate for 3 minutes. The final rinse was performed using a P1000 Gilson pipette to provide a stream of milliQ water. Using this

method, either a single stream of 1 ml of water, or three separate streams making a total rinsing volume of 3 ml was used to rinse the surface. A series of images were taken of both the 800 bp linear fragment, and the much longer 4.2 kbp linearised plasmid (Figure 3-12 and Figure 3-13).

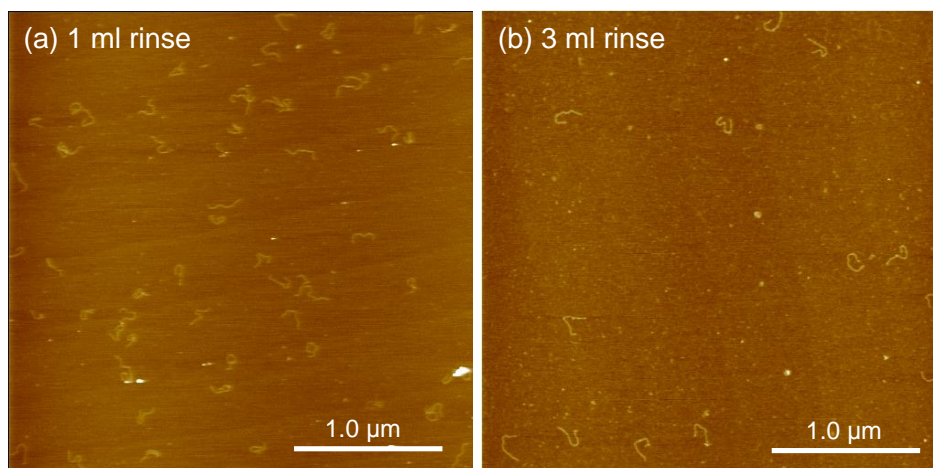


Figure 3-12. Comparison of different rinsing volumes for 800 bp fragments. There is a significant difference between the surface concentrations for the different volumes. This suggests that the DNA is weakly bound and is easily washed off by water.

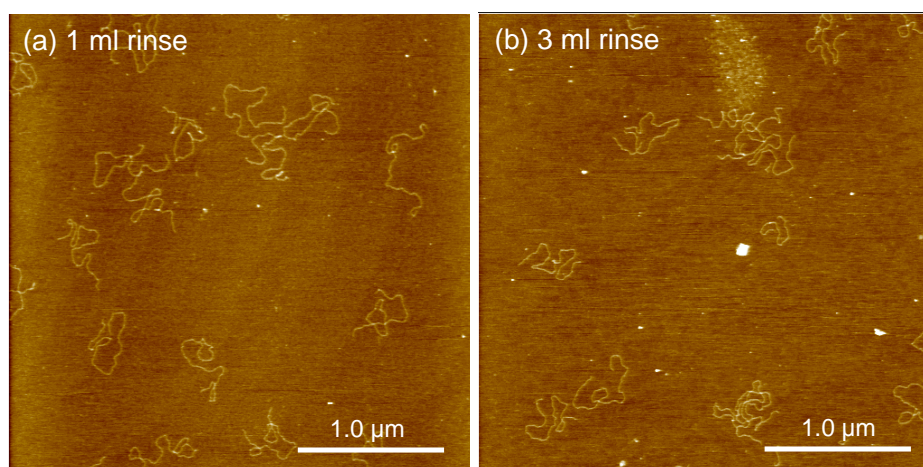


Figure 3-13. Comparison of different rinsing volumes for the linear plasmid fragments. There is less of a difference between the two different methods, indicating that it is harder to remove the molecules from the surface.

Looking at the figures above, it seems clear that with more rinses there are fewer molecules on the surface. To make this outcome quantitative a number of different areas were imaged for each case. For the 800 bp fragment, the average number of molecules in a $3 \mu\text{m}^2$ area was 52 for a 1 ml rinse and 15 for a 3 ml

rinse. When three rinses are performed instead of one, there are roughly only a third of the molecules that were previously seen. This suggests a linear dependence between rinsing volume and DNA surface density for these small molecules.

For the much longer 4.2 kbp fragments, the differences are much less marked: for a 1 ml rinse an average of 16 molecules were seen, while for a 3 ml rinse 12 molecules were observed. It appears that the larger fragments are more strongly bound to the surface, and are less readily displaced off the surface by rinsing. This could be related to the energy cost of breaking the interactions between the DNA, mica and their respective counter-ions. If we consider the mica surface as a lattice of charges [147], and that DNA is bound by counter-ions lying in cavities on the surface, there will be many more points-of-contact for the larger fragment than the smaller. These contacts provide an adhesion force, which is ultimately electrostatic in origin, against the flow of water. Larger molecules remain bound as many more interactions must be broken to cause the molecule to detach from the surface. In terms of AFM studies of DNA-protein complexes, extensive washing has been seen to reduce the number of RNAP· σ^{54} complexes specifically bound to DNA, which was attributed to sliding of the holoenzyme along the template during the rinsing process [179].

3.4.2 Washing with alcohol

Once it appeared that rinsing with too much water can displace weakly-bound molecules of DNA, another washing strategy was attempted. Solutions containing alcohol (usually ethanol or isopropanol) are routinely used to extract DNA for molecular biology studies [176]. The addition of alcohol makes the solution more hydrophobic, by lowering the dielectric constant of the solution, thus reducing the solvating power of the solvent (water) [200]. The reduction of the solubility of the nucleic acid causes a white precipitate to form. The precipitate can be separated from solution by centrifugation, before the pellet is dried and re-suspended in the desired buffer.

The aim here was to see whether the use of alcohol would be beneficial in AFM sample preparation, and was used with the hope of trying to get a cleaner surface with fewer impurities bound to the mica. An experiment was performed whereby 10 μl of open circular plasmid DNA solution (50 fmol) was incubated on mica for 3 minutes, before 1 ml of 50 % isopropanol was used to rinse the surface. The sample was then dried in a weak flux of nitrogen gas. Also, when rinsing with water it may be possible for DNA molecules previously bound to the surface to be dissolved back into solution. Rinsing with an alcohol should prevent this from occurring by preventing surface bound DNA from dissolving back into the bulk solution. .

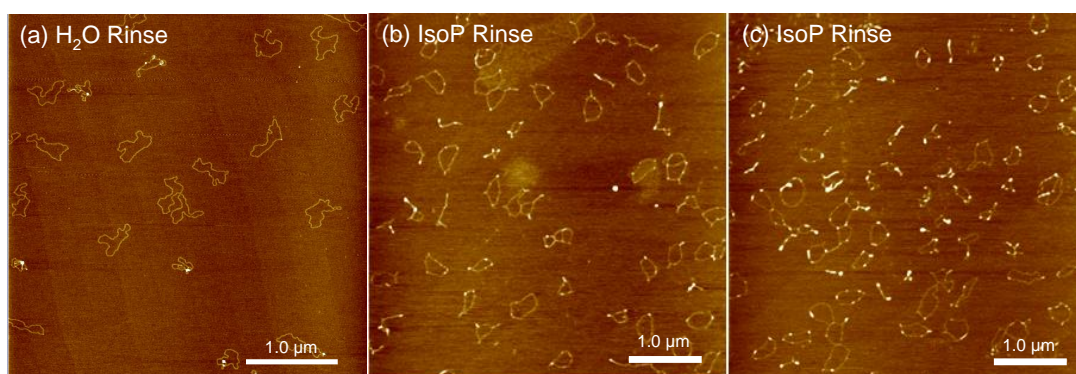


Figure 3-14. Comparison between a conventional rinse with milliQ water, and with isopropanol. (a) The molecules take up a relaxed, surface-equilibrated shape. (b) and (c) Two separate examples of isopropanol rinsing. The alcohol causes the plasmid to have a polygonal conformation, with localised kinking and/or DNA condensation. This is likely an effect of DNA hydration.

Figure 3-14b and c shows the results of two separate preparations, in which the substrate was rinsed with isopropanol. The plasmids show very different conformations from a sample prepared in the conventional manner (Figure 3-14a). In contrast with the curved semi-flexible nature of the plasmid in the conventional preparation, the DNA chain has a much more angular shape after alcohol rinsing. The molecules appear limited in their conformational freedom, and lack any significant smooth bending, indicating that the effective persistence length of the molecules may increase. Also apparent are a number of molecules that have condensed significantly and no longer appear circular in shape. A number of aggregates have also formed, some containing as many as

five molecules, reflecting increased DNA-DNA interaction in alcoholic environments.

Lang *et al.* subjected bacteriophage DNA to various concentrations of ethanol, and visualised the molecules with electron microscopy [112]. At high ethanol concentration he observed polygonal meshes of DNA fibers. The fibers between nodes were straight (similar to observed in the plasmids here), and was attributed to the fact that the molecules were subjected to contractive stress during the adsorption to the imaging substrate. Also found in Lang's study were very condensed structures of DNA, caused by coiling of the DNA helix upon itself, which was a result of dehydration by the ethanol [112]. This dehydration is likely to cause the DNA to convert from B-form to A-form DNA.

Another experiment was carried out to further investigate how alcohol affected plasmid molecules. Samples that had been previously prepared in the conventional manner (using milliQ to rinse the surface) were used. Imaging under AFM showed that these molecules were circular rings of DNA (plasmids). The same sample was then submerged under 50 % isopropanol for 10 s and dried in a weak flux of nitrogen gas. Re-imaging showed that the plasmid molecules had been broken after exposure to the alcohol (see Figure 3-15). The breakage of the chain is caused by contractive stress, and supports a B- to A-form DNA transition. One can still observe the general outline where a circular molecule would have been before, but in its place are a series of different sized linear fragments. This shows that alcohol appears to induce stress into the molecule, and is enough to break certain sections of the DNA backbone. This is a consequence of the interplay between the DNA interaction with mica, and the helix contraction.

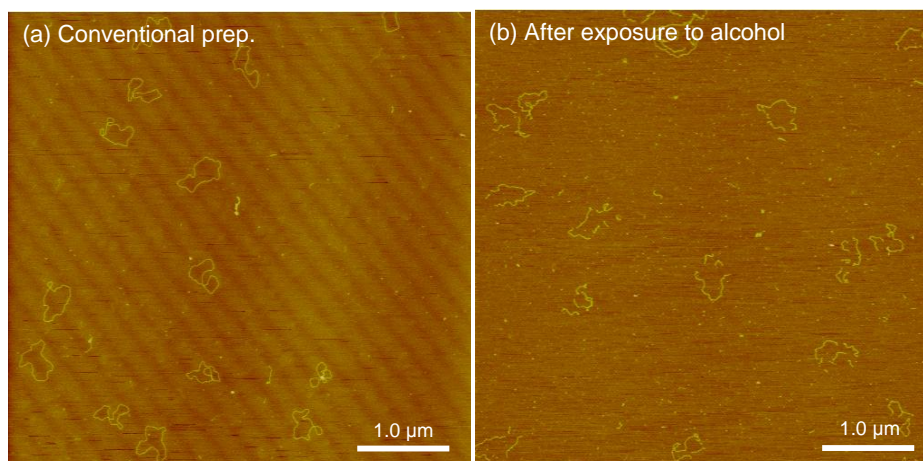


Figure 3-15. Effect of alcohol on a conventionally made sample. Exposure of alcohol leads to the severing of the DNA backbone at a number of places. This is likely a consequence of dehydration and associated B-form to A-form transition of the DNA.

Once it became clear that the isopropanol rinse led to undesirable results, another approach was taken involving the use of ethanol. It appeared that the flow of alcohol across the sample may have been too harsh a process, having a dehydrating effect in the molecule. In the new case, after incubation with DNA solution, the sample was submerged under a small amount of ethanol for 10 s, before it was then dried in a weak flux of gas. Figure 3-16 shows a series of images of 800 bp linear DNA taken after the sample was subjected to various strengths of ethanol. At low to medium alcohol concentrations (20-60 %) the DNA appears well separated and relaxed, not too dissimilar to the morphology seen in the absence of ethanol (0%). At 80 % concentration the DNA begins to appear more angular. Finally, after exposure to a highly concentrated ethanol solution, the DNA takes up a range of complex structures, such as toroids, polygons and aggregates. Also observed were large flower shaped multi-molecular complexes containing tens of molecules (Figure 3-17). It does not appear that the use of alcohol for washing has increased the overall number density of DNA bound to the surface.

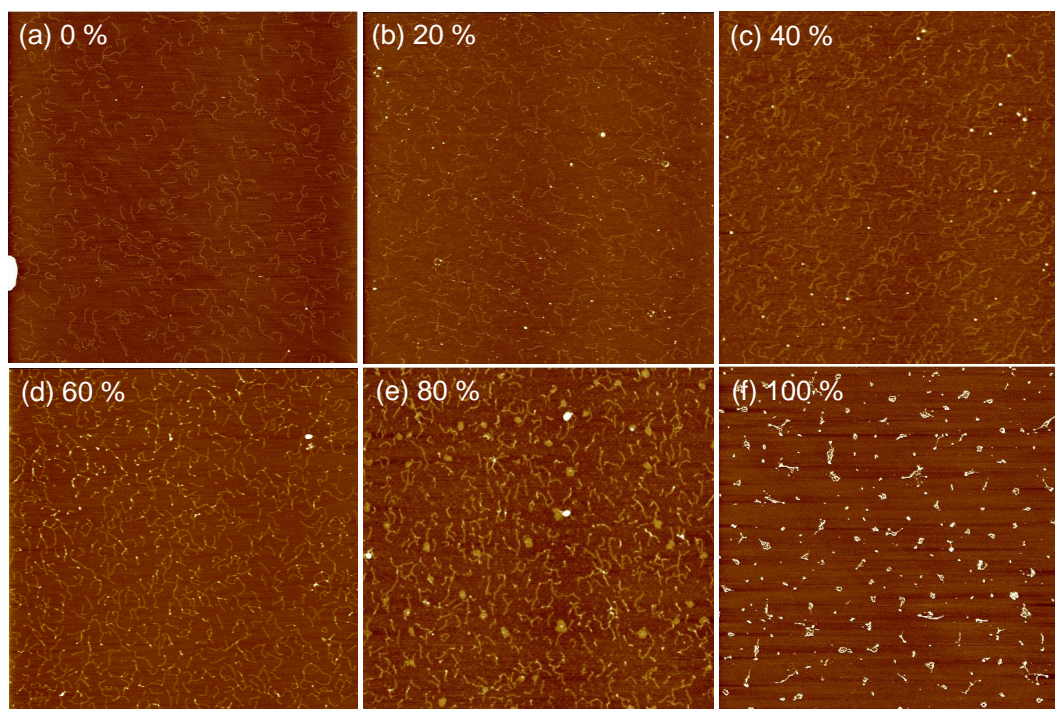


Figure 3-16. Series of images of samples prepared by washing the mica surface with adsorbed DNA with different strengths of ethanol. At low to medium concentrations (20-60 %) the surface appears similar to the conventional sample prep (0%). As the percentage is increased further (80-100 %) the conformation of the DNA begins to change. At 100 % ethanol condensed and toroidal structures can be observed. Scale is $5\mu\text{m} \times 5\mu\text{m}$

On hydrophilic surfaces, such as mica, thin water layers exist that more than likely allow surface-bound molecules to retain a degree of hydration. Accurate AFM measurements of the DNA contour length have shown that the chain exists as a mixture of B-form and A-form DNA [68, 69, 165, 173]. Fang *et al.* also studied the effect of ethanol on DNA confined on mica [68]. In their study, contour length measurements showed a transition from all B-form at 0 % ethanol to completely A-form at 30 % ethanol. It seems that alcohol acts to dehydrate the DNA and change its structure. At higher ethanol concentrations they observed increasingly looped and complex structures, such as toroids and large aggregates.

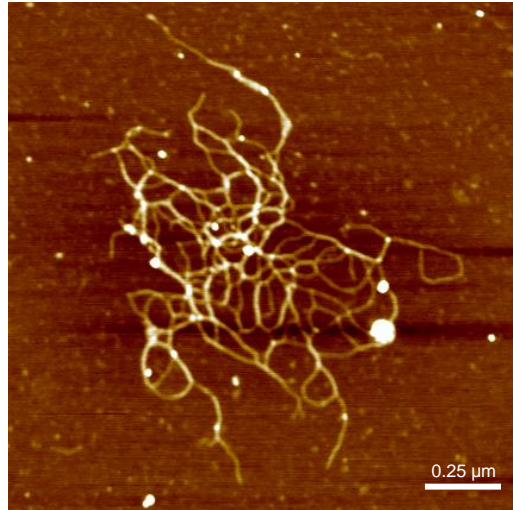


Figure 3-17. Occasionally, after washing the sample with highly concentrated ethanol solutions, complex structures were observed. This flower shaped aggregate contains tens of molecules.

Fang *et al.* also found that the persistence length of the DNA decreased with ethanol concentration [68], meaning the stiffness of the molecule was also decreased. Increased bending of the chain could account for some of the complex structures seen in this study. It has also been shown that alcohol can lead to an increase in DNA supercoiling [112, 200]. Previous studies have shown that water acts to stabilise the DNA double helix. On the other hand alcohol has been seen to have a dehydrating effect on DNA, removing water from the molecule and destabilising the double helix [67, 90]. This destabilisation, coupled with the fact that the counterions present in the deposition buffer are able to screen the negative charge of the DNA backbone, could allow the DNA more freedom to form more complex structures.

It has been shown that a relatively low concentration of ethanol (~30 %) is needed to cause DNA on mica to take up an entirely A-form [44]. It appears that the mica substrate helps to facilitate this change, as in solution the B-A transition occurs at 70-80 % ethanol transition [98]. B-DNA is the hydrated form of DNA, and the structure most commonly found *in vivo*. As a result it is best to study biological processes such as transcription using hydrated AFM samples, in order to get as similar to physiological conditions as possible. The use of alcohol in preparation, and the associated B-A transition that this brings could introduce anomalies into any results. In addition to this the alcohol

rinsing gave no noticeable benefits to the conventional preparations, in terms of increased DNA surface density, or cleanliness.

3.4.3 Drying Investigation

The final step of sample preparation involves drying the substrate in a weak flux of gas, usually nitrogen or argon. This removes the liquid layer leaving surface-bound DNA. An experiment was performed to investigate the effect that the drying step has on how DNA appears under AFM. The aim was to dry the surface from different directions, and then assess whether this influenced the molecular conformation seen under AFM. This involved the use of a mica disc that was marked on its underside with a cross (see schematic in Figure 3-18), and allowed the drying direction to be varied in a controlled way. A drop of 800 bp DNA (30 fmol) in deposition buffer was deposited onto the surface and left to incubate for 3 minutes, after which the surface was rinsed in 1 ml of milliQ water. Following this, the substrate was dried in one of three different directions: in the plane of the disc from left to right or up to down, or from directly above the surface until there was no visible liquid left. Drying was performed at 1 bar of pressure. Figure 3-18 shows a pictorial representation of how the disc was dried and representative images associated with each method. On first sight there does not appear to be any overriding differences between each image.

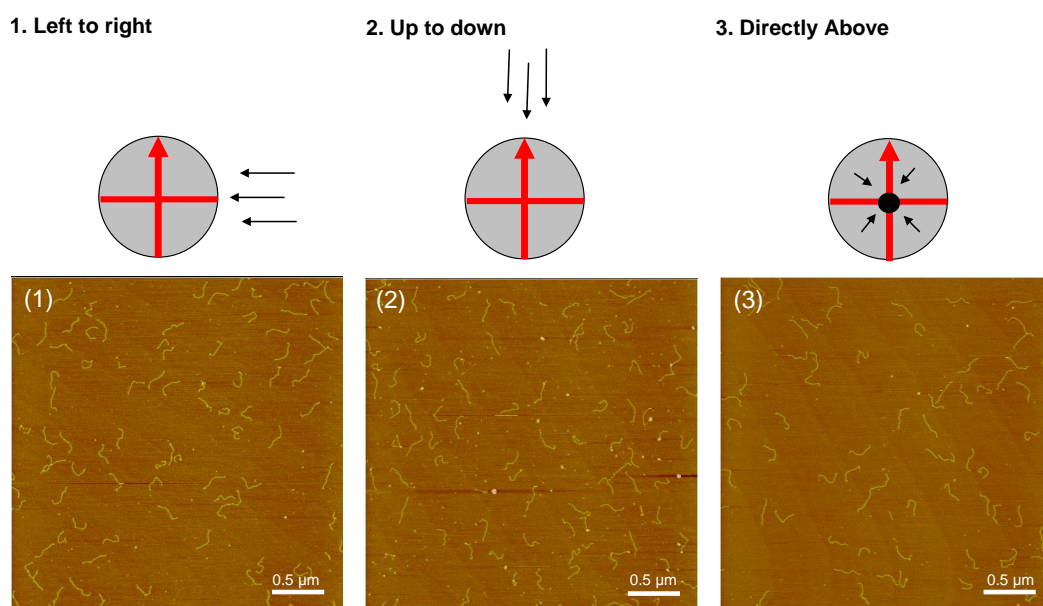


Figure 3-18. Diagram showing each mode of drying. In two methods, the flow was in the plane of the disc, but in different directions. In the third case drying was performed from directly above the disc. There is an accompanying image for each direction, showing the surface after drying under AFM.

To investigate the drying effect more qualitatively a classification scheme was used, in which each molecule was classified as belonging to one of three specific classes denoted by its predominant conformation. Figure 3-19 shows software zooms of three molecules to illustrate the different groupings. The first molecule lies with its backbone predominantly in the vertical axis, the second is horizontally aligned, and the third takes up an irregular curved conformation. A series of images were taken for each drying direction, and a tally of the occurrence of each shape of molecule was taken. Between 200 and 300 molecules were classified for each drying direction.

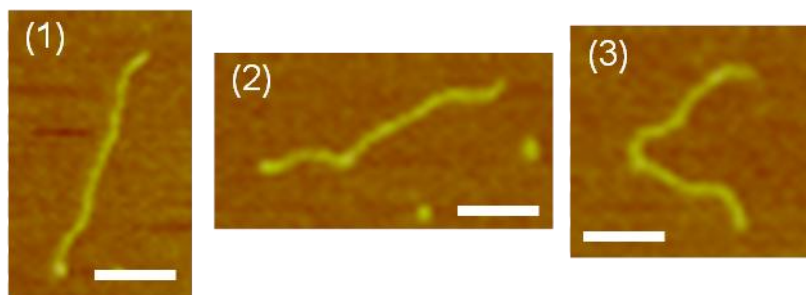


Figure 3-19. Software zooms of individual 800 bp molecules used to illustrate classification scheme used. (1) Molecule lies predominantly in the vertical direction. (2) Molecule lies predominantly in the horizontal direction. (3) Molecule is irregular. Scale bars are 100 nm.

Figure 3-20 shows a plot of the fraction of molecules taking up each particular conformation for each of the drying directions, with the results for horizontal conformation in green, vertical in red, and irregular in blue. The results when drying in the plane of the disc are almost identical. If the flow of gas were to have a combing effect on the molecules one might expect there to be a trend for vertical molecules when the flow of gas went down the disc, and the opposite to be true when drying was done across the disc, with a greater number of horizontally-aligned molecules being observed. However, this does not appear to be the case here.

In general there are similarities between the results for each direction of drying, with more vertically aligned molecules than horizontal being observed. A statistical *t*-test was performed to study whether these differences were significant. The *t* value returned was 6.71 with an associated *p* value of 0.03, whilst the value of *t* required for rejection of the null hypothesis was 2.35. This means that there is a statistically significant difference between the two values, and points to there being a definite effect causing the DNA to take up a preferential orientation on the surface. This effect may be related to the hexagonal crystal lattice structure of mica. This may mean that the DNA with its associated counterions may preferentially bind in one particular orientation on the surface over another direction orthogonal to this. It may be easier for the ions to sit in the mica lattice sites in one position. However, the orientation of the mica crystal is not known in these experiments, so it is difficult to determine the directionality.

There are also fewer irregular molecules than ones with a predominant direction. This may have been down to the use of relatively short 800 bp DNA strands, as they will have a limited range over the conformations they can adopt. In the case where the disc was dried from above slightly more irregular molecules were seen, and could have been a consequence of this mode of drying. However, the differences are again relatively slight.

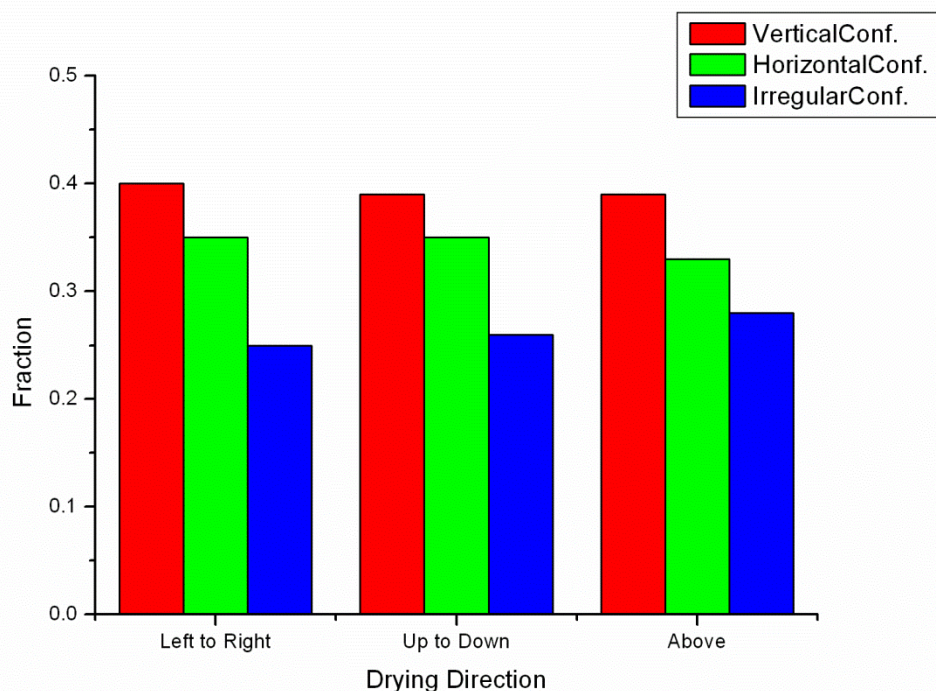


Figure 3-20. Graph showing the fraction of molecules adopting a particular orientation for each drying method. Drying direction does not appear to influence the conformation that the DNA takes up.

It may be possible to displace molecules from the surface if the drying process is too hard. To investigate the possibility of this, the total number of molecules in a number of scans was counted for each drying method. This value was then normalised to give the average number of molecules in a $1 \mu\text{m}^2$ area. The results are plotted in Figure 3-21. Within errors the number density for drying in the plane of the disc is quite similar, irrespective of the direction. The number density decreases when drying was carried out from above the disc, and could be a result of molecular transport towards the outer edge of the disc. The images used here were typically taken around the centre of the disc. The force of gas hitting the centre of the disc may have created a wave of fluid, which

could have carried molecules towards the edges of the disc, and explained the lower number density in the central region. However, it is also possible that these differences are caused by simple sample variability.

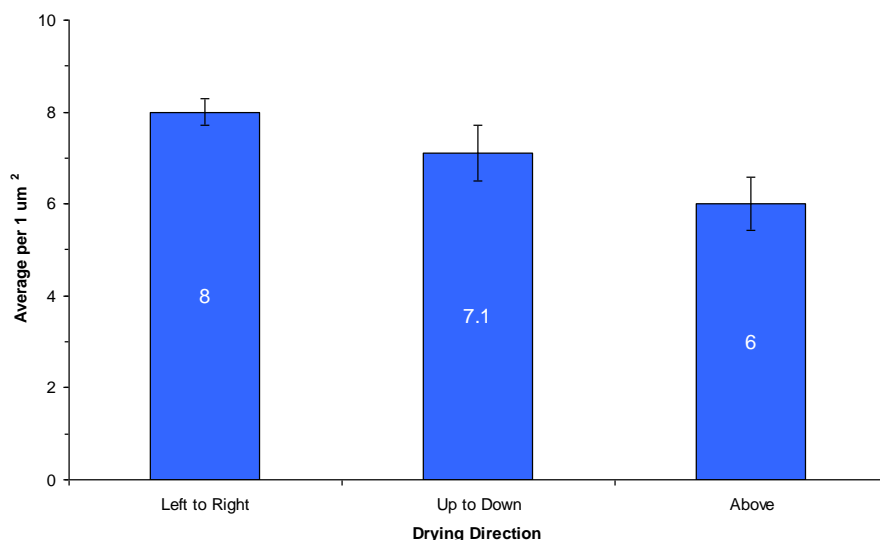


Figure 3-21. Graph showing number density of DNA molecules on the mica surface for the different drying directions. Drying in the same plane as the disc leads to similar results, while drying from directly above leads to a reduction in surface density.

Previous AFM studies have shown that DNA can be straightened and stretched out significantly, having a contour length much greater than expected [116, 198]. This was attributed to flow forces being exerted on the DNA during sample preparation. A major problem when imaging long strands of DNA with AFM is their tendency to become tangled on the surface. One method for aligning long DNA molecules on a surface uses the flow force generated by meniscus motion, otherwise known as molecular combing. Li *et al.* used gas flow to drive forward a drop of DNA solution on a bare mica surface [115]. By carefully controlling the direction of the flow and the speed of the moving interface they were able to align long DNA fragments on the surface. Our results show that the general drying process used in AFM sample preparation does not appear to affect the conformation of DNA. However, the fragments used in our measurements are much smaller than those used in the molecular combing experiment. It appears that the DNA-mica interactions are sufficiently strong to prevent re-aligning of the molecules studied here on the timescale of normal

drying, and the DNA molecules are trapped in a form that represents their conformation prior to drying.

3.5 Effect of surface treatment

In order to image DNA with AFM it is important that the molecule is strongly bound to the imaging substrate, such that the motion of the tip does not displace the molecules. Mica has an atomically flat, layered structure making this material extremely useful for use as an imaging substrate. This does pose a problem for effective DNA adsorption, as both mica and nucleic acids are negatively charged. However, the mica surface can be treated with solutions containing divalent cations. The cations reduce the mutual repulsion between DNA and mica, and give rise to a correlation force which brings the two materials into contact. Recent studies have shown that the binding of DNA to mica is strongly influenced by the absolute and relative amounts of monovalent and divalent cations [152, 153, 167, 194]. We investigated two different methods of DNA binding, both utilising divalent cations. The first method used Mg(II) ions in the imaging buffer as described earlier, and the second used the same buffer but with prior treatment of the substrate with Ni(II) ions. This was achieved by depositing 30 μ l of 2 mM NiCl₂ onto the substrate. After a period of 10 minutes the disc was rinsed and dried, and a DNA solution containing open-circular plasmids was deposited onto the surface as normal.

The varying protocols for sample preparation were seen to lead to differing molecular conformations (see Figure 3-22 for an example of imaging plasmids with both methods). Samples prepared only with Mg(II) in solution for binding took a much more open conformation and the molecules appeared to be spread out onto the surface. In contrast, the conformations of molecules seen using the same Mg(II) containing buffer, but with the additional Ni(II) preparation step are much more condensed on the surface. Each DNA chain crosses over itself a number of times. The variations in conformations between samples that were prepared with and without Ni(II) pre-treatment can be explained in terms of the differing adsorption strengths of DNA to the mica for each interaction. It

has been shown that strongly adsorbed molecules tend to be highly entangled, while loosely bound molecules appear more extended with few crossovers [154].

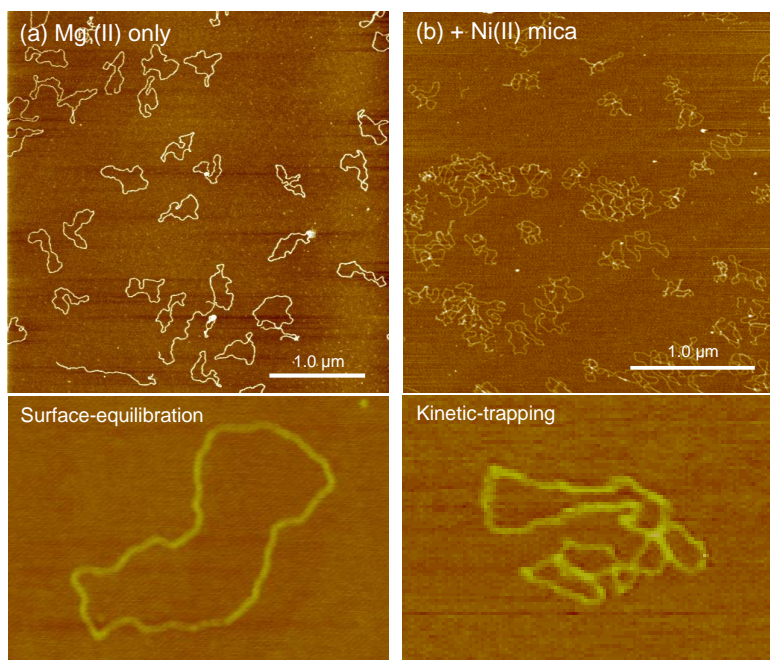


Figure 3-22. The two different sample preparations. (a) MgCl_2 prep. (b) Mg(II) with extra NiCl_2 prep. The differing binding mechanisms affect the molecular conformation of the DNA. When Mg(II) alone is used molecules take up an open, spread-out shape. DNA that is prepared with Ni(II) treatment contains a number of crossovers, despite the fact that this plasmid sample is relaxed open circular DNA. Software zooms of single molecules help to display these major differences more clearly (Images are $800 \text{ nm} \times 500 \text{ nm}$).

3.5.1 Dependence of equilibration on deposition time

We decided to investigate qualitatively how the general imaging buffer used (Tris-HCl (4 mM, pH 7.5) and 4 mM MgCl_2) affects the DNA conformation on the surface. The results above indicate that the use of Mg(II) in buffer to facilitate binding leads to molecules that interact relatively weakly with the surface. As such they are able to retain some mobility on the surface, and relax outwards to form relatively extended shapes in 2D. To study the degree to which DNA can diffuse laterally on the surface, samples were diluted into imaging buffer, and deposited onto the surface with varying deposition times, from a couple of seconds to 10 minutes. In this case two linear DNA fragments were

investigated: a short 800 bp fragment and a longer 4.2 kbp fragment. DNA has been shown to behave essentially as a negatively-charged polyelectrolyte with worm-like chain statistical mechanics [79]. As such polymer chain statistics can be used to study the type of binding. For linear DNA fragments the end-to-end distances of the chains were measured using the Nanoscope software straight line measurement tool and a value of the mean square distance calculated for each deposition time. The experimental values were compared to theoretical values for surface equilibrated molecules $\langle R^2 \rangle_{2D}$ using the following formula [167]:

$$\langle R^2 \rangle_{2D} = 4PL \left(1 - \frac{2P}{L} \left(1 - e^{-\frac{L}{2P}} \right) \right) \quad (3.1)$$

where it was assumed that the persistence length (P) was equal to 53 nm [27], and a base pair repeat of 0.33 nm was used to calculate the contour length (L). This model assumes that DNA behaves as a worm-like chain polymer, and is semi-flexible in nature with a defined persistence length.

Figure 3-23 shows experimentally measured values of mean square end-to-end distance for the short (red points) and long (black points) DNA fragments, prepared using increasing incubation times. Also marked on the plot are theoretical values of end-to-end distance for an ideal surface equilibrated case. These are shown as dashed lines of values 286,000 nm² for the 4.2 kbp fragment and 35,360 nm² for the 800 bp sized fragment.

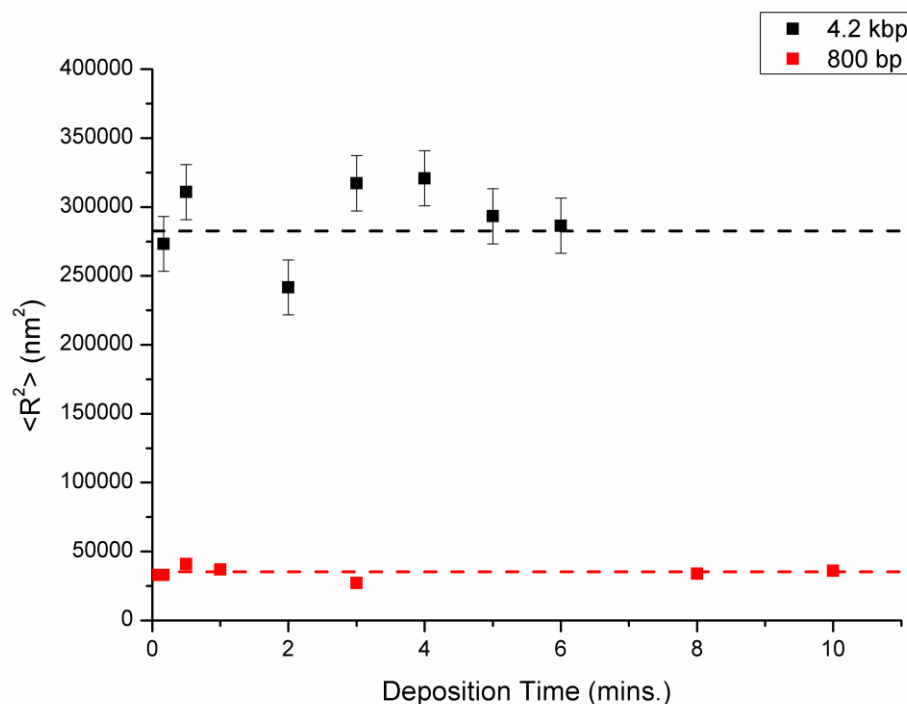


Figure 3-23. Plot showing how the experimentally measured average square end-to-end distance varies with deposition time for two differently sized fragments: 800 bp (red points) and 4.2 kbp (black points). The points all appear close to the theoretical values for surface-equilibrated molecules (dashed lines).

All the points for both samples lie relatively close to the theoretical values of $\langle R^2 \rangle$ for surface equilibrated molecules. Values of $\langle R^2 \rangle$ for a kinetically trapped case would be significantly lower, and allows us to confirm that samples prepared with Mg(II) alone in buffer represent an ensemble of molecules in a lowest energy conformation, existing in a 2-dimensional plane. At the concentrations used incubation time does not influence the average final conformation the molecules adopt. At even the longest length scale the entire DNA backbone can be observed easily. Hence, surface equilibration may prove a better mode of binding with which to study DNA-protein interactions.

If the molecules are weakly physisorbed, as is the case for surface equilibrated molecules, we would expect some degree of lateral diffusion on the surface to be possible prior to the sample being dried. Such diffusional motions are observed when DNA is imaged under bulk fluid [123]. The deposition process consists of two stages. Initially the molecules experience a long range interaction towards the surface, and begin to diffuse towards the surface, as we

observed previously in studies into deposition kinetics. Once at the surface they are free to equilibrate until drying captures them in a particular conformation.

In our experiments, the points of $\langle R^2 \rangle$ remain fairly close to the theoretical values. This is true for both short and long fragments. As such, there appears to be no correlation between deposition time and the conformation the DNA takes up on the surface. Even samples deposited for only a few seconds appear able to equilibrate fully. We observed some variation especially for the longer fragment, but these are likely down to simple sample variability. More significantly we did not observe any time dependence in such fluctuations.

If we were to imagine a situation whereby a DNA molecule arrives at the surface with a complex 3D conformation (as observed when a molecule is kinetically trapped), then we would expect it to spread out over a period of time, as different DNA segments repel each other, until it takes up a 2D equilibrated conformation. As such, we might expect to see a small end-to-end distance at very short deposition times, representing more of the molecules conformation in solution, and for such measurements to subsequently rise over time until the minimum energy conformation was reached. Additionally, we would expect such a transition to occur over a longer time period for a longer DNA fragment as it has a greater area to trace through in order to take up such conformations.

The observation that preparing DNA samples with a buffer containing Mg(II) ions, at a similar concentration and temperature described here, leads to surface equilibrated conformation is supported by work by Rivetti *et al.* [167]. However, when they imaged a 5994 bp DNA fragment, deposited with different deposition times ranging from one to ten minutes, in all cases the measured $\langle R^2 \rangle$ was larger than the theoretical values. Additionally the measured experimental values of end-to-end distance were observed to increase over time, leading them to suggest that the equilibration time for long polymers is of the order of minutes. Shorter fragments were observed to equilibrate in less than one minute. They suggested that excluded volume effects resulting from interactions among different molecules, or among segments of the same molecule (self-avoidance), becomes increasingly important for longer

molecules, meaning that $\langle R^2 \rangle$ values are altered and the equilibration time increased.

We observed the onset of equilibration behaviour much earlier than the previous study, even with large fragments of the order of thousands of base pairs. The fact that we did not observe a strong time dependence on molecular conformation may mean that either the equilibration time is quicker than previously observed, or that the molecule is able to take up an equilibrated conformation in solution sometime before it actually binds to the surface. This case is supported by the work of Sushko *et al* [194]. They present a simple model for the adsorption of DNA onto mica, in which if the adsorption of the molecule is governed by a long-range attraction; as soon as the DNA experiences such a force then the segments of the molecule will gradually adjust their positions to adopt the same equilibrium distance to the surface. Hence, by the time the molecule touches the surface it has already adopted a 2D conformation. The relaxation time of the molecule once it makes contact with the surface may also be fast at all DNA length scales.

They did also say that the longer the molecule, the slower this process would be as the molecule has a large number of degrees of freedom and complicated excluded volume effects [194]. If the deposition time is too short then the molecule would only be partially adsorbed to the surface with the rest of the molecule “sticking” out and forming loops. It appears however that at the length scales we investigated here that equilibration can be achieved for even the shortest deposition times.

3.5.2 Effect of fragment length on deposition

Once it became apparent that surface deposition of DNA in a Mg(II) containing buffer leads to an acceptable number of bound molecules in a surface equilibrated conformation, we aimed to investigate further how pre-treatment with Ni(II) affects molecular conformation. In particular, we aim to study the other extreme of DNA binding, namely kinetic trapping.

In addition we aimed to investigate any possible length dependencies of the DNA binding mechanism, by utilising different divalent cations to achieve the two extremes of binding: surface equilibration and kinetic trapping. Our goal was to see whether a small DNA fragment experiences a similar interaction with the surface as a larger fragment, and whether it could tell us anything about how the divalent cations help form complexes with the surface. To this end we utilised a range of different sized linear DNA fragments, some prepared by PCR and others obtained through direct enzymatic digestion of plasmid DNA. The fragments ranged in size from quite small 200 bp fragments to much larger 4.2 kbp sized samples. All samples were linear, and diluted into standard imaging buffer (Tris-HCl (4 mM, pH 7.5) and 4 mM MgCl₂).

This study called for each sample to be examined with and without pre-treatment of the mica substrate with Ni(II) ions. This pre-treatment was carried out by transferring 30 µl of 2 mM NiCl₂ onto the surface. A period of 15 minutes elapsed before the surface was rinsed with milliQ water to remove residual buffer salts, and dried with a weak flux of nitrogen gas.

All samples were prepared in an identical way for AFM. DNA samples were prepared for AFM by depositing 10 µl of 30 fmol DNA solution onto freshly cleaved muscovite mica and incubating for 3 minutes at room temperature. This incubation time should be sufficient enough for the DNA to take up its favoured conformation on the surface. After this, the samples were rinsed in milliQ water, before being dried in a weak flux of nitrogen gas.

Two different methods were used to attempt to bind DNA to mica substrates: (1) Mg(II) ions were included in the incubation buffer from which DNA was deposited on to otherwise untreated ("bare") mica; (2) pre-treatment of the mica with a Ni(II) solution followed by incubation in Mg(II) containing buffer as described above. The former method results in the appearance of DNA molecules with characteristic conformations related to surface equilibration in the 2D plane of the mica, while the latter method is expected to result in kinetically-trapped molecules. [153, 167]. Each fragment size was prepared

with and without Ni(II) pre-treatment and a number of scans were taken of each case.

Qualitatively the two protocols for sample preparation led to differing molecular conformations. These were similar to the plasmids imaged previously (Figure 3-22). Samples prepared with Mg(II) in solution took a much more spread-out conformation. In contrast, the conformations of molecules seen after Ni(II) preparation were much more condensed and occupy a smaller area. Especially evident for the longest DNA fragments are that each DNA chain crosses over itself a number of times. Figure 3-24 shows typical images for each fragment, with binding facilitated by both cations.

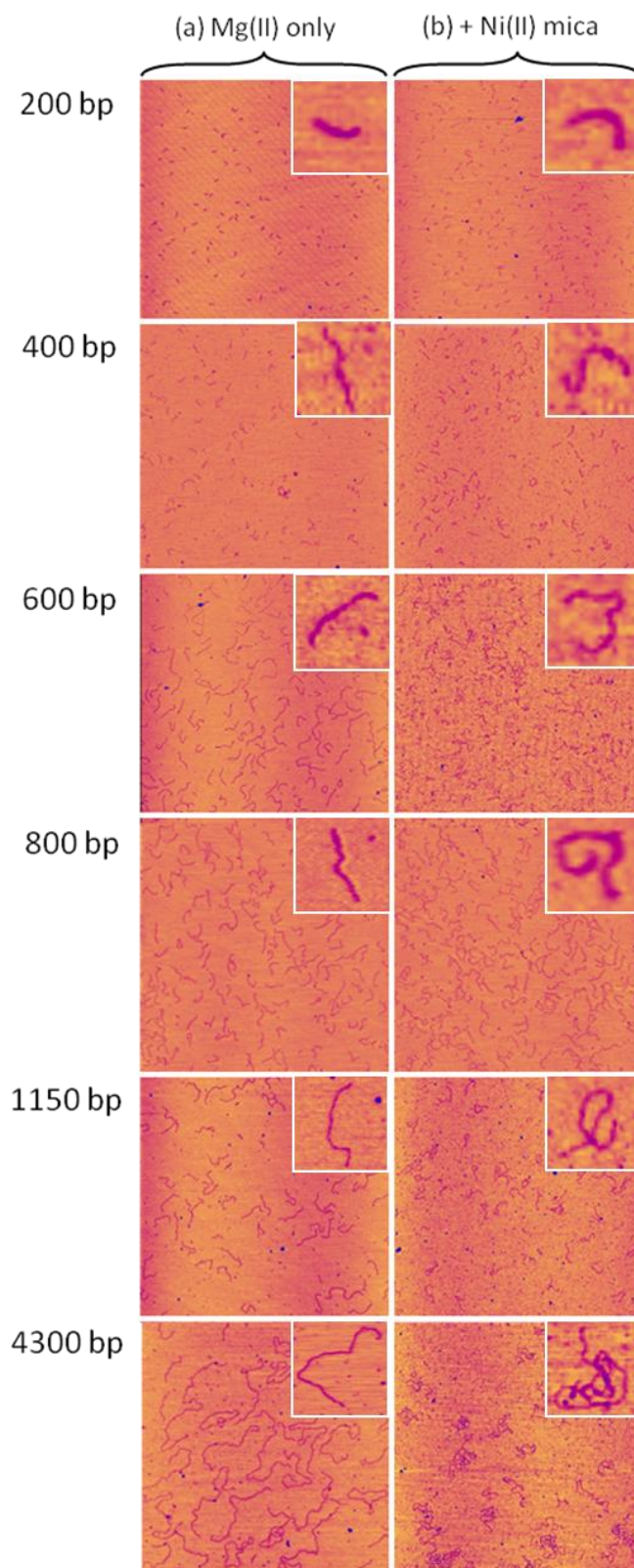


Figure 3-24. Typical AFM images of each fragment prepared with (a) Mg(II) only on the left-hand side, and (b) additional Ni(II) pre-treatment of mica on the right-hand side. Each image is 3 μm × 3 μm. An example of each length at each preparation method is shown in the inset of each image.

To get a more accurate picture of the binding mechanisms at play end-to-end distances measurements were taken. Polymer chain statistics can help to elucidate how the DNA conformation relates to how the molecule binds to the surface. Rivetti demonstrated that the average square end-to-end distance for each binding mechanism obeys different statistical models, and has a typical value for a given fragment size (given by Equation 3.1 for the surface equilibration case) [167]. In this case we also aim to investigate whether kinetic trapping of molecules occurs. In which case the average square end-to-end distance is given by,

$$\langle R^2 \rangle_{3D} = 2PL \left(1 - \frac{P}{L} \left(1 - e^{-\frac{L}{P}} \right) \right) \quad (3.2)$$

On average, one would expect a fragment that has a certain degree of mobility, as in the surface equilibrated case, to have a larger end-to-end distance on average than a molecule that is pinned at one point and unable to diffuse laterally, as occurs during kinetic trapping.

Table 3-1 and Table 3-2 shows the average end-to-end square distances for each fragment prepared via the two different binding mechanisms, together with theoretical values relating to surface equilibrated and kinetically trapped binding cases. Immediately it is noticeable how much smaller the value is for the Ni(II) pre-treated samples as opposed to those bound to bare mica. This demonstrates the stronger attractive forces that Ni(II) pre-treatment provides. When looking at the experimental values it becomes clear that for the DNA prepared with Mg(II) alone there is a general agreement with the theoretical values for DNA of the same lengths, which are surface equilibrated. These are plotted in Figure 3-25, together with a curve representing the theoretical case of surface equilibration. The value for the largest fragment is slightly high, but still within one standard deviation of the theoretical value. The discrepancy was most likely caused by this large fragment exhibiting the greatest variation in end-to-end distance, a variation that is magnified when the value is squared.

Table 3-1 Experimental values of the mean square end-to-end distance for samples prepared using Mg(II) only. These are compared to the theoretical case of surface equilibrated molecules.

Fragment Size (bp)	$\langle R^2 \rangle$ Experimental: Mg(II) only (nm ²)	$\langle R^2 \rangle_{2D}$ Equilibrated theoretical (nm ²)	Standard error (nm ²)	n
200	3,050	3,577	60	186
400	11,200	11,981	300	150
600	23,000	22,975	700	130
800	36,000	35,358	1,000	200
1149	55,000	58,608	2,000	219
4300	310,000	278,356	20,000	215

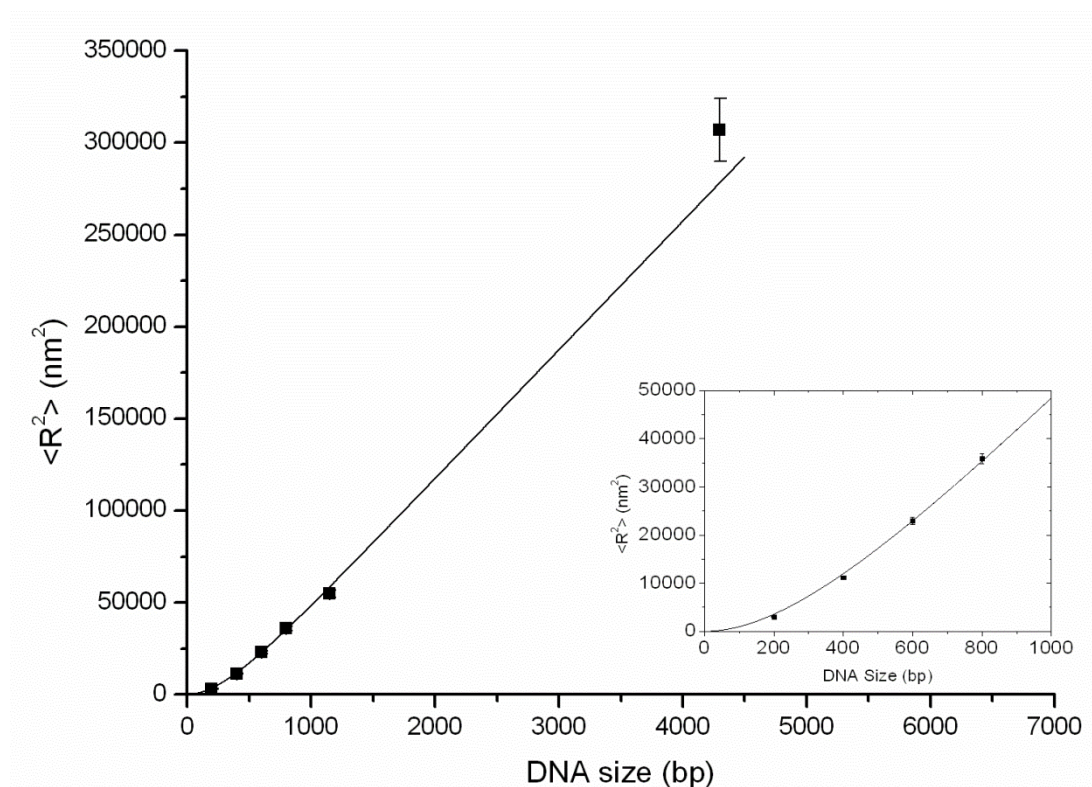


Figure 3-25. Experimental values of square end-to-end distances (points) of samples prepared with Mg(II) alone, plotted with a curve of the theoretical values for surface equilibration. The inset shows a close-up of the region from 0-1000 bp where most of the measurements were clustered, and the curve displayed the greatest rate of change.

Figure 3-26 shows the end-to-end squared values of samples prepared with Ni(II) pre-treatment of mica plotted with a binding curve representing the theoretical values of kinetically trapped molecules. For the samples prepared with Ni(II), the larger fragments show very good agreement with the theoretical values typical of kinetically-trapping mediated binding. However, below 600 bp the measured values show significant deviation from the theoretical case. The inset of Figure 3-26 shows a closer view of the measurements for the smaller fragment sizes, where the deviations from the theoretical case are observed. The discrepancies between experimental and theoretical cases actually increase as the fragment under investigation becomes smaller. For example, the 400 bp displayed a 0.25 fractional deviation from the expected value, whilst this value increased to 0.45 for the 200 bp fragment. As the smallest fragments contained the smallest spread of data this discrepancy is unlikely to have been caused by natural sample variability, but rather points to a physical effect relating to the mica surface and its associated counter-ions.

Table 3-2. Experimental values of the mean square end-to-end distance for samples prepared using an additional Ni(II) pre-treatment step. These are compared to the theoretical case of kinetically-trapped molecules.

Fragment Size (bp)	$\langle R^2 \rangle$ Experimental: + Ni (II) (nm ²)	$\langle R^2 \rangle_{3D}$ Trapped theoretical (nm ²)	Standard error (nm ²)	n
200	3,640	1,997	60	195
400	7,900	5,893	300	202
600	11,200	10,336	600	195
800	15,000	14,936	700	240
1149	23,000	23,076	2,000	157
4300	96,000	99,431	10,500	74

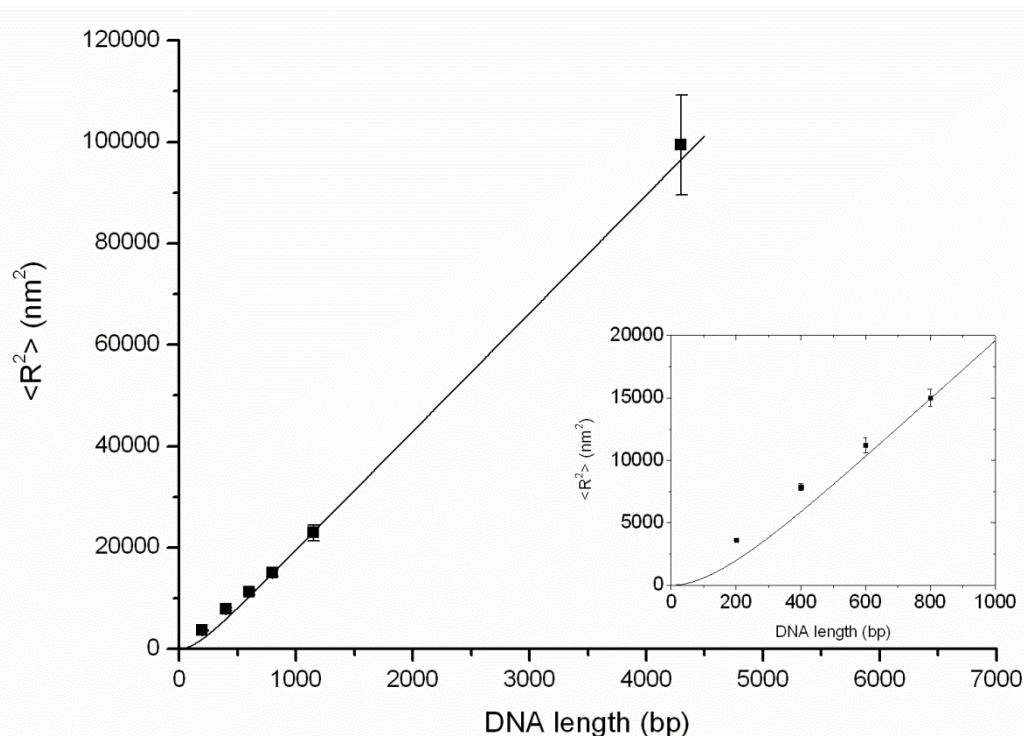


Figure 3-26. Experimental values of square end-to-end distances (points), of samples prepared with Ni(II) pre-treatment of mica, plotted with a curve of the theoretical values for kinetic trapping. The inset shows a close-up of the region from 0-1000 bp. Below a fragment size of 800 bp the points begin to deviate from the expected values. The magnitude of the difference increases toward smaller fragment size.

3.5.3 Discussion

AFM images of identical fragments prepared with and without Ni(II) pre-treatment of mica appeared qualitatively different. The differences between each preparation method appeared more marked at higher fragment sizes, when numerous chain crossings were observed under Ni(II) treatment. The variations in initial conformations between samples that were prepared using Ni(II) pre-treatment and those that simply contained Mg(II) in the deposition buffer, can be explained in terms of the differing adsorption strengths, a result of separate adsorption interactions occurring. Strongly adsorbed molecules tend to be highly entangled, while loosely bound molecules appear more extended with few crossovers [154].

To explain these major differences in conformation between Mg(II) and Ni(II) samples we must consider the mechanisms of binding. Effective binding of DNA is shown to be possible only by introducing positively charged divalent ions into the system [84, 197, 205]. The cations are able to form a diffuse atmosphere of counter-ions close to a surface, which act to partially neutralise the negatively charged phosphate backbone of the DNA, and the mica surface. Correlations between the counter-ion clouds can lead to the situation where a net attraction exists between the DNA and mica, promoting the adsorption of the molecule [152, 153].

It has also been shown that different cations allow for varying degrees of DNA surface binding, with transition metal ions being the most effective [84]. In particular, it appears that Ni(II) were among the best candidates among the transition metal ions, and is effective at balancing the surface charge of the mica and DNA surfaces [2, 189]. These factors act to increase the correlation force, and bring the DNA together with mica. On the other hand Mg(II) ions tend to bind non-specifically with DNA [59], leading to weaker interactions.

These interactions are eventually responsible for the conformation that the molecule takes up on the surface. In the binding model of Sushko *et al.* [194] if the adsorption of the molecule is governed by long-range forces, the molecules

form a 2D conformation on the surface. The adsorption is due to non-specific van der Waals interactions; the molecule remains relatively mobile and is not restricted in its lateral movement. However, if short-range forces are responsible for the adsorption the molecule adopts a projected 3D \rightarrow 2D conformation on the surface (kinetic-trapping). At distances shorter than the Debye length, the short-range ion-correlation force is significant. Below this length the DNA is able to “see” the negatively charged mica. In order for the molecule to adopt a projected conformation, the surface-charge density of the substrate should be high, so as to ensure significant attractive ion-correlation forces.

We observed direct evidence for each interaction regime depending on the preparation method. In the first case (Mg(II)) a long-range weak interaction is responsible. The DNA has time to attach to the surface and relax and spread-out. In the second instance (Ni(II)), a short-range interaction is responsible and strongly pins the molecules to the surface. Once the molecules make contact with the surface they are frozen in place, and as such the flat conformation on the mica is a 2D projection of the molecules 3D configuration in solution. These were confirmed by comparing experimentally measured values of the average end-to-end distance with theoretical values for each extreme of surface binding. The differences are easiest to observe for the longer fragments, and are shown side-by-side in Figure 3-27.

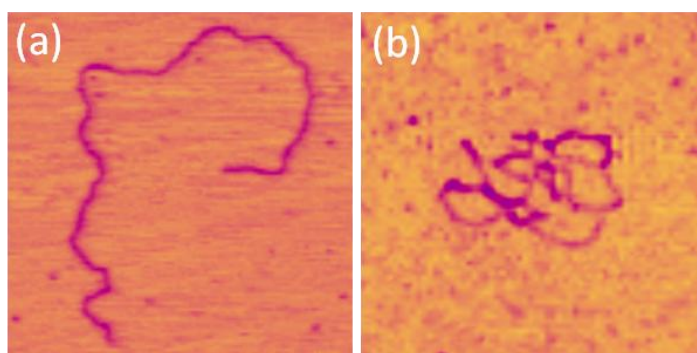


Figure 3-27. Software zooms of two linear 4.2 kbp fragments prepared with (a) Mg(II) only and (b) additional pre-treatment of mica with Ni(II). These demonstrate the differences between surface equilibrated and kinetically trapped molecules. The former takes up a spread-out conformation, whilst the latter is pinned at one location, and contains numerous chain crossovers. Images are 800 nm \times 800 nm.

The fact that every $\langle R^2 \rangle$ point for samples prepared with Mg(II) in buffer line up very well with the theoretical binding curve points to the fact that there is a homogeneous distribution of magnesium ions on the surface and the DNA. As such, all molecules behave similarly and the ensemble average of $\langle R^2 \rangle$ tends to the theoretical value, representing surface equilibrated molecules. There is no length dependence, and small fragments are able to equilibrate in an identical way to large chains.

However, the situation changes somewhat with the additional step of Ni(II) pre-treatment for mica. Here the points show an increasing deviation from an ideal kinetically trapped case, below a certain DNA size threshold. This behavioural change occurs somewhere between 800 bp and 600 bp, after which the DNA exists in an intermediate conformation that is between kinetically trapped and surface equilibrated. As this effect is dependent on chain length it appears to suggest that there are certain inhomogeneities in the Ni(II) treated surface, of some characteristic size and related the structure of mica.

Mica is a silicate based mineral which has a layered structure. One of its main characteristics is the perfect cleavage of the basal layers. The ability to cleave mica means that it is perfect for use as an AFM substrate, providing a fresh clean surface each time an experiment is to be performed. The layers in mica are bound together through fairly weak interactions mediated by K(I) ions. These are able to fit into small recesses in the mica surface. It has been suggested that when the surface of mica becomes hydrated some K(I) ions can be released into solution, leaving the surface negatively charged [150]. The vacancies left behind by K(I) can be filled by ions of similar size, that are present in the solution [84, 147]. In the case of surface pre-treatment attempted here, this will be Ni(II) present in the NiCl₂ solution. If all the vacant lattice sites were able to be filled, one would expect to have a surface with a uniform surface density of Ni(II). The cleaving of the top mica layer is expected to remove roughly half of the K(I) ions on the uppermost cleaved layer, leaving the freshly clean surface half occupied [139]. If these ions were able to remain held in their sites once NiCl₂ is added to the surface, then the Ni(II) ions would only be able to interact with sites that are unoccupied, in which case the surface would

become an patchwork of the two different ions. The space between the Ni(II) domains will be occupied by ions of K(I), H(I) or Mg(II) in a dynamic equilibrium, the exact composition of which is difficult to determine. This composition, however, will be the same as that where the mica has not been pre-treated with Ni(II).

A patchy mica surface could go some way to explaining the deviation from perfectly kinetically trapped behaviour seen at small length scales. A reduction in the Ni(II) ion density would reduce the effective charge of the surface, diminishing the ion correlation forces that lead to kinetic trapping.

To only see such behaviour below a certain transition point in DNA length would appear to suggest that any patches of ions would have to be relatively small in size. As the change in ideal behaviour is seen below chain lengths of roughly 267 nm (800 bp) then the size of such patches must be lower than this. Additionally, the fact that the deviation from the expected values of $\langle R^2 \rangle$ increases as the fragment size decreases, may suggest that the size of the Ni(II) patches begins to approach that of the DNA chain length. If we imagine a simple situation whereby half of the Ni(II) treated surface is occupied by divalent cations and that they form a simple polka dot pattern, then a long molecule will easily spread across a number of Ni(II) patches. A chain of intermediate length may be partially on a Ni(II) site and partially on an unoccupied site. A very small fragment of a size approaching that of the patches may be either directly over or completely away from the divalent cation sites. See Figure 3-28 for a schematic diagram explaining these effects.

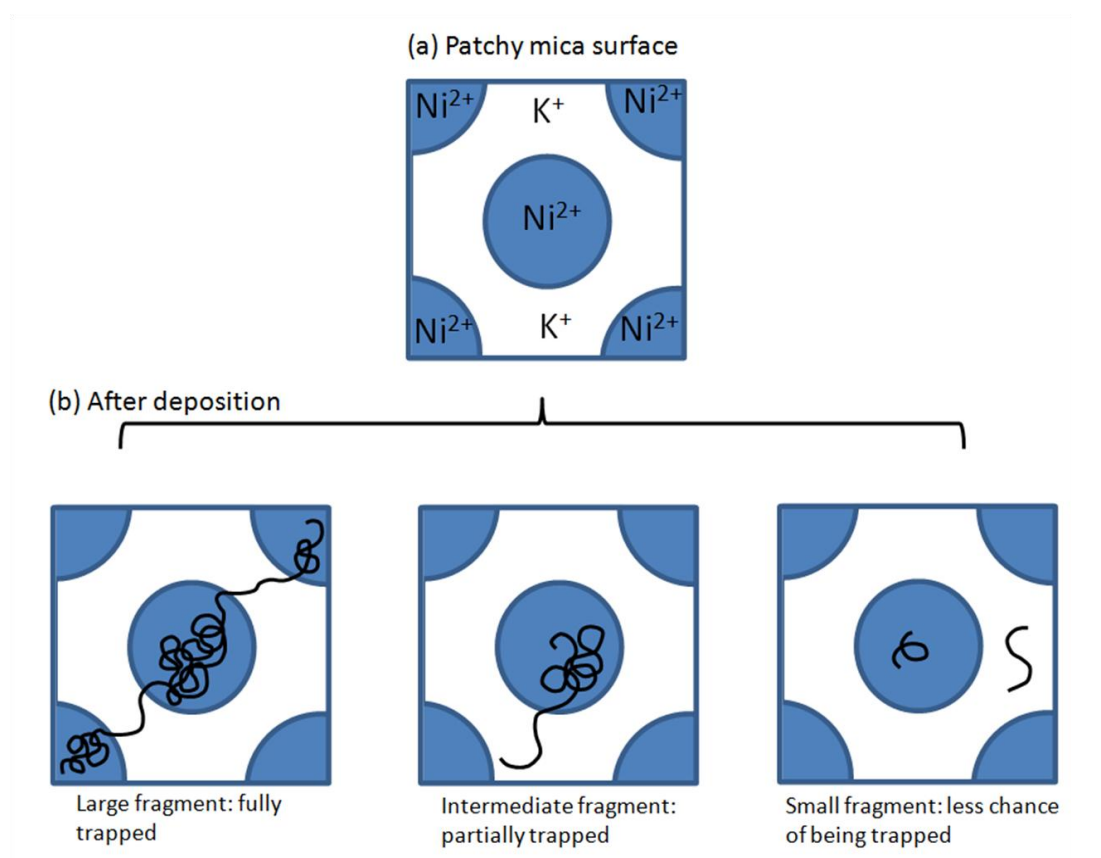


Figure 3-28. (a) Schematic diagram of an idealised possible structure of the ion-exchanged mica surface, containing regions of K(I) and patches of Ni(II). (b) Representations of how different DNA fragments may appear. Large fragments can bind easily across more than one region of Ni(II), meaning they are fully trapped. As the fragments become smaller there is less of the chance of a molecule interacting with a Ni(II) region. As such, there are less fully trapped molecules and a corresponding increase in end-to-end distance.

Due to the presence of Mg(II) in the deposition buffers for both sample preparation methods, it is expected that a molecule coming down on an area entirely uninhabited by Ni(II) ions would not experience the large attractive force at a short-range from the surface, that is necessary for trapping. The attraction would be of the weak, long-range kind that would lead to surface equilibration once the molecule made contact with the surface. As such the end-to-end distances would be greater than expected.

It appears that for longer molecules, the cooperative effect of different segments interacting with perhaps multiple Ni(II) patches is enough for the molecule to be kinetically trapped. This is a consequence of the fact that the length scale of

heterogeneity on the mica is very small compared to the molecular size. As the molecule becomes smaller the chances of strongly interacting with a Ni(II) patch, thus facilitating strong binding, becomes less likely. This is the most likely explanation for the fractional deviation from the expected behaviour increasing for smaller DNA fragments.

We propose that exchanging the freshly cleaved surface of K(I)-mica with divalent transition metal ions leads to a patchy charged surface, where domains of K(I) and the exchanging ion are present in phase separated domains. This is based on the observation of the behaviour of DNA molecules in the AFM under Mg(II) containing buffer, absorbed to Ni-mica (i.e. muscovite mica pre-treated with a NiCl₂ solution). As the length of the DNA decreases, the deviation away from kinetic trapping towards conformations indicating 2D equilibration becomes more marked, indicating that the hypothesis of an ion-exchanged mica with localised patches of Ni(II) ions is a realistic one.

3.6 Conclusions

The work presented in this chapter centred on preparing high quality DNA samples, by systematically studying the different aspects of AFM sample preparation. A range of different DNA fragments were purified and used as test samples. Samples were prepared for imaging using the sessile drop method and the deposition and binding onto a mica substrate was investigated. During which it was confirmed that the transport of molecules was governed by a diffusion mechanism.

The excess DNA solution is rinsed off using water to minimise the amount of salt precipitation on the mica surface. It was found that increasing the rinse volume could displace weakly bound molecules from the surface. In practice, it is impossible to eliminate all salt formation on the mica, therefore other ways of cleaning away all salt and yet retaining the DNA are desirable. Since alcohols precipitate DNA, we investigated of the effect of alcohol rinses in substitution for a water rinse. Rinsing with alcohol at present gives no clear benefits, and at higher concentrations led to the DNA taking up irregular structures. Such an

approach, however, may be beneficial for high resolution imaging using low concentration alcohol washes, where the conformation of the DNA is relatively unaffected.

Samples are dried in a weak flux of nitrogen gas. The direction of drying relative to the sample, was found to have no significant influence on molecular conformation of the sample studied.

Mica was used as a substrate for the DNA, and was treated with a solution of positive ions, typically Ni(II), thus allowing binding of the molecules from an Mg(II) rich solution. Non Ni(II) treated mica led to the surface equilibration of DNA molecules. The chains spread out and adopted a minimum energy conformation, making it easy to observe the entire length of molecule. This may be more desirable than the surface treatment with nickel ions, which gave kinetically trapped molecules with many crossings. This surface bound morphology in kinetic trapping will make accurate determinations of protein position on DNA templates more difficult. Additionally, for longer DNA fragments, which are more prone to being sheared in solution and broken down, it may be hard to confirm that the molecule is of the expected size, i.e. it may be difficult to take contour length measurements if the molecule is severely contorted.

Through the binding behaviour, and final conformations of DNA molecules on mica surfaces it was suggested that muscovite mica ion-exchanged with transition metal ions, such as Ni(II), can be heterogeneous. The size of the patches is on the nanoscale and influences the binding of the DNA at short length scales (< 800bp).

Chapter 4

4 Humidity controlled AFM of DNA

4.1 Introduction

Imaging of biomolecular systems *in vitro* by atomic force microscopy (AFM) has been on-going for almost two decades now. The success of imaging isolated, weakly physisorbed molecules deposited on supporting surfaces (e.g. mica), has been achieved routinely following the development of dynamic modes of AFM [19, 21, 196, 220]. The intermittent contact of the AFM tip normal to the sample surface effectively eliminates shearing interactions present in contact mode, where the tip constantly tracks over the surface. Biomolecules such as DNA and protein, can apparently withstand compressive forces more readily than shearing interactions [221], which is not surprising for visco-elastic systems with large water content. For structural studies of biomolecules, imaging is often performed in air, where problems of molecular motion in liquid conditions are obviated. On hydrophilic surfaces, such as mica, thin water layers exist that allow surface-bound molecules to retain a degree of hydration [68, 165, 173]. Surface force apparatus measurements and AFM force measurements demonstrate that the amount of water present on the surface is correlated with the environmental humidity [17, 40, 42, 93, 107, 122, 212].

Almost all biomolecular imaging in air has been performed under ambient conditions of laboratory humidity, which is usually between 30 and 40 % relative humidity (RH). Following the introduction of intermittent contact mode AM AFM, AFM imaging in contact mode was largely abandoned. This was despite reports that DNA could be reliably imaged in contact mode provided that the capillary neck interaction was removed by using dry conditions [66, 199, 205]. However, even under these dry conditions, DNA contour length measurements indicated that double-stranded (ds) DNA probably retained a B-form unless the surface was also cooled [69]. It has also been established that

capillary neck interactions may influence AM AFM measurements, depending upon the AFM operational parameters, due to changes in the size of the capillary neck [18, 19, 42, 122, 203, 212, 222]. The outcomes of varying the RH while imaging biomolecular systems are largely unexplored to date.

In this chapter, we perform a systematic study on the effects of changing the relative humidity in the imaging chamber of the AFM while scanning over dsDNA physisorbed to mica. DNA on a mica surface behaves essentially as a negatively-charged polyelectrolyte with worm-like chain statistical mechanics [79]. Recent studies have shown that the binding of DNA to mica is strongly influenced by the absolute and relative amounts of monovalent and divalent cations [152, 153, 167, 194]. Here, we investigate cation mediated binding further. A range of different DNA test samples were imaged using two methods of binding (1. Using Mg(II) in buffer, 2. With additional Ni(II) pretreatment), with the aims of investigating the effect of imaging at high humidity in air (Section 4.2).

4.2 Samples and preparation

A number of DNA fragments were used as test molecules to try to understand the behaviour of double-stranded DNA (dsDNA) bound to mica under conditions of varying environmental humidity. Four types of DNA molecules were chosen to compare certain physical characteristics: (i) Length of linear molecules. (ii) Effect of topology between linear and circular forms. (iii) Effect of introducing strain (i.e. supercoiling) in circular forms. The samples used were as follows: (1) Short Linear 800 bp (SL) (2) Long Linearised pBR322 (4.2 Kbp) (LL) (3) Open circular pBR322 (4.2 Kbp) (OC) (4) Supercoiled pBR322 (4.2 Kbp) (SC). After purification, all DNA samples were diluted ten-fold into imaging buffer containing 4 mM Tris-HCl (pH 7.5) and 4mM MgCl₂.

Two different methods were used to attempt to bind DNA to mica substrates: (1) Mg(II) ions were included in the incubation buffer from which DNA was

deposited on to otherwise untreated (“bare”) mica; (2) pre-treatment of the mica with a Ni(II) solution followed by incubation in Mg(II) containing buffer as described previously. It is well-established that the former method results in the appearance of DNA molecules with characteristic conformations related to surface equilibration in the 2D plane of the mica, while the latter method is expected to result in kinetically-trapped molecules, where the conformation on the surface is more akin to a 2D projection of the 3D conformation in solution [153, 167]. We demonstrated both extremes of binding previously in Chapter 3, but in this section we would like to pay particular attention to how each scenario responds to conditions of increased humidity.

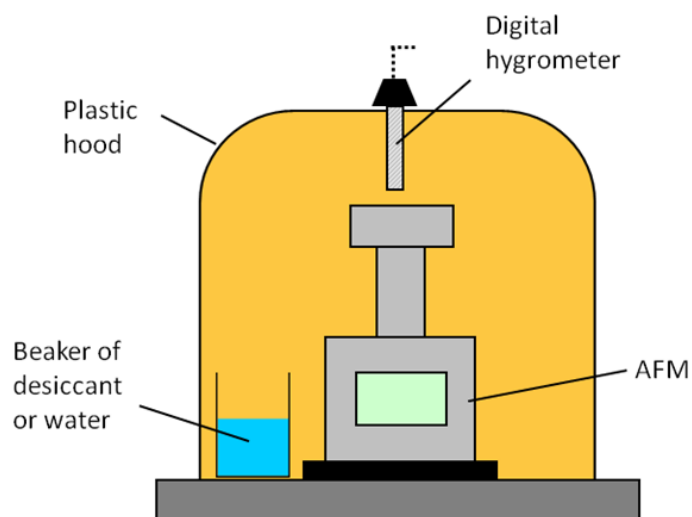


Figure 4-1. Experimental set-up used for controlled-humidity imaging. The AFM was placed inside a hood to isolate it from the laboratory environment. The humidity was raised and lowered by the addition of a beaker of water or desiccant respectively. The humidity inside the hood was determined through the use a digital hygrometer.

The relative humidity was controlled by placing the AFM inside a low vacuum plastic hood (Nalgene, Rochester, NY), allowing complete isolation from the surrounding laboratory environment (Figure 4-1). Low humidity was achieved by placing a beaker of desiccant (self-indicating silica gel) into the hood, while high humidity was obtained by replacing the desiccant with a beaker of water. To change the humidity in the chamber the volume and area of the water in one or more beakers were altered accordingly. The experiments were conducted while the RH was continuously rising or falling. During one image acquisition

time the RH would not change by more than 1%. A digital hygrometer, inserted into the plastic chamber, was used to measure the relative humidity and temperature. The temperature remained constant in the chamber at 23.5 ± 0.5 °C and the quoted instrumental error for RH was $\pm 3\%$ RH. Images were obtained in series from low to high relative humidity. Care was taken to always image the same area using the same scan and frame direction, in order to limit the effects of instrumental drift. In between taking images, the cantilever was withdrawn around 20 μm from the surface and the humidity was altered, and left to reach a stable value before another image was taken.

4.3 Results

4.3.1 Linear fragments

First we will consider the effects of humidity on imaging the linear fragments. It was possible to resolve complete molecules for both the long (4.2 kbp) (LL) and short (800 bp) (SL) DNA fragments even when the humidity was close to 100 % RH(see Figure 4-4 and Figure 4-6). This was true for both classes of sample preparation (with and without Ni(II)). In addition to this, image contrast remained good throughout imaging, and no significant change in either height or width of the molecules was observed during the image sequence (see Figure 4-2 and Figure 4-3). The average height of the dsDNA was 0.40 ± 0.05 nm and the average width 25.6 ± 0.7 nm.

This height is smaller than the expected diameter of B-form DNA (2.0 nm) but it is typical for AM AFM measurements in air [137]. The reasons behind this underestimation by AFM at a single molecule level could be manifold including sample deformation [29] and/or salt deposition [197], or a consequence of the local probe-sample geometry[175]. Santos *et al.* modelled the tip-surface-sample interaction and found that once the size of a feature becomes smaller than the effective area of interaction between tip and sample, the AFM cannot differentiate between interactions resulting from the substrate and those from the sample. As a result height information is spread out laterally across the tip-sample interaction area resulting in the loss of height for features of a nanoscale

dimension such as DNA [175]. Large height losses occur even when there is no sample deformation, and the true height cannot be obtained directly.

An increase in DNA width may have implied an increase in tip-force and increased sample deformation, as observed in early contact mode AFM studies of DNA at ambient humidity [205]. The absence of these effects at high RH, coupled with the overall good resolution help to illustrate the benefits of AM AFM over contact mode. In this case, it appears the effect of capillary forces is diminished, a result of the reduced tip-sample contact time. The higher than expected widths of DNA is a result of the fact that AFM tips are not infinitely sharp, having a typical radius of curvature of around 10 nm, meaning that the final image becomes a convolution of the actual surface topography and the shape of the tip. The tip convolution effect depends on the geometry of the objects imaged, and leads to the overestimation of the lateral dimensions of three-dimensional objects. In this case the side wall of the tip interacts with DNA first rather than the apex, after which the system acts to raise the height of the tip to follow the topography, leading to a broadening of the feature. Both the height and width of features may be improved by using sharper tips.

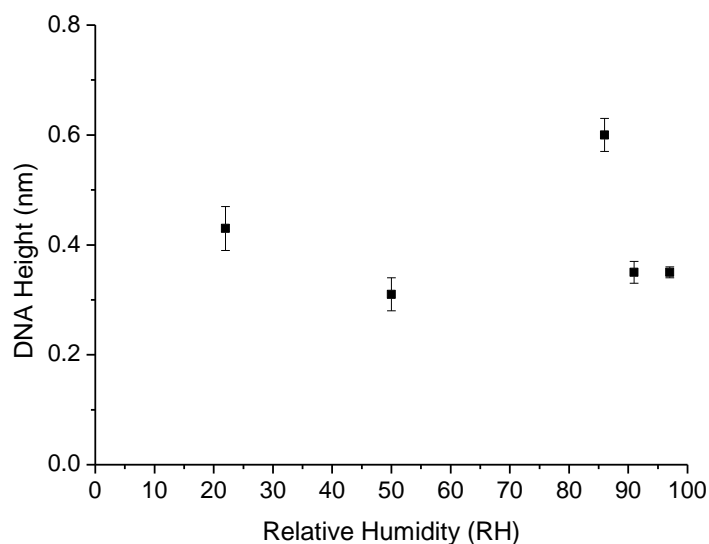


Figure 4-2. Average DNA height versus relative humidity (RH) for individual double-stranded long linear (LL) DNA helices lying against the mica surface (n=50 in each point). There is no apparent correlation between measured height and the RH when imaging by repulsive force regime AM AFM.

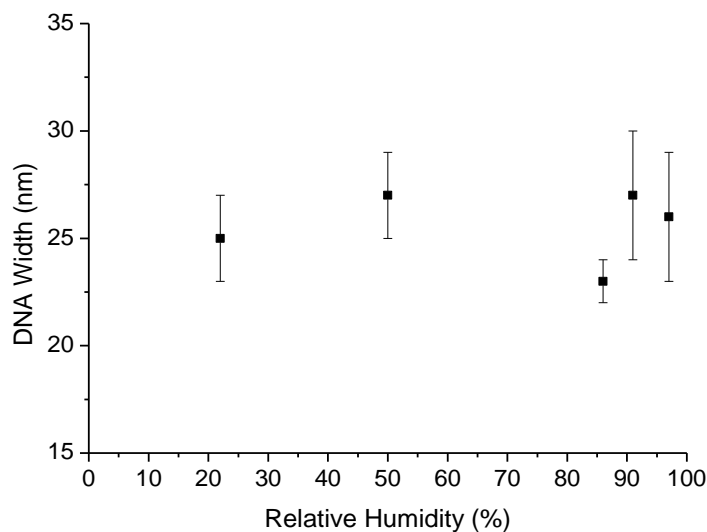


Figure 4-3. Average DNA width versus relative humidity (RH) for individual double-stranded long linear (LL) DNA helices lying against the mica surface (n=50 for each point). There is no apparent correlation between measured width (full width at half-height) and the RH when imaging by repulsive force regime AM AFM.

The varying protocols for sample preparation led to differing molecular conformations. Linear DNA prepared only with Mg(II) in solution for binding took a more open conformation and the molecules were equilibrated onto the surface. In contrast, the conformations of molecules seen using the same Mg(II) containing buffer, but with the additional Ni(II) preparation step are much more condensed on the surface. Especially prevalent for LL is that each DNA chain crosses over itself a number of times (Figure 4-4b). These observations were consistent with binding interactions according to the models of surface equilibration and kinetic trapping respectively. For linear chains, it is possible to ascertain how the molecule behaves during deposition by measuring the end-to-end distances and comparing them to theoretical values for equilibrated $\langle R^2 \rangle_{2D}$ and trapped $\langle R^2 \rangle_{3D}$ cases (see Table 4-1 for a summary of theoretical and experimental values).

As expected the samples prepared with Mg(II) alone had shorter end-to-end distances. For the short linear (SL) fragment the mean square end-to-end distance was $35,000 \pm 2000 \text{ nm}^2$, which compares well to the theoretical value for surface equilibrated molecules ($35,360 \text{ nm}^2$). For the long linear (LL)

molecule, it was $250,000 \pm 40,000 \text{ nm}^2$, which, within errors, agrees with the theoretical value for equilibration of $282,000 \text{ nm}^2$. With regard to samples prepared with additional Ni(II) treatment, where kinetic trapping is expected to be the mechanism of binding, for the SL fragment the mean square end-to-end distance was found to be $15,000 \pm 700 \text{ nm}^2$ and compares well with the expected value of $14,936 \text{ nm}^2$. Additionally, the mean square end-to-end distance was found to be $80,000 \pm 10,000 \text{ nm}^2$ for the LL fragment, and compares well to the expected value of $94,000 \text{ nm}^2$ for kinetically trapped molecules. As such, we can be confident that samples prepared with Mg(II) alone are surface equilibrated, whilst those prepared with the additional Ni(II) treatment are kinetically trapped. These values support the results of the previous chapter.

Table 4-1 Comparison of theoretical values of average square end-to-end distances with for each binding case, with experimental values for the two linear fragments. Samples prepared with Mg(II) are equilibrated, whilst additional Ni(II) treatment leads to trapping. Values are in nm^2 .

Sample	$\langle R^2 \rangle$ Mg(II) only	$\langle R^2 \rangle_{2D}$ equilibrated	$\langle R^2 \rangle$ +Ni(II) mica	$\langle R^2 \rangle_{3D}$ trapped
SL	$35,000 \pm 2000$	35,360	$15,000 \pm 700$	14,936
LL	$250,000 \pm 40,000$	282,000	$80,000 \pm 10,000$	94,000

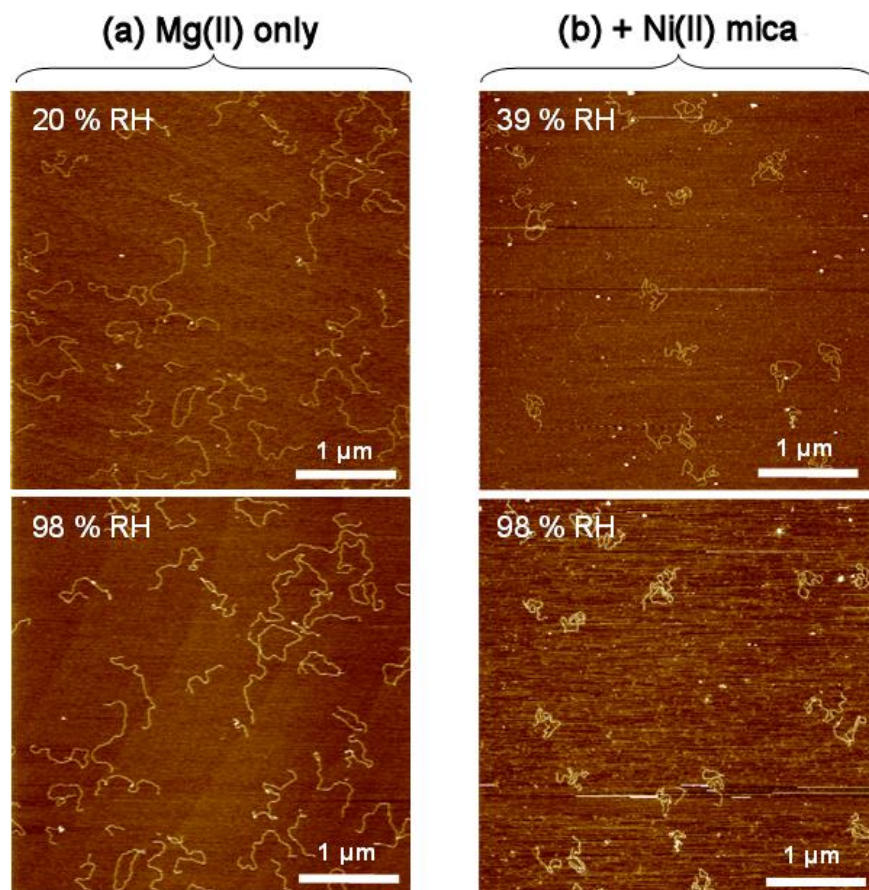


Figure 4-4. Images of linearised plasmid (LL) at both low and high humidity of the same region prepared with (a) Mg(II) alone and (b) Ni(II). No changes in conformation at high relative humidity are observed in either case.

The larger 4.2 kbp fragment showed no variation in molecular conformation when imaging was performed at increasing humidity. This was the case irrespective of whether or not Ni(II) treatment had been used (Figure 4-4). Figure 4-5 depicts software zooms of LL molecules at low and high humidity to facilitate comparison of molecular conformations.

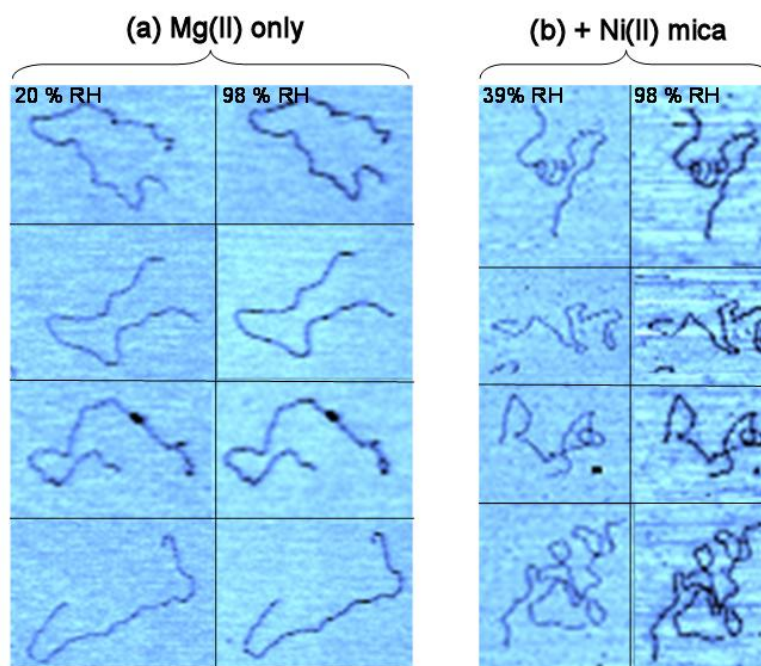


Figure 4-5. Software zooms of individual LL molecules, depicting molecules imaged from low to high humidity on (a) Mg(II) only prepared samples, and (b) with additional Ni(II) pre-treatment. No change in conformation is observed in each case.

However, for the SL 800 bp DNA fragment some different behaviour to the larger fragment was observed (Figure 4-6). Samples prepared with Ni(II) displayed the same results as the longer linear(LL) fragment; with the molecules remaining stationary, even at very high humidity. However, when the small fragments were deposited using Mg(II) in buffer only . a number of the DNA fragments displayed molecular perturbations at high humidity; regions on the molecules where local chain-chain interactions had occurred, leading to a “condensed-state” (marked by arrows on Figure 4-7) were typical. As well as this condensation behaviour, more wholesale changes were observed on many molecules. Where large sections of a molecule have moved across the surface, examples are circled in Figure 4-7. In addition to this many molecules also appeared to be more “kinked”.

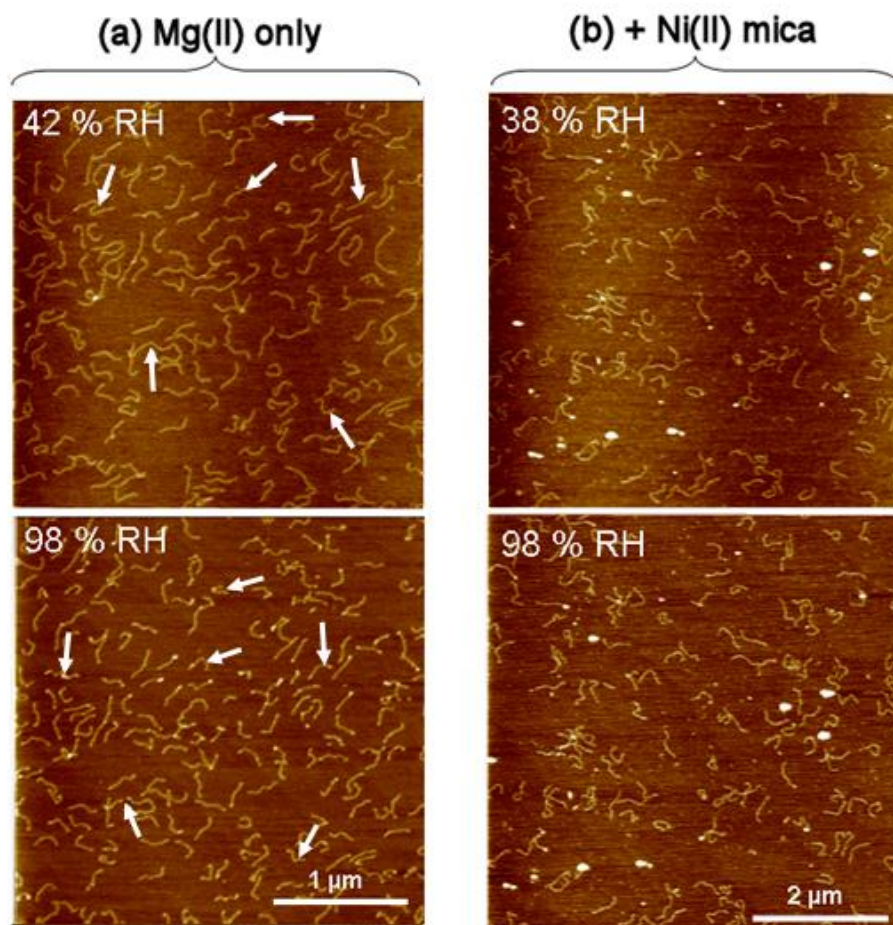


Figure 4-6. Images of 800 bp linear fragment (SL) at both low and high humidity prepared with (a) Mg(II) and (b) Ni(II) pre-treatment. With Ni(II) treatment no differences are seen between imaging at low humidity and high humidity. When the sample is prepared using Mg(II) a number of molecules begin to change shape, adopting a more condensed structure and/or movements in the position on the surface. A number of these changes are indicated with arrows.

Re-arrangements that involved in-plane clockwise rotation of DNA segments were approximately equal to those requiring an anticlockwise rotation ($n_{\text{Total}}=22$). It appears, therefore, that the structural changes do not occur in a preferred direction and are not influenced by the scan direction of the AFM tip. For condensation effects, globular features appeared on either end of the molecules in equal measure.

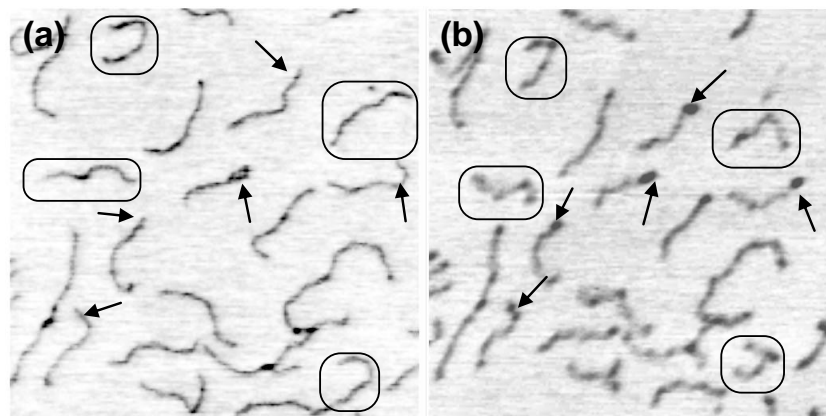


Figure 4-7. Software zoom of an area depicted in Figure 4-6 taken at (a) low humidity , and one at (b) high humidity. Widespread changes in conformation can be observed. A large proportion of molecules exhibit whole chain rearrangements (indicated by circles). In addition to this, intra-chain condensation has occurred in a number of molecules (indicated by arrows). This is typified by molecules curling up at the end of the chains and the visualisation of globular structures.

Contour length measurements were performed on the SL fragment prepared with Mg(II) only at both low and high humidity. At low humidity, the average contour length was 255 ± 1 nm ($n=40$), which corresponds to a base pair repeat of 0.328 ± 0.002 nm, consistent with B-form DNA. At high humidity, the molecules were separated into two groups: (1) Molecules that had undergone a conformational change, but that had not condensed. (2) Those that contained a condensation effect. Molecules that underwent a simple conformation change displayed a small reduction in contour length to 241 ± 2 nm. We conclude that this is due to a reduction of curvature of the molecule and increased kinking at high humidity. Height measurements of molecules that appeared more kinked at high humidity ($n=20$) were taken. At low RH the average height was 0.246 ± 0.007 nm, while at high RH this increased to 0.31 ± 0.01 nm. This may indicate that the molecules may have partially detached from the surface. The condensed molecules displayed a large reduction in contour length, to an average of 190 ± 5 nm, indicating that these represent the aftermath of whole re-arrangements of the molecule.

4.3.2 Circular Molecules

Second, we now consider what happens when a linear molecule is topologically closed. Both the supercoiled (SC) and open circular (OC) components of plasmid pBR322 were investigated. Once again, the circular plasmids were seen to adopt different structures, depending on whether or not the Ni(II) pre-treatment step was performed (Figure 4-8). In the absence of Ni(II), both SC and OC plasmids adopted a spread-out conformation, typical of surface equilibration. The conformation of samples prepared with Ni(II) were quite different, being more condensed with a number of cross-overs, indicative of kinetic trapping as the mechanism for adsorption.

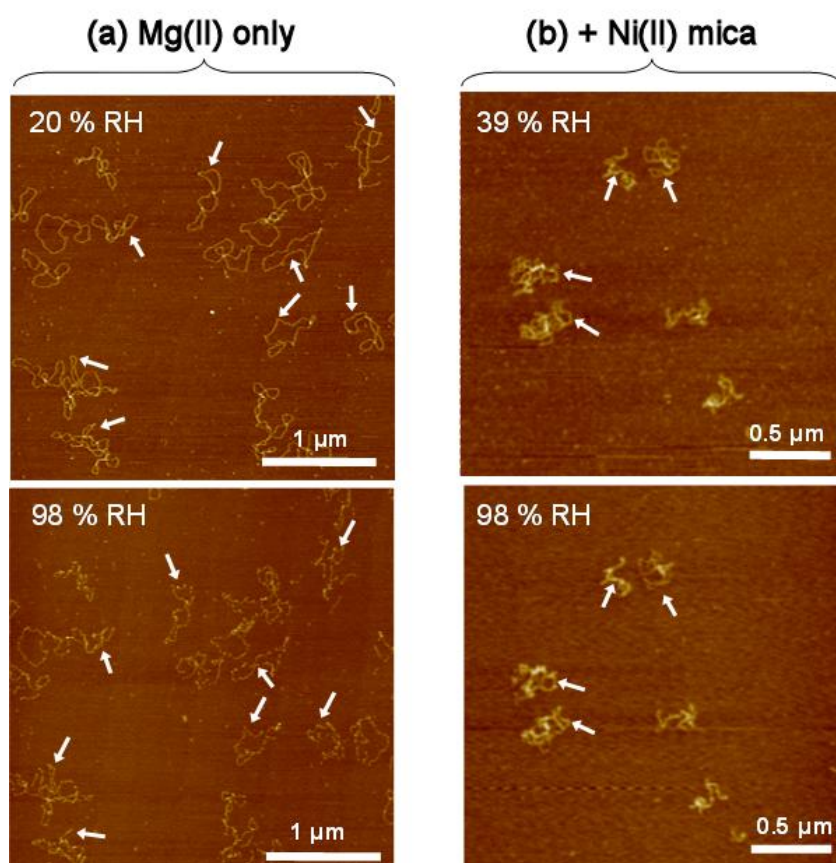


Figure 4-8. Comparison of DNA imaging at low humidity (top) and high humidity (bottom) for closed circular (SC) plasmid prepared with (a) Mg(II) only and (b) with additional Ni(II) pretreatment. At high humidity the DNA begins to regain mobility and local conformation changes can be observed. A number of these changes are indicated with white arrows. Conformational changes were observed for both preparation methods.

When SC plasmids were used, a number of interesting effects were observed when imaging at high humidity. Comparisons of the same imaging area at values across the humidity range demonstrated marked differences in DNA molecular conformation at increased humidity. This was true whether or not the mica surface was pre-treated with Ni(II) prior to sample deposition. Figure 4-8 shows a direct comparison between low and high humidity with and without Ni(II) pretreatment. White arrows highlight examples of conformational changes, which were observed across the entire scan area. Software zooms were performed on a number of exemplar molecules to further illustrate these conformational changes in more detail (Figure 4-9). These images have been colour-inverted to facilitate observation of the DNA chains.

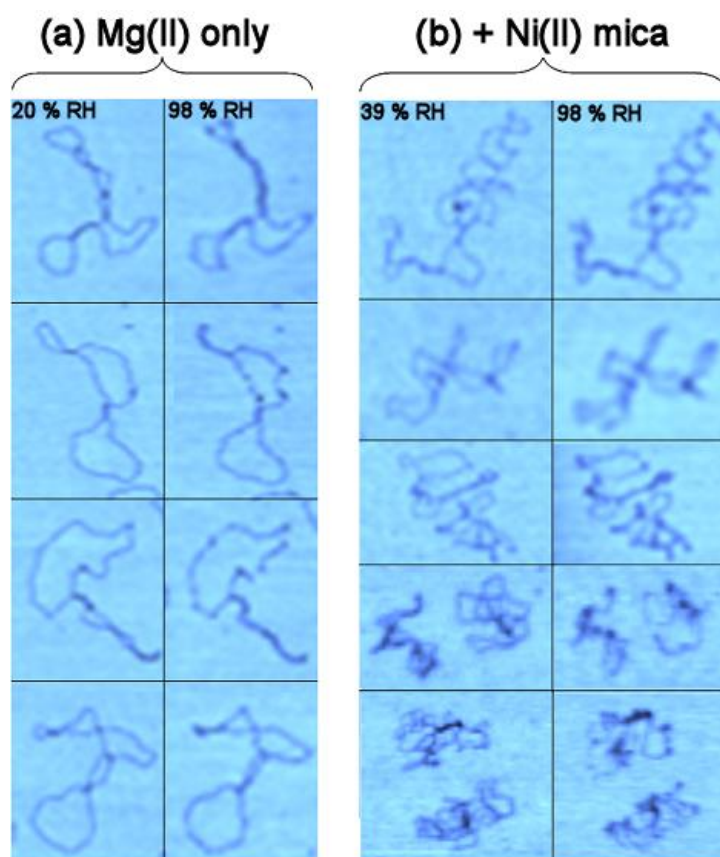


Figure 4-9. Software zooms of a number of SC DNA molecules depicted in Figure 4-8 showing the molecules imaged at low humidity (left) and high humidity (right) and illustrating the molecular perturbations. Both Mg(II) only (a) and Ni(II) pre-treatment (b) preparations are shown.

Wholesale changes in shape across the entire molecule were not usually evident; hence, we attribute these changes to DNA conformational changes and localised chain movements. Local DNA-DNA intra-chain interactions (condensation) were also evident but most noticeable were the large numbers of points where the curvature of the DNA molecules became kinked. For the Ni-mica sample, there are cases where the relative positions of different parts of the molecules have completely shifted.

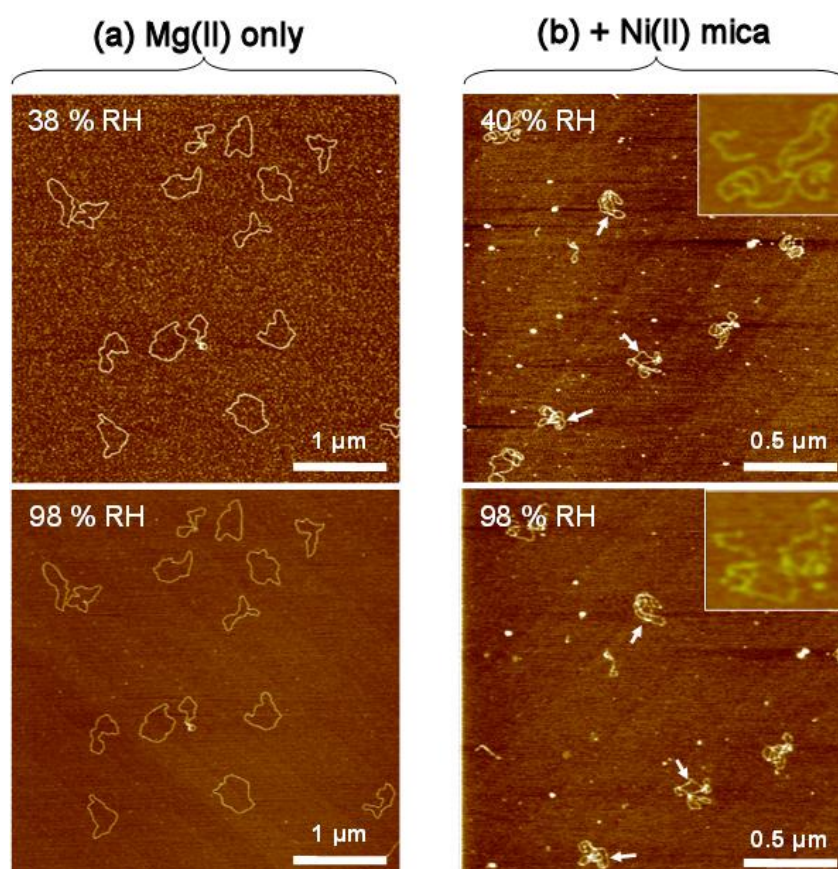


Figure 4-10. Images of relaxed (OC) plasmids at both low and high humidity prepared with Mg(II) (a) on mica and (b) Ni-mica. In (a) no changes of conformation can be observed, while in (b) small, localised conformational changes can be observed at high humidity. A example of such a change is shown in the inset of each image.

Conversely a relaxed plasmid (OC), with no supercoiling in solution by virtue of possessing a single nicked strand, prepared with Mg(II) alone showed no variation in conformation with increased relative humidity (Figure 4-10a and Figure 4-11a). This suggested that the local conformation changes and the increased kinking seen for the SC plasmid samples at high humidity were

topologically driven (Figure 4-9). On Ni-mica, small regions of the OC plasmid were observed to move at high humidity (Figure 4-10b and Figure 4-10b). These conformational changes associated with very high humidities were not as pronounced as those seen for the supercoiled SC plasmids (e.g. compare Figure 4-9 with Figure 4-11) and typically involved a reduction in curvature of regions of the DNA chain and again some localised kinking.

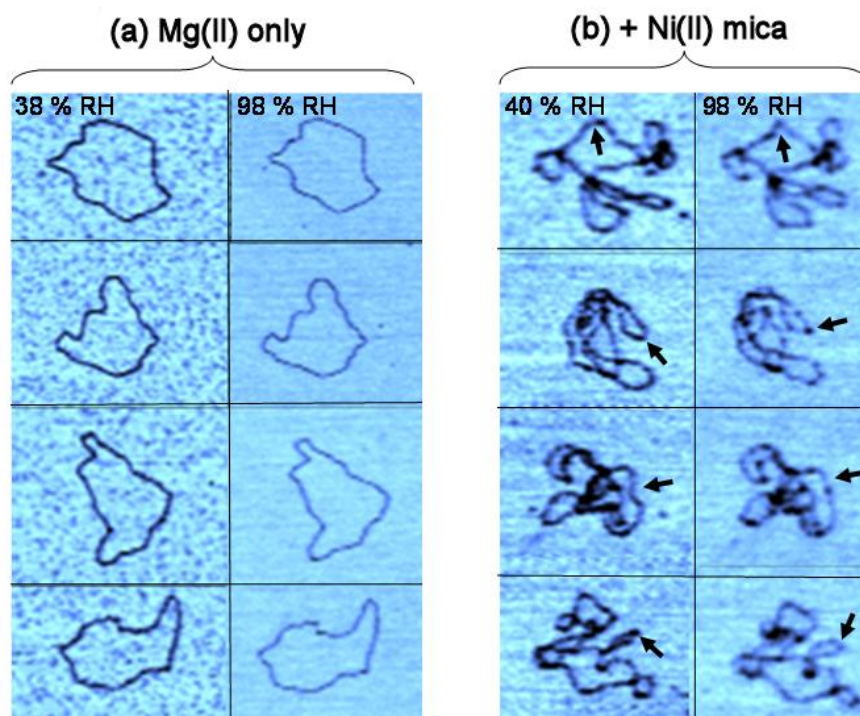


Figure 4-11. Software zooms of relaxed (OC) plasmids at both low and high humidity prepared with (a) Mg(II) on mica and (b) Ni(II)-mica. In (a) no changes of conformation can be observed, while in (b) small, localised conformational changes can be observed at high humidity (indicated by arrows), a consequence of local strain introduced in the molecules during the adsorption process.

4.3.3 Summary of results

Three DNA fragments from pBR322 plasmid and one linear 800 bp fragment were used as test molecules to understand the behaviour of dsDNA bound to mica under conditions of varying environmental humidity. Each sample was imaged from conditions of low humidity to high humidity (see Figure 4-12 for an example of a complete series). The binding mechanism of the DNA to the mica surface was also tested, and its effect on the behaviour of the molecular

motion investigated. The main discovery was that at very high RH (>90%) molecular motion and conformational changes occurred on the mica surface for some molecules under some binding conditions (e.g. see zooms in Figure 4-12).

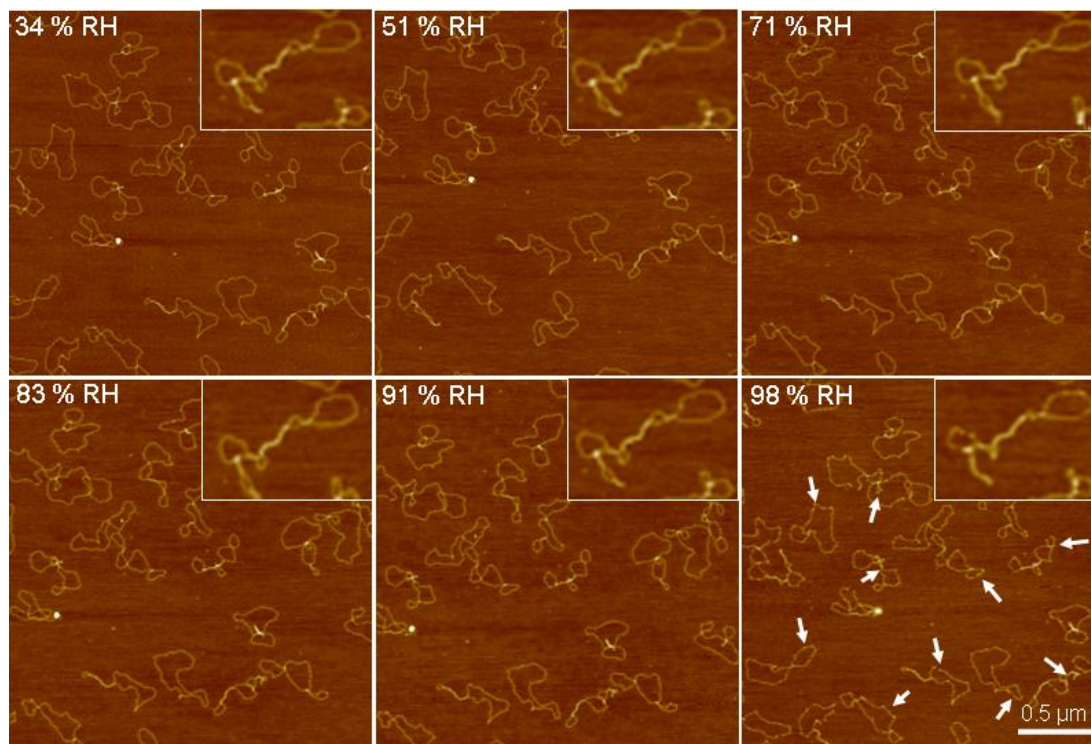


Figure 4-12. Series of AM AFM images of supercoiled (SC) plasmid prepared with Mg(II) on mica. The relative humidity (RH) is gradually increasing and shown in the top left of each image. Only once a very high humidity is reached (~ 98 % RH) can the DNA molecules gain some mobility and local conformational changes can be observed (see inset for details on one molecule). Other examples of these changes are indicated with white arrows.

Table 4-2 summarises the DNA molecules and binding conditions used and records whether or not any DNA conformational changes were observed at high RH. For longer dsDNA molecules (4.2 kbp) with no supercoiling strain, no backbone motion was seen on any length scale. Note this is valid for LL under either preparation method and OC in the equilibrated binding situation. Where short DNA fragments were used (800bp), then some molecular motion can occur when the DNA is surface equilibrated but not when kinetically-trapped. When supercoiled molecules are used, conformational changes can occur regardless of the sample preparation method. In all cases, the changes seen are local chain movements and not mass transport across the surface.

Table 4-2. Summary of the occurrence of local conformational changes or local motion of dsDNA molecules on mica at high RH (> 90%) from two sample preparation conditions. A tick signifies that observable DNA backbone motions occurred at high humidity, while a cross signifies that no conformational changes were seen at any humidity. More ticks indicate that conformational changes are larger scale and more widespread across that particular sample.

DNA \ Binding method	Mg(II) in solution (Equilibrated)	Mg(II) + Ni(II) mica (Trapped)
Linear 800bp (SL)	✓	×
Linear plasmid pBR322 (LL)	×	×
Open circular plasmid pBR322 (OC)	×	✓
Closed circular plasmid pBR322 (SC)	✓✓	✓✓✓

We observed conformational switching to occur in general at 95% RH or above. The transition from one conformational state to the final one occurs from one image to the next as the humidity is increased, the molecules being completely stable before and afterwards. The conformational switch, therefore, occurs on time-scales much faster than the image acquisition time (~ 5 mins), and this change was seen to be irreversible even when the RH was further increased or decreased again (See Figure 4-13).

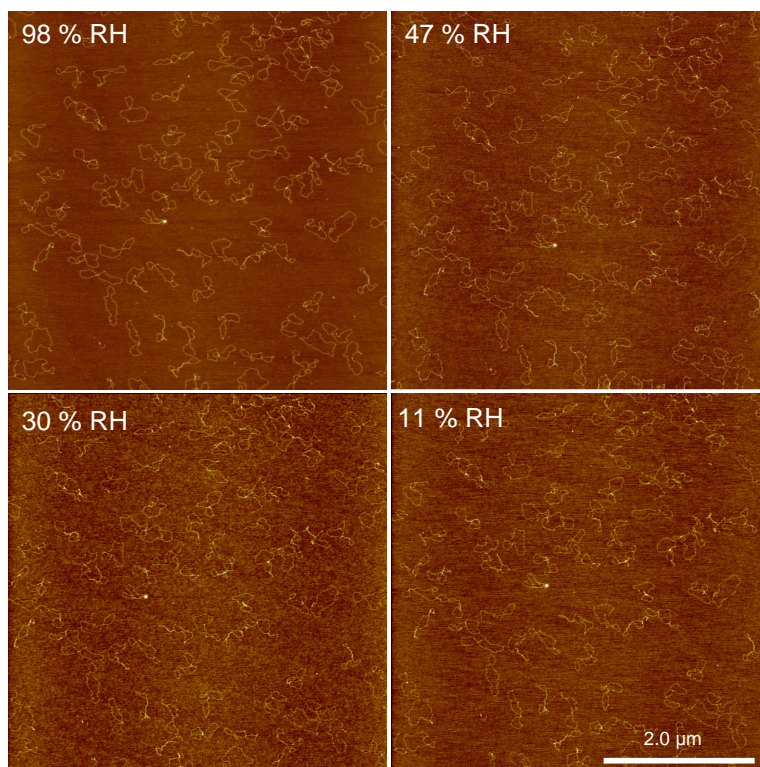


Figure 4-13. Series of AFM images showing the same region of a sample of closed circular (SC) plasmid pBR322 prepared with Mg(II) on mica. The relative humidity (RH) is gradually decreasing as shown in the top left of each image. No changes in DNA conformation are observed once the changes at high humidity (~ 98 % RH) have occurred. While the background image noise increases at lower RH, no changes in the apparent dimensions of the DNA occur.

4.4 Discussion

4.4.1 Mica water layer

These results show that the imaging of samples in a controlled-humidity environment can lead to irreversible conformational changes in the molecules at sufficiently high RH. Subtle changes in the behaviour occur depending on whether a binding mechanism is used that favours surface equilibration or kinetic trapping [167]. To explain these results we must first consider how the use of a mica substrate can influence the formation of surface water layers, and hence the local environment directly surrounding the surface-bound DNA.

The drying step in the sample preparation protocol effectively provides a “snapshot” in time and leaves the DNA molecule immobilised on the surface. Previous

work using techniques such as sum-frequency-generation (SFG) vibrational spectroscopy, scanning polarisation force microscopy (SPFM) [136], and also ellipsometry [17] give insights into the formation of water on mica surfaces. SFG vibrational spectroscopy has revealed that there is incomplete water coverage on mica below 90% RH, with an increase in ordering of the layer to become a more ordered hydrogen-bonding network at 90% RH [136]. These outcomes are largely in agreement with the ellipsometry measurements, which yield sub-nanometer thicknesses up to about 95% RH [17]. The picture may change somewhat when the AFM tip is introduced to the surface, because of capillary condensation effects [40]. In the presence of the tip, water can condense and/or migrate into the tip-sample junction increasing the local film thickness [213], and possibly affecting the local water structure.

At very high humidity (> 97 % RH) multi-layer water films can form on mica with a thickness greater than 1 nm in the absence of any capillary condensation [136], as was also observed in the ellipsometry study by a large increase in the thickness of the adsorbed film [17]. Interestingly, the humidity at which the thickness of the water layer becomes comparable to the diameter of dsDNA is about 97% RH. We observe conformational switching to occur in general at 95% RH or above. The presence of a fluid-like water layer, with sufficient thickness, clearly allows surface bound molecules a degree of mobility.

During the acquisition of the series of images at different RH it was noticed that a step in the background surface was often seen (black arrows Figure 4-14). These steps are too large to be digitisation and too small to be steps in the mica and moreover, the position of the steps changes from image to image and the steps are often curved and/or ruffled in appearance. We surmise that these steps are partial layers or droplets of water that are forming on the surface as the RH increases. The heights of these steps are typically 0.2 nm consistent with a water monolayer and analysis of a series of images suggests that the free interface is moving at about 30-40 nm/min.

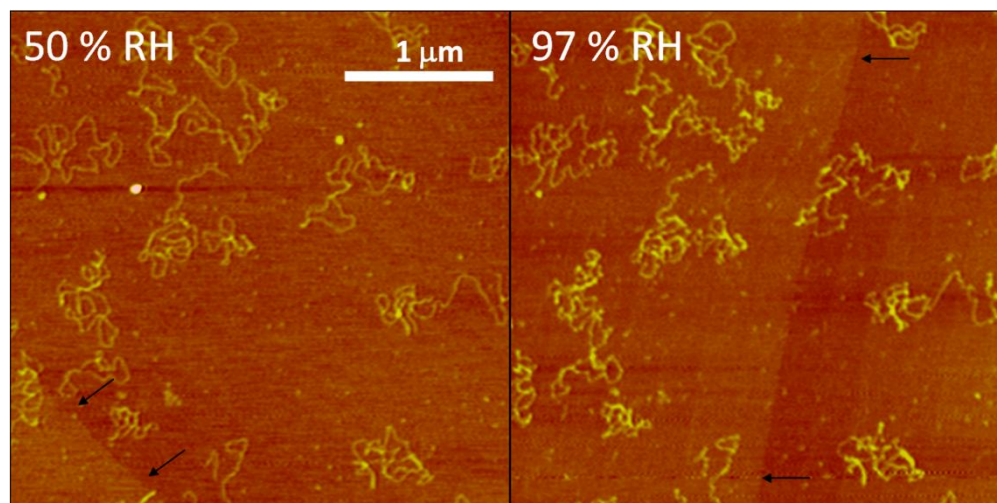


Figure 4-14. On some images a step in the background was observed (black arrows). The feature was seen to move when the same area was imaged over a period of time. Measurements of the height of such features are comparable with that of a water monolayer. These droplets appear to be localised in specific areas on the surface, but are larger than the scan size leading to the straight features seen here.

The mechanisms that affect the initial conformation of the DNA on mica have been explained in detail previously [153, 194]. At the low concentrations of salt used, surface equilibration was observed in the absence of Ni(II) (Mg(II) only), whilst when Ni(II) pre-treatment of mica was used, the molecules appeared to be kinetically-trapped based on the observed molecular conformation and end-to-end distances of linear fragments. Mg(II)-only samples were spread-out and contained few crossings, whereas on Ni-mica the DNA molecules contained many crossings. In the case of linear DNA molecules, a smaller end-to-end distance was also observed. This result is similar to previously reported experimental studies [153, 194], and can be attributed to the degree of charge of the mica surface. On a weakly charged surface, DNA adopts a 2D conformation, whereas a highly charged surface will cause the molecules to be projected.

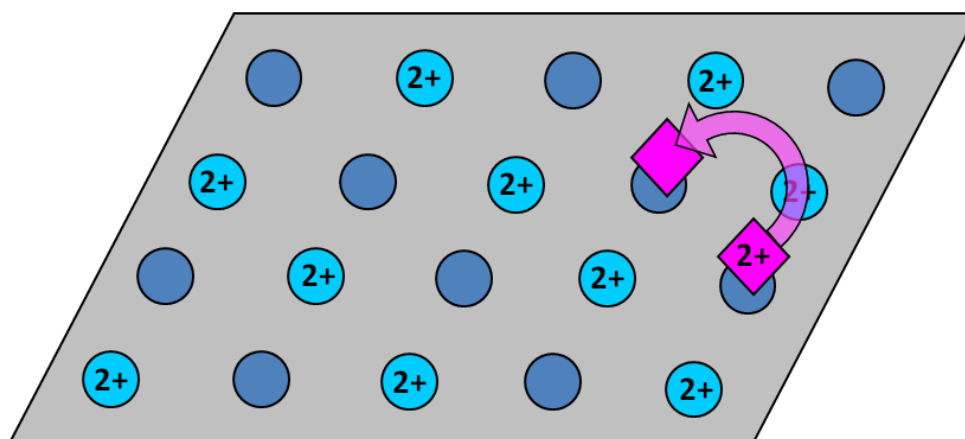


Figure 4-15. Schematic diagram of the lattice model of mica proposed by Pastre *et al.* [153]. If the divalent cations are of a suitable size then they bind into sites on the surface. Cations associated with the surface are shown as blue spheres, and those associated with DNA are shown as pink diamonds. A net attraction exists when these ions take up a staggered orientation. For movement of a DNA molecule, the electrostatic interactions between the molecule and the mica surface must be broken, leaving the cation free to move to another lattice site (indicated by arrows).

Pastre *et al.* modelled the DNA and mica surface as charged planes and hypothesised that if their respective counterions were to adopt a staggered configuration, they could be shared between the two surfaces [153]. Lateral diffusion of DNA molecules on the mica surface is prevented by large frictional forces, related to electrostatic interactions between DNA and mica. If we consider the mica surface as a lattice of charges [147], for DNA segments to move laterally divalent counterions condensed on the DNA must “jump” from one staggered position to the next [154] (see Figure 4-15). For this to occur an energy barrier must be overcome.

4.4.2 Linear Fragments

In our studies at high RH we begin to observe changes in conformation for the 800 bp fragments (SL) when Mg(II) alone was used for DNA binding. No such changes were observed when Ni(II) pre-treatment was used. Differences in binding mechanisms will be an important contributing factor in explaining the observed changes. As the liquid layer forms, DNA becomes more mobile, straining the binding interactions between molecule and substrate. The

conformational changes in the absence of Ni(II) are thus related to the relative strengths of binding of Mg(II) and Ni(II) into the mica cavities. In the presence of Ni(II), the cations remain strongly bound in the original mica cavities and the DNA is unable to move. However, Mg(II) ions did not remain irreversibly bound with increasing RH, but are much more mobile and able to break free of the cavities in the mica surface. The counter-ions associated with the DNA are thus able to migrate to different lattice sites in the mica or to positions between adjacent DNA strands. Where sufficient ionic translocations occur then the DNA molecule would undergo a large-scale change in conformation (Figure 4-16a).

Moreover, in the same case a number of molecules were observed to “condense” and form globular structures. We suggest that the DNA molecules curled up because the divalent Mg(II) cations promote DNA condensation through counter-ion correlations [24, 108]. It is an open question at present what role the tip plays in this process, although as we have seen, the presence of the tip may induce capillary condensation and increase the amount of water in the vicinity of the AFM tip. Tip-induced folding of DNA has previously been observed by AFM [94]. It appears that as a consequence of a thicker water layer, the Mg(II) ions move away from interactions between DNA-mica to DNA-DNA interactions. The condensation effect always occurred at one end of the molecule, and therefore may be related to molecules which are not fully bound to the mica surface at the time of sample drying.

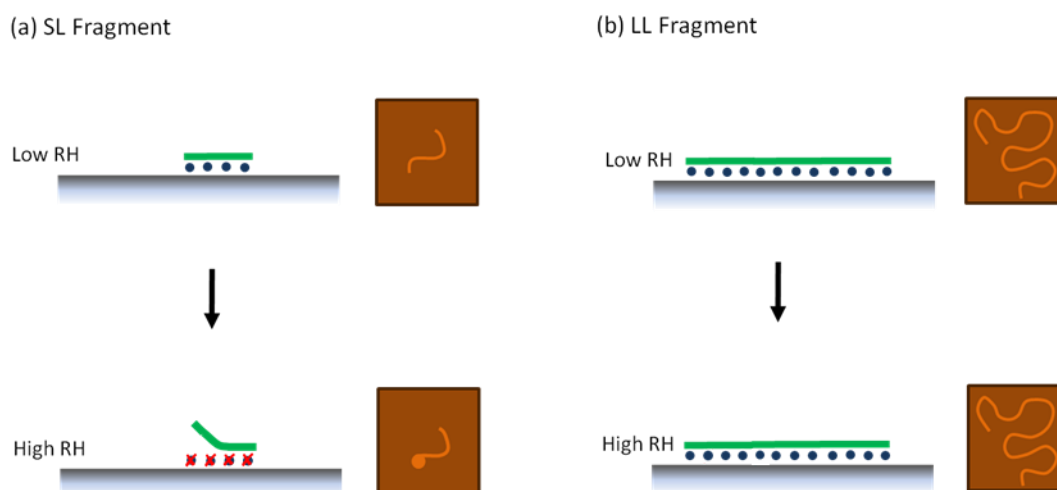


Figure 4-16. Schematic representation of the behavior observed when surface equilibrated linear fragments are imaged from low to high humidity. The DNA is represented as a green rod, and counterions as blue spheres. To the right of each schematic is a pictorial representation of what a typical molecule may look like under AFM. (a) The small SL fragment has fewer interactions with the surface. These become strained and can break at high RH (as depicted by red crosses), and the molecule can partially break free of the surface. This leads to re-arrangements, in this case DNA condensation. (b) The LL fragment has many more attachments to the surface. Even conditions of high RH, the molecule remains tightly bound due to the cooperative nature of the correlation force. Each interaction between a counterion is stabilised by many more nearest-neighbor interactions. In this situation, no conformational changes are observed.

No obvious molecular changes were observed for the longer LL fragment with increasing RH, even up to an RH of 98 %, whether in the presence or absence of Ni(II). We postulate that this is due to the cooperative nature of the binding mechanism (i.e. counter-ion correlations). These results suggest that these longer fragments are more strongly bound to the surface than the smaller fragments. This could be related to the energy cost of breaking the interactions between the DNA, mica and their respective counter-ions. Integrating over all sites of contact for a longer molecule will give a larger interaction energy. Considering the lattice model of mica, there will be many more points-of-contact for the LL fragment than the SL (see Figure 4-16b). Movement of the molecule requires ion translocation to other lattice sites in the mica. The frictional cost to move many condensed cations appears too great, even on loosely absorbed Mg(II) prepared samples.

4.4.3 Circular molecules

Now we turn to the circular DNA molecules. Provided that drag is not significant, linear molecules are able to relieve any stress imposed upon them by rotating in space [143]. In the case of circular molecules however, which can exist in various topological conformations, supercoiling can induce large amounts of strain into the molecule [14]. The degree to which a circular DNA molecule is supercoiled can be defined by the conserved linking number; $L_k = T_w + W_r$, where T_w is the twist and W_r is the writhe (i.e. how the chain coils in space). Strain energy can therefore be partitioned between helical twist and writhe (i.e. global conformation), depending on the circumstances. Molecular dynamics studies that have considered supercoiled DNA adsorption have suggested that the surface can act to restrict the configurational freedom of the DNA, altering structural properties such as twist and writhe [71, 204].

We observed the greatest occurrence and largest extents of conformational changes for the supercoiled plasmid (SC). Supercoiled plasmids appeared to temporarily regain mobility and switch from one stable conformation to another, although no molecules became completely detached from the surface. Some areas of the molecules were seen to open out while others appeared to condense. These results are similar to conformational changes seen in other AFM studies, when supercoiled DNA is imaged under bulk aqueous liquid [140, 223]. Unlike imaging in bulk liquid, the whole of the molecule can be clearly resolved in humidity-controlled AFM without movements of segments away from the surface. A prevalent effect is molecules becoming more kinked, which may be a consequence of only localised detachment of segments from the surface in between pinning points, where the binding interaction is the strongest.

The supercoiled plasmid (SC) shows the largest changes in conformation when bound to the Ni-mica, compared with the Mg(II) case. This is expected, since a kinetically-trapped configuration is closer to the 3D solution form, whereas surface equilibrated molecules collapse into 2D. It has been shown previously

that supercoiled molecules in the surface equilibrated case can partition writhe into local twist to relieve strain and adopt a 2D conformation [37]. In contrast, we postulate that the 3D configuration allows large rotational motions of sections of the DNA molecules to relieve supercoiling strain energy.

Individual molecules displayed differing responses to the increased humidity. Some DNA loops opened out, while other areas were observed to condense. These differences can be explained by considering local differences in surface adhesion or high local tensions [223]. Nagami *et al.* postulated that along the length of a supercoiled DNA molecule there might be sections that are more strongly attached to the surface than others [140]. As the molecules begin to gain mobility, loosely attached parts open up, causing supercoiling tension to accumulate in other sections. The condensed counterions coupled with the high tension might drive other sections of the chain towards a condensed state (see Figure 4-17).

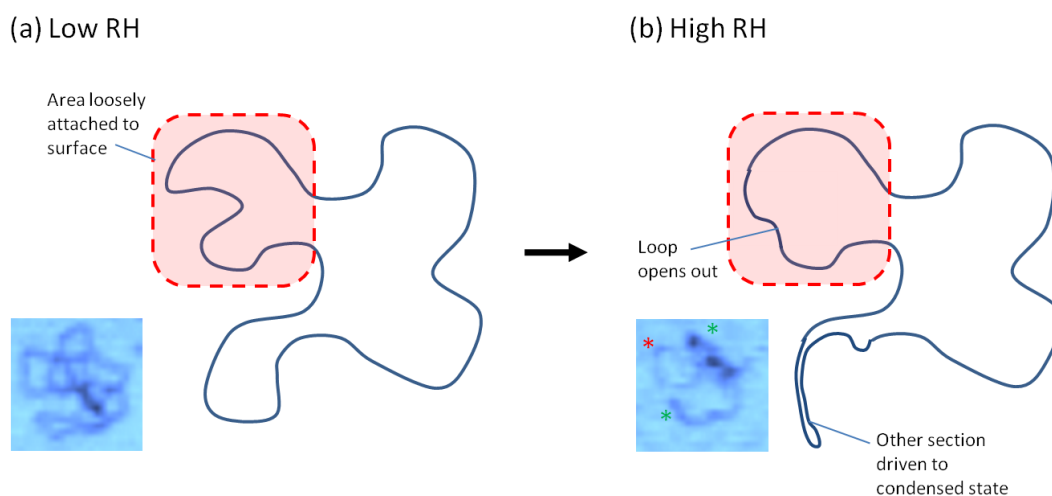


Figure 4-17. Schematic representation of how SC molecules behave when imaged at high RH. Different segments react differently depending on how strongly they are bound to the surface [140]. Here a loop is depicted opening out. This event increases the tension in neighboring segments, which are then driven towards a condensed state. Also shown is a software zoom of a supercoiled plasmid under AFM. At high RH (b) the opening out of a loop is indicated by a red asterisk, whilst increased condensation is shown by a green asterisk.

For the relaxed plasmid (OC), samples prepared using Mg(II) adopt a spread-out conformation typical of surface equilibration. No deviation from this initial

conformation was observed, even at very high RH. This again is thought to be related to the large number of binding sites between the molecule and the surface, meaning a large energy barrier must be overcome to detach segments from the surface. Moreover, stress induction would not be an issue for the relaxed sample by virtue of it having a single, nicked DNA strand. It is thus able to relax onto the surface in 2D without any topological constraints. As such, as the humidity increases, there is no strain built up to drive topological rearrangements, and the molecule remains irreversibly bound due to its many interactions with the surface.

This situation changes, however, when the OC molecules are deposited on Ni-mica. In this case, the relaxed plasmid adopts a kinetically-trapped conformation, where certain segments of the molecule become pinned, while the rest of the molecule collapses down on top of these points. This effectively makes the molecule a 3D projected object, and means that there are pockets of aqueous fluid underneath DNA chains that are stacked on top of each other, giving rise to a weaker, longer range interaction between a DNA chain segment at a higher level and the surface. More importantly perhaps, the binding trajectory of the molecule onto Ni-mica will lead to configurational constraints being imposed on the molecule. Kinetic trapping may induce stress into the molecule, as it is made to cross over itself in order to fit into a smaller area on the surface, a consequence of random initial attachment and pinning. As the humidity increases, small movements are observed in a number of molecules that could be caused by local variations in surface adhesion, or through the molecule releasing strain, which had arisen in the adsorption procedure. No movements were observed in the LL fragment, which also exhibited numerous crossovers on Ni-mica, suggesting that topologically closing the fragment allows strain to build up during binding. This is an example of a frustrated process, where the molecule needs to compromise between internal strain energy and surface binding energy. These perturbations ultimately arise due to the build-up of stress in the open circular chain during deposition, again a topological constraint which is subsequently released at high humidity. Note that the extent

of the changes due to strain build up during surface binding are qualitatively less than those induced by inherent supercoiling.

4.4.4 Differences between linear and circular fragments

It would appear that the conformational changes seen when imaging is undertaken at high RH occur via two different mechanisms for the linear and circular molecules, as evidenced by their different responses, both in terms of overall motions and to the different preparation methods.

The linear molecules only display changes when the interaction of the DNA with the surface is sufficiently weak, i.e. in short, surface equilibrated molecules prepared with Mg(II) only. No changes were observed in kinetically trapped DNA. Hence, the re-arrangements of the molecules are mediated by the strength of the counter-ion correlation force, and the mobility of the divalent cations used.

The circular molecules displayed differing behaviour as a response to the increased water layer. Irrespective of the preparation method used, the supercoiled fragments displayed topological re-arrangements, as did kinetically trapped OC molecules. In fact, kinetically trapped molecules, hence more strongly bound to the surface, displayed larger readjustments than equilibrated molecules. This is the opposite of what is seen in the linear molecules. These facts indicate that superhelical stress stored in the molecule prior to binding, and the subsequent release of these constraints as the molecules regain some freedom at high RH, is the underlying mechanism driving the irreversible conformational changes observed here.

4.4.5 Possible use of humidity-controlled AFM in transcription studies

By using the techniques described here it should be possible to distinguish between molecules that are supercoiled and those that are not, by observing an associated conformational change when imaging at high RH. As such, the method could be used to study the importance of DNA supercoiling in

transcription regulation. To actively transcribe a gene a polymerase must track the DNA's helical pitch, rotating relative to the central axis of the chain backbone. This motion is akin to screwing a nut on bolt. Liu and Wang proposed that if the RNAP had a large transcript, and other ancillary proteins, such as a ribosome, it would experience a significant amount of viscous drag, which would hamper active elongation. In response to these observations, they proposed the 'twin-supercoiled domain model', in which if DNA rotates relative to RNAP then the advancing polymerase would generate positive supercoils in the DNA template directly ahead of it, and negative supercoils behind [121] (Figure 4-18a).

This would have no effect for transcription initiated on a linear template originating from a single promoter, unless the template was constrained at both ends. The molecule could simply rotate in space, thus dissipating any superhelical tension. However, in the case of transcription originating from two convergently aligned promoters, positive supercoils can be propagated if the RNAP acts as a kind of "topological barrier", preventing the dissipation of twist. This could lead to the build-up of helical stress in the region between the promoter sites, which could lead to stalling of any translocating RNAPs (Figure 4-18b).

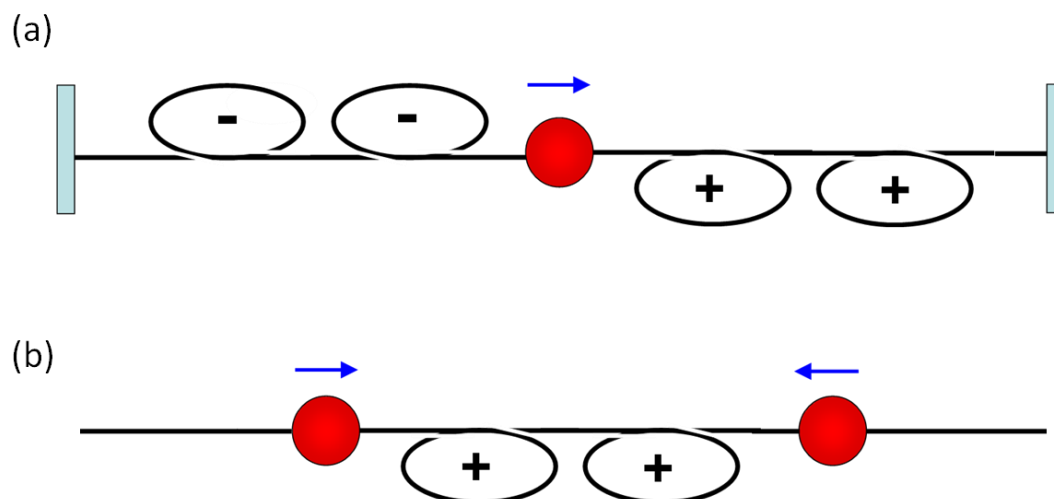


Figure 4-18. Models of supercoiling produced by elongating RNAP originally proposed by Liu and Wang [121]. (a) A linear DNA molecule constrained at both ends. The active transcription by a single RNAP leads to positive supercoils ahead of the enzyme, whilst leaving negative supercoils in its wake. (b) Model of what might occur during convergent transcription. The two RNAPs approaching each other could create a build-up of positive supercoiling in the inter-polymerase space, which could affect the efficiency of transcription.

The presence of a region of high tension, a consequence of convergent transcription, could possibly be observed by imaging at high humidity. A possible experiment could involve starting transcription elongation with a subset of nucleotides. The RNAPs will stall on the first occurrence of the missing base, at some point downstream of the promoter. The distance they travel may be such so as to cause supercoiling to be propagated in the inter-polymerase region. This could be verified by increasing the humidity to close to 100 % RH. On the supercoiled plasmids examined here conformational changes seem to be driven by the release of supercoiling strain. If the imaging of DNA-RNAP complexes at high RH was accompanied by a corresponding conformational change in the chain region between the proteins then it would confirm the presence of positive supercoiling. The use of linear fragments of a certain size may mean that some changes, such as those seen in the SL at high RH would also occur, thus confusing interpretation of the results. However, these effects could be negated by also using kinetically trapped samples. Ni(II) pre-treatment stopped small linear fragments from moving, but still allowed re-arrangements in supercoiled samples.

4.5 Conclusions

In this study four model DNA fragments and two sample preparations were compared; binding of DNA to untreated muscovite mica in the presence of Mg(II) and pre-treatment of mica with Ni(II) ions (Ni-mica) prior to addition of DNA in the presence of Mg(II). For long linear plasmid (pBR322) molecules, no conformational changes were observed in either condition. For short linear DNA fragments (800bp), the ends of molecules folded over at high RH when only Mg(II) was present but not when Ni(II) pre-treatment had occurred. For open-circular pBR322, very small local changes in the DNA backbone were seen for Ni-mica but not for bare mica with Mg(II). For closed-circular pBR322, large relative movements in the local backbone positions were observed at high RH in both conditions, but most particularly on Ni-mica.

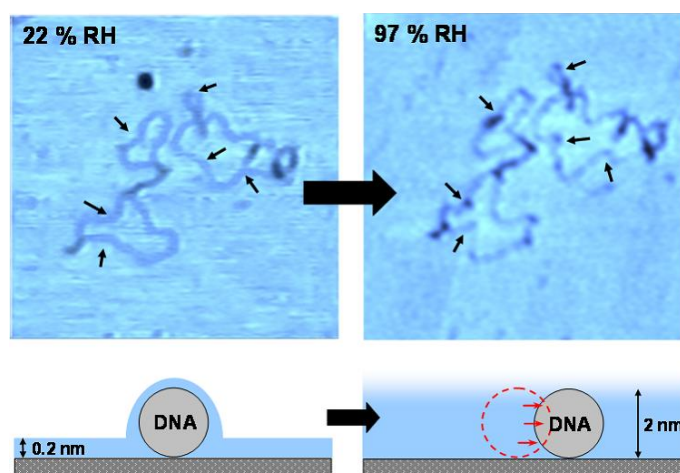


Figure 4-19. Summary showing typical conformational changes in SC supercoiled DNA as the humidity is increased, with a corresponding increase in the thickness of the surface water layer. Topological rearrangements in the DNA are observed to induce DNA condensation, kinking of the backbone, and changes in the backbone trajectory. A number of these changes are highlighted with arrows.

It appears likely that the scanning of the tip does not induce the conformational changes, as such changes are not observed across every sample. Rather, both the presence and the extent of the movement were found to be dependent on the molecule studied and the mode of binding.

In summary, it is suggested that achieving very high humidities for AFM imaging applications in air is not only possible, but is desirable in the case of studying DNA structures and dynamics because it gives two possible conformational states for a given topoisomer (see Figure 4-19). By imaging under two extremes of surface binding states while varying the humidity in the AFM, it should be possible to distinguish between different topoisomers, enabling the AFM to characterise conformations of circular DNA including catenanes and knots. Moreover it can alleviate problems associated with AFM imaging under bulk liquid where modulating the binding interactions is challenging and following DNA processes *in situ* inevitably leads to loss of information about the DNA backbone positions during the experiment. Humidity-controlled AFM imaging is a novel and desirable method for investigating topological properties of DNA on surfaces and will assist investigation of topology changes in DNA induced by DNA-dependent proteins or other molecules.

Chapter 5

5 Study of transcription interference in DNA: E. Coli RNAP complexes

5.1 Introduction

In this chapter we use AFM to probe the outcomes of transcription events on linear DNA templates containing two promoters. It is possible for both promoters to independently recruit RNAPs. Hence, there is the possibility for two RNAPs to interfere with each other's progress (a state of transcriptional interference (TI)). We investigated two distinct promoter arrangements on two separate DNA fragments: pDSU which contained its promoters in a convergent orientation, and pDSP which had its promoters in a tandem orientation.

AFM is particularly useful for obtaining direct structural information. By using *in vitro* transcription reactions to form DNA-RNAP complexes that represent different stages of the interaction, and depositing the resultant complexes onto a mica substrate for AFM imaging in air, it became possible to study the interplay between two RNAPs independently transcribing the same stretch of DNA. By measuring the positions of RNAP on the template it is possible to distinguish between different stages of the reaction by studying open promoter complexes (OPCs), stalled elongation complexes (SECs) and collided complexes (CCs). By studying distributions of the RNAPs positions on the template it becomes possible to obtain a "snap-shot" of the collision event, and infer the consequences of such an occurrence. We observe that after interacting RNAPs in both the tandem and convergent templates remained on the template, often in close proximity to each other, a result that was indicative of RNAP stalling. We also found that one RNAP can push another along the template, leading to a situation of backtracking in the convergent configuration.

5.2 Sample preparation

For studies into convergent transcription the plasmid pDSU (6136 bp) was used (Figure 5-1). This plasmid contains two λ_{PR} promoters situated in a convergent arrangement, and separated by 338 bp. The template does not contain a terminator sequence, and so the chance of two RNAPs colliding with each other is increased. The plasmid underwent enzymatic digestion by HindIII for two hours at 37 °C to produce a linear fragment 1149 bps in length and containing both promoter sites, plus three other extraneous fragments. Gel electrophoresis was performed to separate the components by virtue of their different migration rates. The fragment of interest was visualised under UV light as a discreet band, and subsequently excised from the gel and purified using the QIAquick Gel Extraction Kit (Qiagen, Valencia, CA), as per the manufacturers instructions.

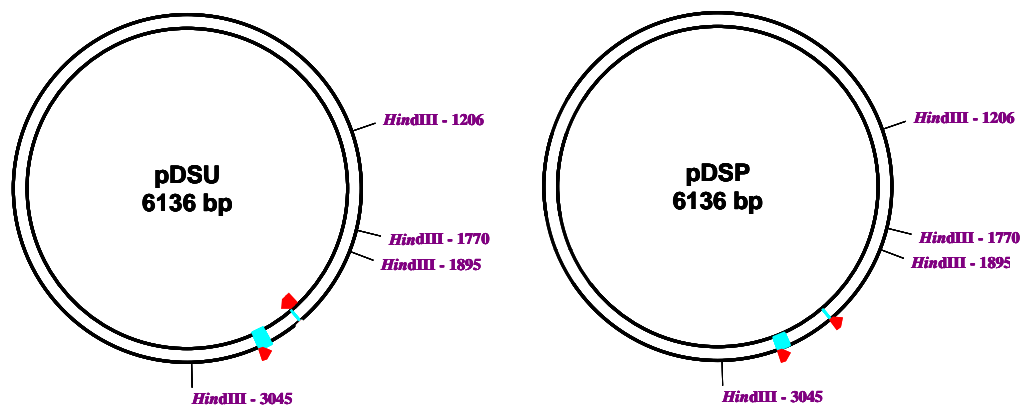


Figure 5-1. pDSU and pDSP plasmid constructs. Both plasmids can be digested with Hind III to produce linear templates (1149 bp) containing two distinct promoters. pDSU contains two promoters situated in a convergent arrangement on opposite DNA strands. pDSPs promoters are oriented in the same direction, in tandem, and on the same strand. The direction of the promoter is indicated by red triangles.

To compare transcription originating from a different promoter arrangement the plasmid pDSP was used. This was identical to plasmid pDSU except that one of the promoters has been turned around, such that they are both oriented in the same direction, on the same DNA strand. A linear transcription template,

also 1149 bps in length, from plasmid pDSP was prepared using the same restriction digestion and purification method as applied to pDSU.

Figure 5-2a shows a schematic of the 1149 bp fragment used in the convergent transcription experiments, and includes the positions of the two oppositely-aligned promoter sites. The hatched region directly downstream of each promoter site depicts areas on the template that are deficient in the base cytosine. This allows actively transcribing complexes to be stalled at the end of these regions by a process of nucleotide omission, thus allowing the processes of transcription initiation and the eventual collision to be investigated separately. The distances from the start of the promoter regions and the ends of the template are different for each site. This asymmetry means that promoter-bound RNAPs can be simply identified by virtue of their positions from the ends of the chain. It also means that three measurements can be defined from each complex (Figure 5-2b): the inter-RNAP distance (the contour length along the DNA backbone between the centres of the two RNAPs), and two arm lengths (the contour length from the centre of each RNAP to their closest respective end of the template). The smaller of these two values is termed the short-arm distance, whilst the larger is the long-arm.

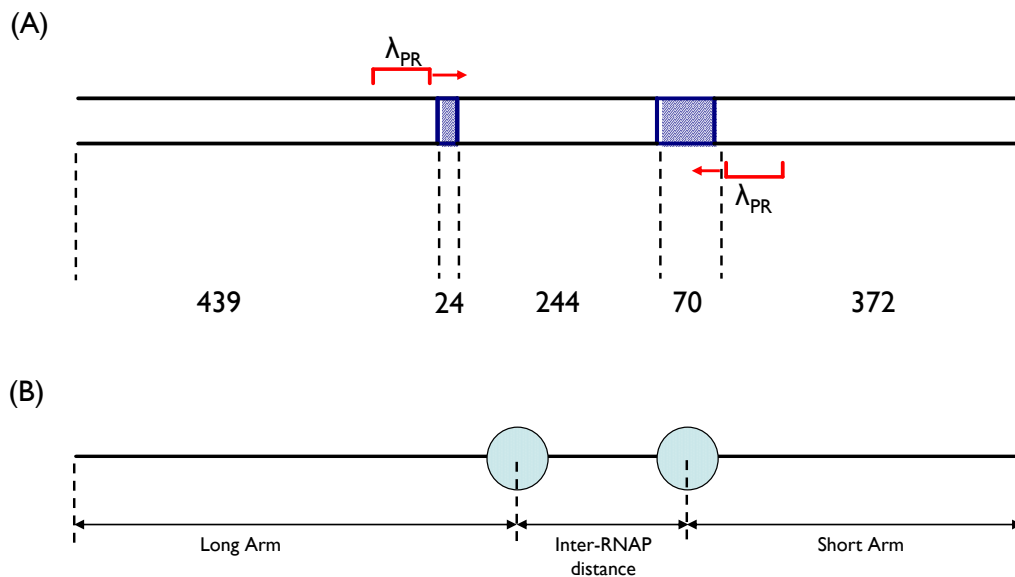


Figure 5-2. (a) Schematic representation of the 1149 bp fragment used in the convergent transcription experiments. This contains two λ_{PR} promoters separated by 338 bp. Sequences deficient in C bases, directly downstream of each promoter, are shown by blue hatching. Also shown is the distance in base pairs from each promoter to the end of the template. (b) The asymmetry of the template allows three measurements to be defined for each complex. The centre of mass of each RNAP is taken as the intersection of each measurement.

An *in vitro* transcription reaction was performed using a protocol derived from similar AFM studies into transcription complexes [47, 50, 166, 168] (see Figure 5-3 for schematic describing process). The initial stage involved taking 200 fmol of the 1149 bp *Hind III* digested fragment of pDSU and incubating it at 37 °C for 15 minutes with 400 fmol of *E. Coli*· σ^{70} holoenzyme (Epicentre, Madison, WI, USA), and 10 μ l of transcription buffer (20 mM Tris-HCl (pH 7.9), 50 mM KCl, 5 mM MgCl₂, 1 mM DTT). This initial step enables the formation of open promoter complexes (OPCs), which consist of template DNA with RNAPs situated at the two promoter sites, but which are unable to begin actively transcribing the chain.

The second stage of transcription elongation can be initiated through the addition of NTPs to the reaction mix. Stalled elongation complexes (SEC) were formed by adding ATP, GTP, and UTP to a final concentration of 200 μ M and incubating at room temperature for 15 minutes. Due to the absence of CTP, the

RNAPs will stall at the first occurrence of the base cytosine on the coding strand, downstream of each promoter site.

Once all four NTPs are added to the reaction mix then it is expected that transcription elongation can proceed unhindered until RNAPs travelling in opposite directions begin to interact with each other. Collided complexes (CCs) were formed in one of two ways. In the two-step method of collision formation, 200 μM of CTP was added to the mix containing SECs, before being left to incubate at room temperature for 15 minutes. This allows the processes of stalling and re-initiation to be studied. The one-step method involved adding ATP, CTP, GTP, and UTP (100 μM each) to the reaction mix containing OPCs and incubating at room temperature for 15 minutes at room temperature.

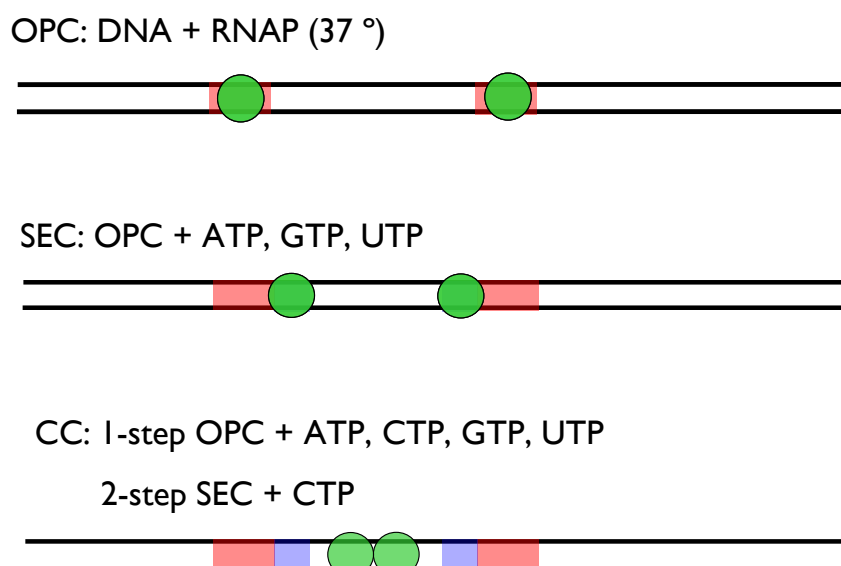


Figure 5-3. Schematic diagram showing a summary of the different classes of complex formed for investigations into convergent transcription. RNAPs are shown as green spheres, the promoter region as red rectangles, and the C-less region as blue rectangles.

All reaction mixes were studied separately by AFM, and were diluted 10-fold into imaging buffer (4 mM Tris-HCl (pH 7.5), 4 mM MgCl_2), to facilitate efficient binding to the substrate. Complexes were then deposited onto bare muscovite mica, and imaged in ambient conditions ($\sim 35\%$ RH) using tapping mode AFM. Complexes were analysed in the Nanoscope software (Digital Instruments, Santa Barbara, CA, USA) and using ImageJ (NIH, USA). Complexes containing two RNAPs were analysed in the following manner. Measurements of contour

lengths were performed manually by tracing the DNA backbone as a series of connected straight lines. Three contour length measurements were taken as described previously (Figure 5-2b): the short arm distance, inter-RNAP distance, and the long-arm distance. The centre of mass of the 2D projection of each RNAP was taken as the intersection between one measurement upstream on the template, and the next downstream. By studying how these three measurements change relative to each other during the different stages of the reaction it will be possible to gather information about the collision process during convergent transcription. Additional analysis was performed for the OPC reaction mix, which only had 0 and 1 RNAP bound to the template.

5.3 Convergent pDSU template

5.3.1 Open Promoter complexes (OPCs)

OPCs were formed by incubating template DNA containing the convergently aligned promoters with *E. Coli* RNAP holoenzyme. The RNAPs are able to form stable interactions with the promoter region, but due to the absence of NTPs are unable to move away from their positions to begin active transcription. Figure 5-4 shows representative AFM image of OPCs, containing software zooms of individual complexes. A number of classes of complex were observed: templates were seen with 0, 1, 2 and multiple RNAPs bound. Complexes containing more than two enzymes arise from non-specific DNA-protein interactions and were not included in any analysis. A complex was denoted as an OPC if it contained two RNAPs positioned towards the centre of the template such that two arms were observed, and with a separation consistent with the expected separation for the inter-promoter distance ($\sim 1/3 \pm 10\%$ of the total contour length).

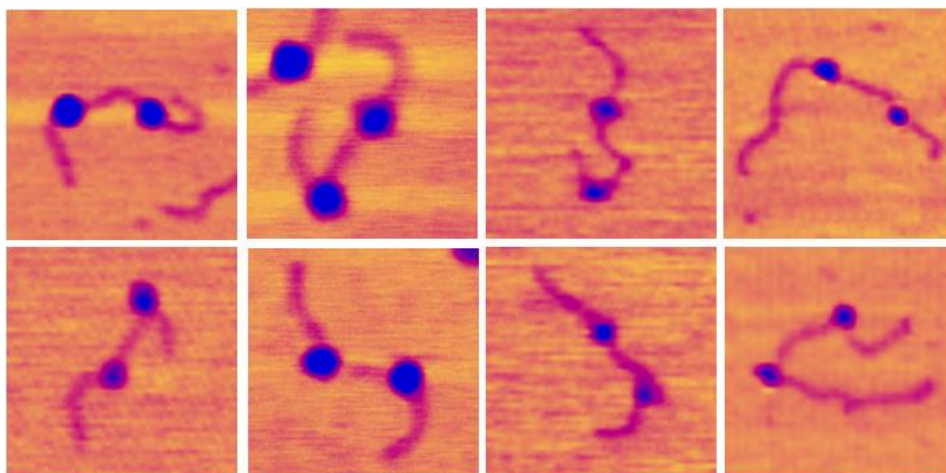


Figure 5-4. Montage of individual software zooms depicting typical open promoter complexes. The RNAPs appear as two globular structures situated in the central region of the template. They are separated by a small distance, consistent with the promoter separation. Images are 300 nm × 300 nm.

A comparison of the total contour lengths of templates with 0, 1 and 2 RNAPs bound as OPCs was undertaken. Bound RNAPs are expected to reduce the contour length by a characteristic amount, arising from specific contacts being formed between the DNA and protein upon OPC formation. Figure 5-5 shows the distribution of contour lengths for templates with varying numbers of RNAPs attached. The bin widths of the histogram were calculated automatically by the graphing software Origin, by taking into account the range of the data and the number of data points. These data sets were fitted to a Gaussian distribution, and it can be seen that as the number of bound RNAPs increases the contour length is reduced accordingly. The average contour length for the bare DNA fragment is 388 ± 3 nm ($n = 50$). This corresponds to a base pair repeat of 0.34 nm, and corresponds well with the value for B-form DNA. Templates with 1 and 2 RNAP bound showed a reduction in overall contour length to 358 ± 3 nm ($n = 50$) and 330 ± 4 nm ($n = 45$) respectively. This corresponds to an overall reduction of 30 nm for templates with a single RNAP bound, and 58 nm for those with two. It is interesting to note that the overall reduction in contour length is similar each time for each bound *E. coli* RNAP, which thus acts to reduce the contour length by a characteristic amount. The distribution of DNA contour length with two RNAPs bound has the greatest

range, and could indicate that if an RNAP binds first, it may alter the ability of a second RNAP to wrap the DNA normally. It may also be a surface binding effect.

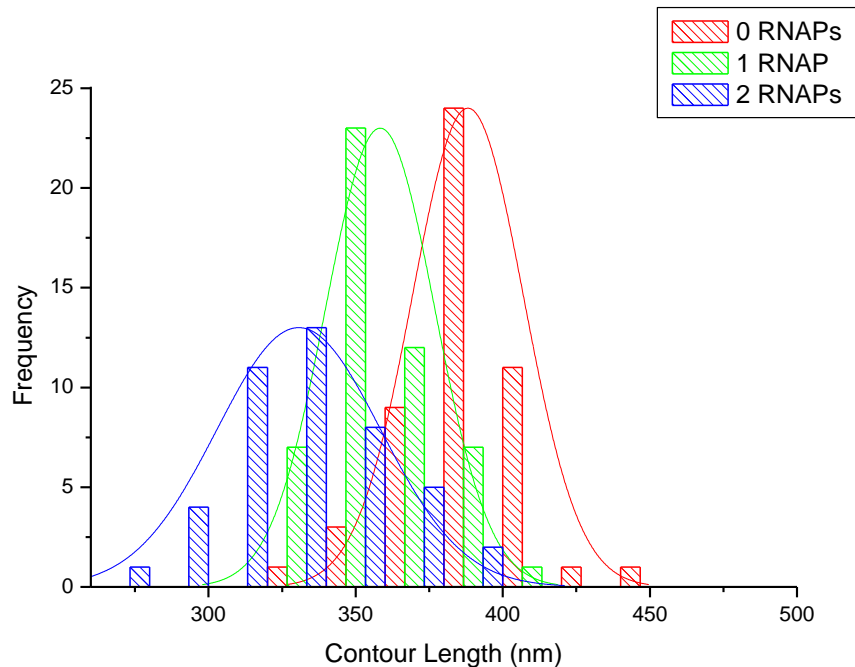


Figure 5-5. Compaction of DNA by RNAP. Contour length measurements for DNA with 0 (red), 1 (green), and 2 (blue) RNAP bound. Each promoter-bound RNAP reduces the contour length by a characteristic amount (~ 30 nm). This is consistent with a model in which promoter DNA is wrapped once around the RNAP holoenzyme core.

Observations of RNAP- σ^{70} OPCs with AFM have shown that the template always appears bent in the region bound to the polymerase [160, 168]. This is a consequence of protein-induced DNA wrapping. In the images of OPCs obtained with AFM the DNA often appeared to be significantly bent as it traversed around the RNAP, relative to downstream and upstream regions. The actual bend angle was quantified by drawing tangents to the incoming and outgoing regions of the DNA chain directly surrounding each RNAP for complexes with two RNAPs bound, in the manner displayed in Figure 5-6a. These results are plotted in Figure 5-6b. The average value was found to be $68^\circ \pm 4^\circ$. The distribution is broad, and could be a result of the finite tip size, broadening each RNAP making it appear larger than it actually is. This causes the obscuring of DNA directly neighbouring the RNAP itself, and any changes in DNA curvature in these regions would be missed, leading to inaccuracies in the bend angle

distribution. A large spread could also be a reflection of the natural variability of the complex.

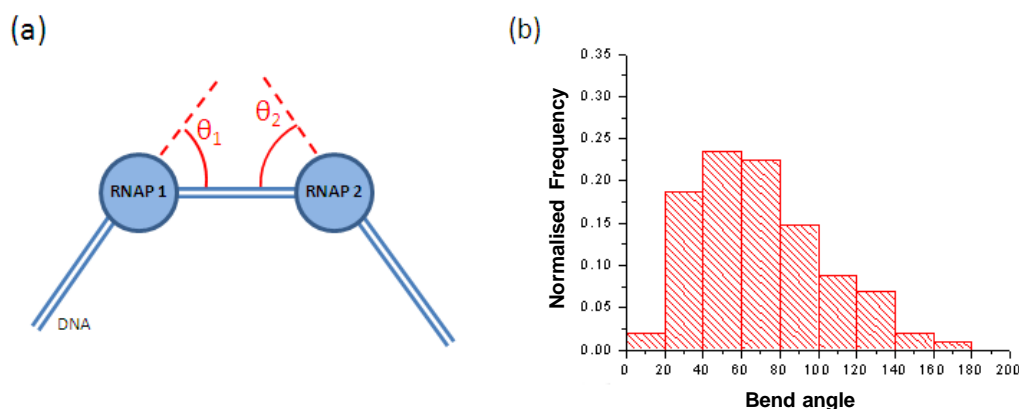


Figure 5-6. (a) Schematic diagram demonstrating how RNAP induced DNA bend angles were measured. (b) OPC bend angle distribution ($n = 92$). Whilst there is a large spread in bend angles (ranging from 0° to 180°), the distribution is centred near 60° with a standard deviation of 23.95° .

To further investigate this result, bend angles were measured from DNA chains containing a single RNAP. As the expected binding positions are known, once the average values (as measured from templates containing two RNAPs) for the long-arm and short-arm lengths are known, it becomes possible to classify singly-bound RNAPs as being specific or non-specific. RNAPs were identified as being specific if the distance from the protein to the closest chain end was typical of either of the average values for the arm lengths measured previously (91 nm for the short arm and 127 nm for the long arm, both $\pm 1\sigma$). These are expected to represent promoter-bound complexes. If the RNAP did not satisfy this criterion they were said to be non-specifically bound. These were found to be DNA templates with RNAPs either in the very central region of the chain, or very close to either of the two ends. Contour length measurements of the total chain length were taken to test this method of classification. If an RNAP is specifically bound, then one would expect DNA wrapping, indicated by a consequential reduction in contour length. This is indeed the case with the average value being $358 \pm 3 \text{ nm}$ for the specific case, compared with $375 \pm 4 \text{ nm}$ for non-specifically bound complexes. The specific complex compares well with the $\sim 30 \text{ nm}$ reduction in contour length per RNAP observed previously on

templates containing two RNAPs (one specifically bound at each promoter). It is interesting to note that the non-specific case displays a small reduction in contour length relative to free DNA, showing that it is still possible for RNAP to form some contacts with DNA. Any small bend angle observed could act to reduce the apparent DNA length, if the chain through the protein is not held in a 2D configuration and then obscured by the protein.

Figure 5-7 shows the bend angle distributions for both groups, in addition to software zooms of individual complexes typifying each class. What becomes apparent for the specific case is that the distribution is monomodal with an average at $67^\circ \pm 5^\circ$, a value that was roughly similar to that obtained from measurements of templates with two RNAP bound. This suggests that these group of complexes containing one RNAP were specifically bound at the promoter (OPCs). We see the appearance of a low angle spike for the RNAPs classified as being non-specifically bound, along with a general shift in the distribution to lower angles and an average of $38^\circ \pm 7^\circ$. It appears therefore that unspecific binding is typified by small protein-induced bend angles. On the other hand, in the case of RNAP that is specifically bound at the promoter region, DNA is made to wrap around the protein core leading to a characteristic bend angle.

It is therefore possible that the population of RNAPs exhibiting a small bend angle observed in the twin system may represent unspecifically bound protein misidentified as promoter-bound complexes. However, it is also possible that the presence of two RNAPs bound to two different promoters may lead to a situation whereby one enzyme exerts an influence on its neighbour. For example if one RNAP was to bind to the mica surface first it may limit the conformational freedom of the DNA surrounding the remaining RNAP once it becomes attached, leading to a change in DNA bending, especially once bound to the surface.

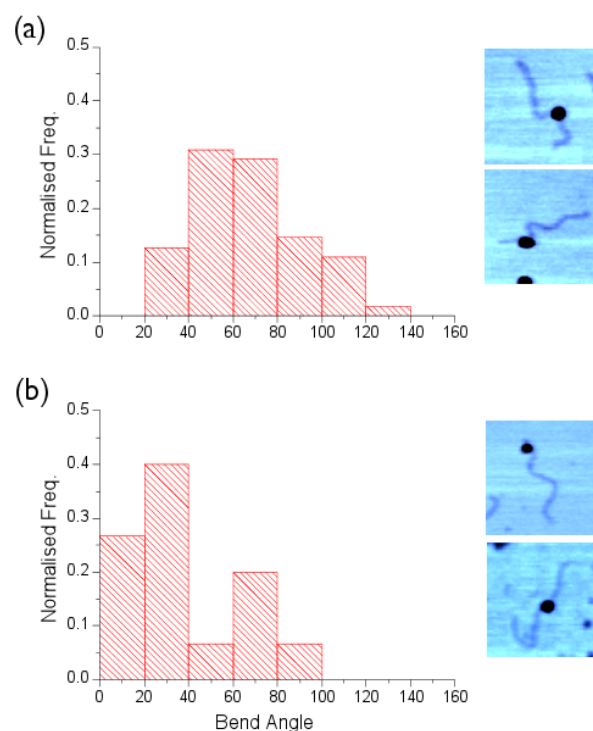


Figure 5-7. Protein-induced bend angles measured for (a) specific promoter-bound ($n = 55$) and (b) unspecific binding events ($n = 30$). Individual software zooms to the right of the graph show exemplar molecules belonging to each class. The unspecific molecules display a shift to lower bend-angle distributions. Images were taken in air, at 35 % RH.

5.3.2 RNAP size

Further confirmation of OPC DNA wrapping can be provided by measuring the effective size of the protein. If DNA is wrapped around the entire circumference of an RNAP then we would expect to see an increase in the size of the complex (see insets of Figure 5-8). Measurements of RNAP size were taken by thresholding the AFM images to remove the background and DNA chain, picking out only the protein in the Gwyddion image analysis software package. RNAPs bound to promoter DNA were picked out as grains after thresholding and their areas were calculated automatically by the program, and, by assuming that the protein was essentially circular, an effective radius was calculated. The same analysis was performed on RNAPs that were not bound to DNA (situated on the background substrate). The average radius for the DNA bound RNAPs was 18.9

± 0.3 nm, compared to a value of 14.3 ± 0.3 nm for the surface-bound RNAP alone (see Figure 5-8 for distributions).

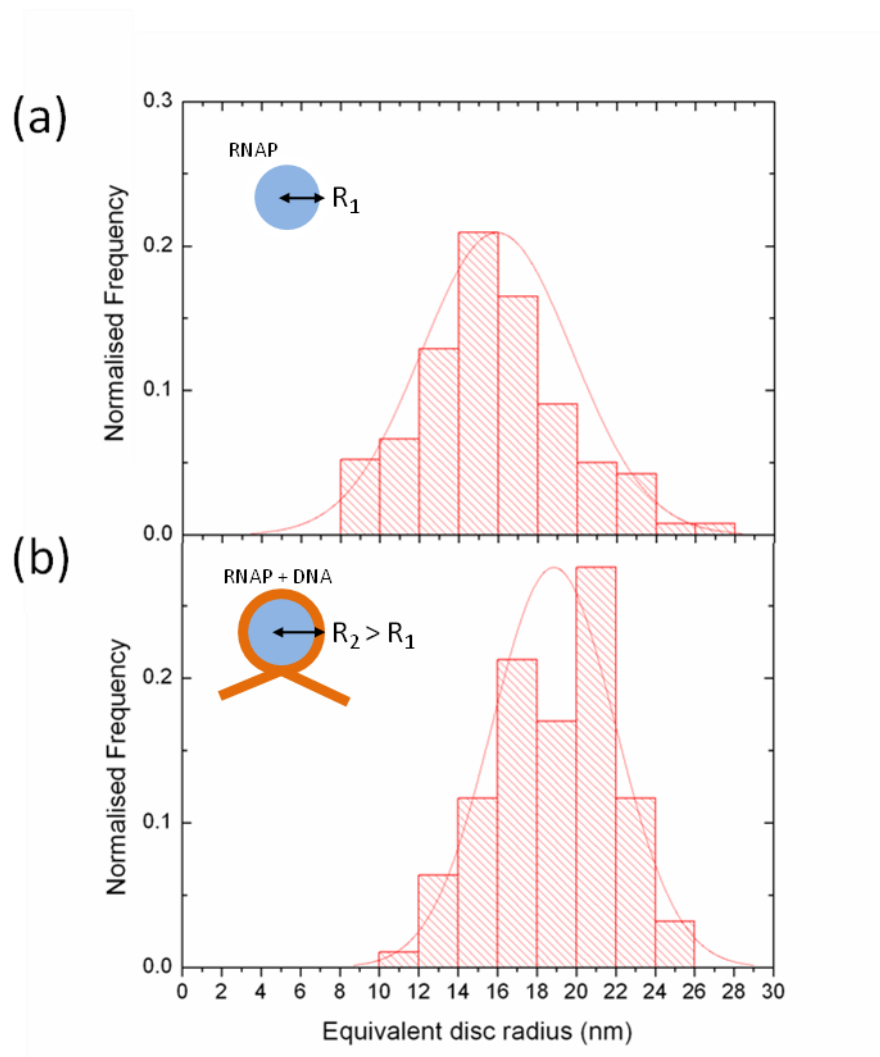


Figure 5-8. Equivalent disc radius for RNAPs that are (a) Bound to the mica substrate only, and (b) associated with DNA as an OPC. The RNAP has a greater effective size when it is associated with the template (see insets), as illustrated by a shift to a higher distribution. This suggests that DNA is wrapped around the RNAP core. Also shown in the inset of each graph is a schematic diagram of how DNA wrapping would affect the size of the RNAP.

There was no difference in size for RNAPs associated with the promoter belonging to the short arm and the long arm. These values for RNAP size are larger than values of RNAP structure obtained in crystallographic studies (~ 10 nm), but these discrepancies can be explained by the tip convolution effect. The RNAP associated with DNA was found to exhibit a shift to a larger radius when

bound to DNA (see Figure 5-8). The differences between the two values (4.6 nm) are roughly comparable with twice the expected diameter of B-form DNA (~2 nm). We might expect the extra radius to be exactly equal to the accepted size of DNA, but inaccuracies resulting from the automated analysis procedure, which assumes that the RNAP is a perfect sphere, may be responsible for the extra size. However, these results support the model that DNA is wrapped around the outer region of the RNAP.

5.3.3 Stalled Elongation Complexes (SEC)

Transcription was initiated using a sub-set of NTPs, in which ATP, GTP and UTP was added to a reaction mix containing already-formed OPCs on template pDSU. The absence of CTP means that the transcribing complex will stall at the first occurrence of the base cytosine. The stall distance is different for each promoter, being +24 and +70 nucleotides away from each transcription start site; however the base sequence is identical around each stall site. Hence, the stalling behaviour should be similar for each RNAP. Outwardly the appearance of SECs is not too dissimilar to that of the OPCs, with the template containing two RNAPs situated in the central region of the chain, and separated by a small distance (see Figure 5-9).

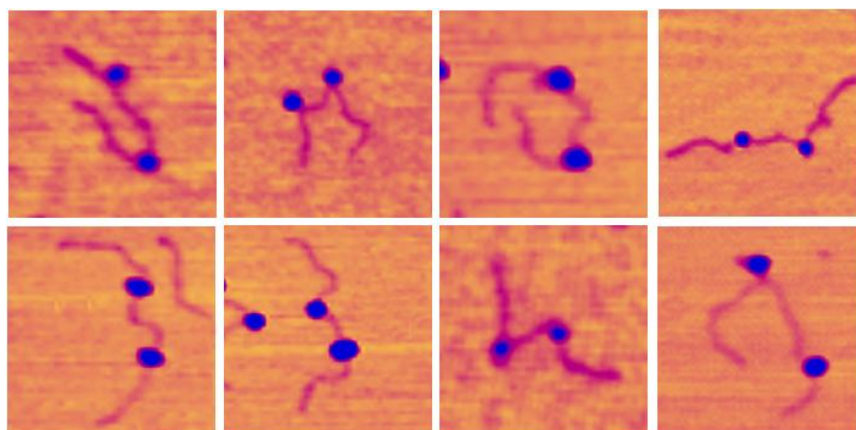


Figure 5-9. Montage of individual software zooms depicting typical stalled elongation complexes (SECs). With the naked eye these look similar to OPCs. However, contour length measurements indicate that the RNAPs are able to begin moving in the expected directions until they eventually stall. Images are 300 nm × 300 nm.

Measurements of the separation, in addition to the arm lengths, were taken, with a view to confirming that the RNAPs had travelled a distance coinciding with the stall length. The total contour length was seen to increase by 32 nm to 362 ± 3 nm, when compared with the OPC measurement of 330 nm. Relative to free DNA this corresponds to a reduction in contour length of 13 nm per RNAP. Additionally, the average values for both the short arm and the long arm were observed to have increased, whilst the inter-promoter distance decreased. The actual increase from the corresponding arms for the OPC case was similar for both the arm measurements (~ 20 nm), and the reduction in the inter-promoter distance was 10 nm.

The average protein-induced bend angle was seen to reduce slightly to $62^\circ \pm 3^\circ$, with a median of 54.9° (see Figure 5-10 also). However, what is apparent is that there is a significant population of SECs with little bend angle, with around 45 % of measurements being in the range 0-40°. Low bend angles were found to be an indication of unspecific complexes, as measured from OPC data (See Figure 5-7).

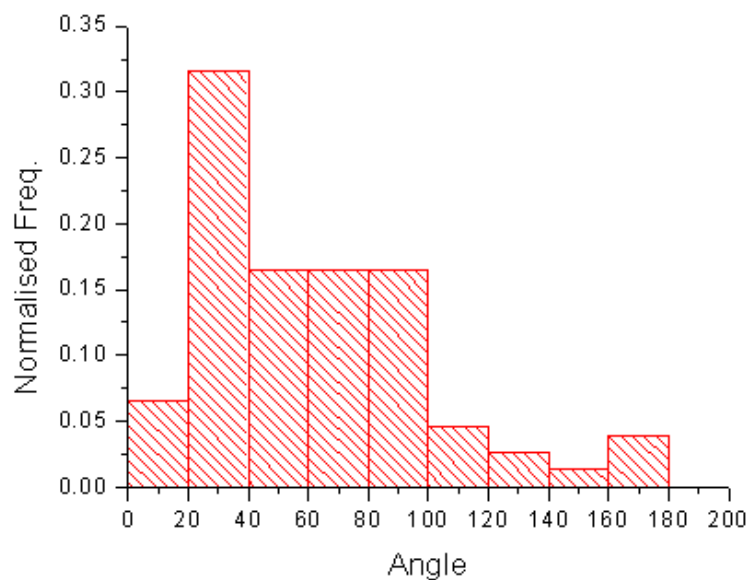


Figure 5-10. SEC bend angle distribution. Whilst the average bend angle is only slightly reduced when compared to the OPC case, there is a large shift towards smaller bend angles, with a particular spike being seen in the range 20-40°. This suggests that the DNA is not as extensively wrapped as in a promoter-bound complex.

The increase in total contour length, coupled with a reduction in bend angle, seem to indicate that the RNAPs are able to make the transition into an actively transcribing unit. The reduced wrapping can be attributed to the loss of contacts between the DNA and polymerase during the transition from initiation to elongation [168]. Significant sequence-specific contacts with the promoter region are essential for initiation, and are the main contribution to wrapping. It is believed that some of these interactions are lost upon promoter escape [54]. The apparent DNA bend angle is reduced as DNA wrapping is reduced, allowing the RNAP more freedom to move as it transcribes the chain.

The template contains two promoters on opposite DNA strands, and can be considered as a general model of nested genes found *in vivo*. A nested gene is completely enclosed within a larger host gene. Hence, there are two distinct promoter regions (one for each gene) both of which can recruit an RNAP and become transcriptionally active. Further evidence that transcriptional activity is possible on this model nested gene template is provided by measurements of the arm lengths. Both arms increased in size, whilst the distance between them decreases, consistent with the RNAPs moving away from their promoters and beginning to transcribe the chain in the directions expected. The average movements on each arm were quite similar. This is unexpected as one stall site is roughly three times further away from the transcriptional start site for the small arm than the long arm. However, the end-terminals of the DNA template are not labelled in these experiments so it is possible that the short-arm could be misidentified as the long-arm, and vice versa. This discrepancy could be a result of the fact that after the RNAPs move away from the promoter to the stall sites the differences between the expected lengths of the two arms are much smaller (463 for the long arm vs. 442 bp for the short arm). Assuming B-form DNA a difference of 21 bp would correspond to a distance of only 7 nm. The tip broadening affect could also have an influence, as RNAP appears larger. Considering the positions of the stall sites from the DNA sequence, a reduction in the inter-RNAP distance of around 31 nm (94 nt) would be expected. The lower than expected reduction (10 nm) could be explained by result of a change in the partition of wrapping between upstream and downstream DNA,

on maturation to an elongating complex. If the downstream DNA wrapped in an OPC was subsequently released on the transition to an SEC then the net effect would be an increase of DNA in the inter-RNAP region, meaning that the apparent reduction in the RNAP separation would be less than expected once transcription has started,

5.3.4 Collided complexes (CCs)

Collisions between transcribing RNAPs were induced in two different ways. In the first case, the missing NTP (CTP) was added to a mixture containing SECs, meaning that the collision is formed by a two-step process. This was expected to allow the stalled RNAPs to reinitiate, before they proceed along the template and experience a collision with another RNAP, or run off the end of the chain. The base sequence around the stall sites is identical in both cases, and as such that the behaviour of stalling and re-initiation should be similar for each promoter.

Transcription was also induced in a simple one-step process, in which all four NTPs were added to the reaction mix containing OPCs. In this case transcription should proceed unhindered past the stall site described previously.

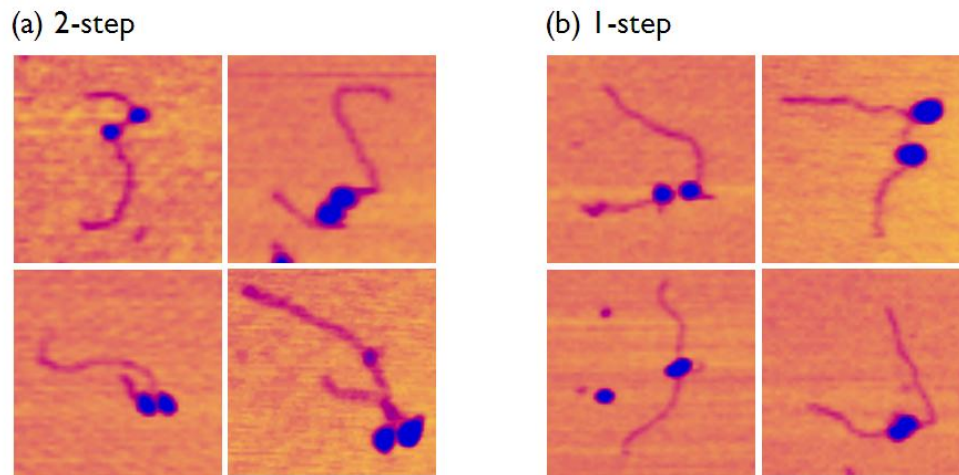


Figure 5-11. Montage of individual software zooms depicting collided complexes prepared via (a) a 2-step process with a stalled intermediate (b) a 1-step process. The RNAPs are expected to transcribe towards each other until a collision event occurs. In both cases, molecules containing two RNAPs in very close proximity to each other were often observed. Their separation is almost always much smaller than observed for OPCs and SECs. Image size is 300 nm × 300 nm.

Figure 5-11 shows a montage of individual complexes that represent the results of collisions between RNAPs initiated by (a) a two-step process, and (b) a one-step process. When imaging CCs formed using both strategies, templates containing two RNAPs bound in close proximity to each other were often observed, indicating that they may have collided. Also, the RNAPs were often situated asymmetrically towards one end of the template ($P = 0.55$), past the promoter at the short arm. These initial observations would appear to indicate that these complexes represent a distinct result of transcription interference, where RNAPs are observed in close proximity past one of the starting points, were not generally observed when imaging OPCs and SECs. The fact that RNAPs were observed close to the end of the template appears to indicate that they were able to backtrack along the template past their starting point at the promoter. It would seem that one particular effect of a collision is that the RNAPs stall against each other. That both RNAPs were still situated on the template after collision suggests that transcription of a stretch of DNA or a gene may be possible.

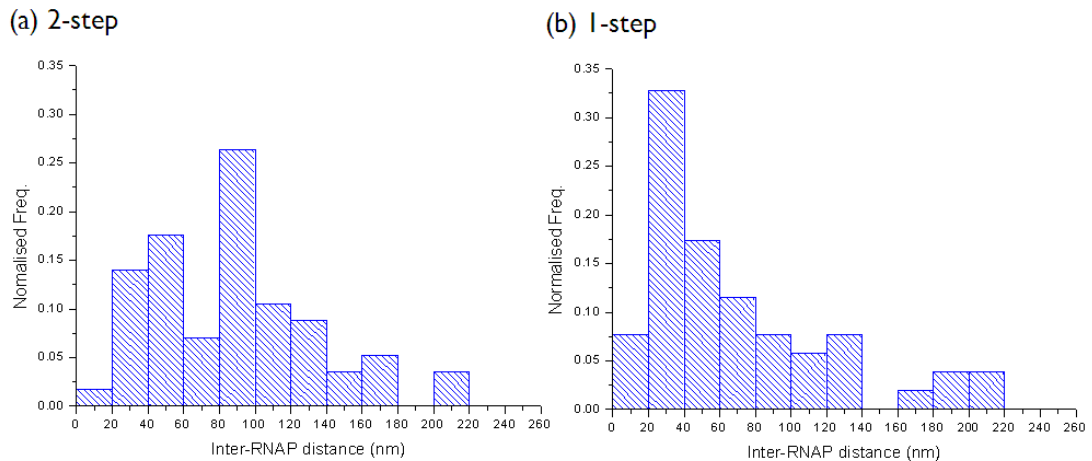


Figure 5-12. Inter-RNAP measurements from CCs prepared via (a) a 2-step process (n =67) and (b) a 1-step process (n =62). The distributions vary between methods. The 2-step process exhibits a large peak in the range 80-100 nm, whilst in the 1-step process there is a much greater proportion of complexes with a smaller separation. This difference is attributed to be a result of an inability of stalled complexes to re-initiate once the remaining NTP is added. Both methods show a broad distribution with a significant fraction at small separations.

To further quantify the effect of the collision and arrest of two convergently transcribing RNAPs, and to study differences between the 1-step and 2-step methods of CC formation, measurements were taken of the contour length along the DNA between the two polymerases (see Figure 5-12). Both distributions display a proportion of molecules that have a small separation. The overall range of distances is similar for both cases, suggesting that the behaviour of RNAPs after transcription has occurred is similar in both cases. However, there is a large peak centred in the range 80-100 nm for the 2-step case (Figure 5-12a), that is absent in the 1-step case (Figure 5-12b), where the majority of RNAPs are in close proximity to each other. This extra peak is consistent with the expected separation between the two stall sites, as calculated from the sequence data, suggesting that when CCs are formed by a 2-step process a significant proportion (~45 %) of the paused complexes remain unable to resume transcription, and become transcriptionally arrested. In the 1-step process, transcription is able to proceed unhindered past each stall site, and there are no significant peaks relating to either stalled or promoter bound complexes.

5.3.5 Visualisation of RNA

It was rare ($\sim 10\%$) that any nascent RNA was observed exiting the polymerase. RNA is able to form base pairs between different, distal regions of the same chain, allowing it to fold up into compact secondary structures. The nascent chain should remain in contact with RNAP during elongation and could perhaps become obscured by the broadening effect of the tip. However, a few images were obtained of structures that may represent RNA, in which small chains or globular features were observed to exit the RNAP in a different direction to the incoming and outgoing chain lengths. Only occasionally were such tertiary features observed exiting both RNAPs. A montage of these complexes, with the feature of interest highlighted, is shown in Figure 5-13. Structural information can be obtained by mapping the location of such features, relative to upstream and downstream regions of DNA. The RNA exit point was typically ($P = 0.81$) observed on the opposite side to the smallest angle formed by the incoming and outgoing DNA arms, as measured to be $62 \pm 3^\circ$. The average orientation of the transcript was measured, and it was found that the angle between the RNA and upstream DNA was equal to $140 \pm 20^\circ$ ($n = 25$). As such, we can suggest a structural model of the elongating complex (Figure 5-13b).

To gain more evidence that the features observed were of RNA transcripts the length of the feature protruding from RNAP was measured and compared to the RNAPs position on the template. The length of a transcript would be directly coupled to transcription and the distance that the RNAP had travelled from its promoter. However, in only 21% of cases was the length of the transcript similar to the distance (within 10% of the expected value) that the RNAP had travelled from its start site. Typically the feature had a smaller length than the distance of RNAP travel. The reasons for this are that the finite size of the AFM tip means the RNAP is seen to be broader, and creates a shadow that obscures the area surrounding the protein, including the RNA chain exiting it. Additionally RNA is able to fold up into more compact secondary structures. To examine this, the DNA sequence from the transcription start site at the

promoter at the long arm across the inter-promoter region was inputted into the RNA secondary structure prediction tool Mfold. Figure 5-13c shows the predicted structure that an RNA molecule would take up if it was formed from transcription across this particular DNA sequence. This shows that RNA transcripts formed after transcription across this template would be likely to fold up into a compact structure, formed of a number of DNA stems and loops.

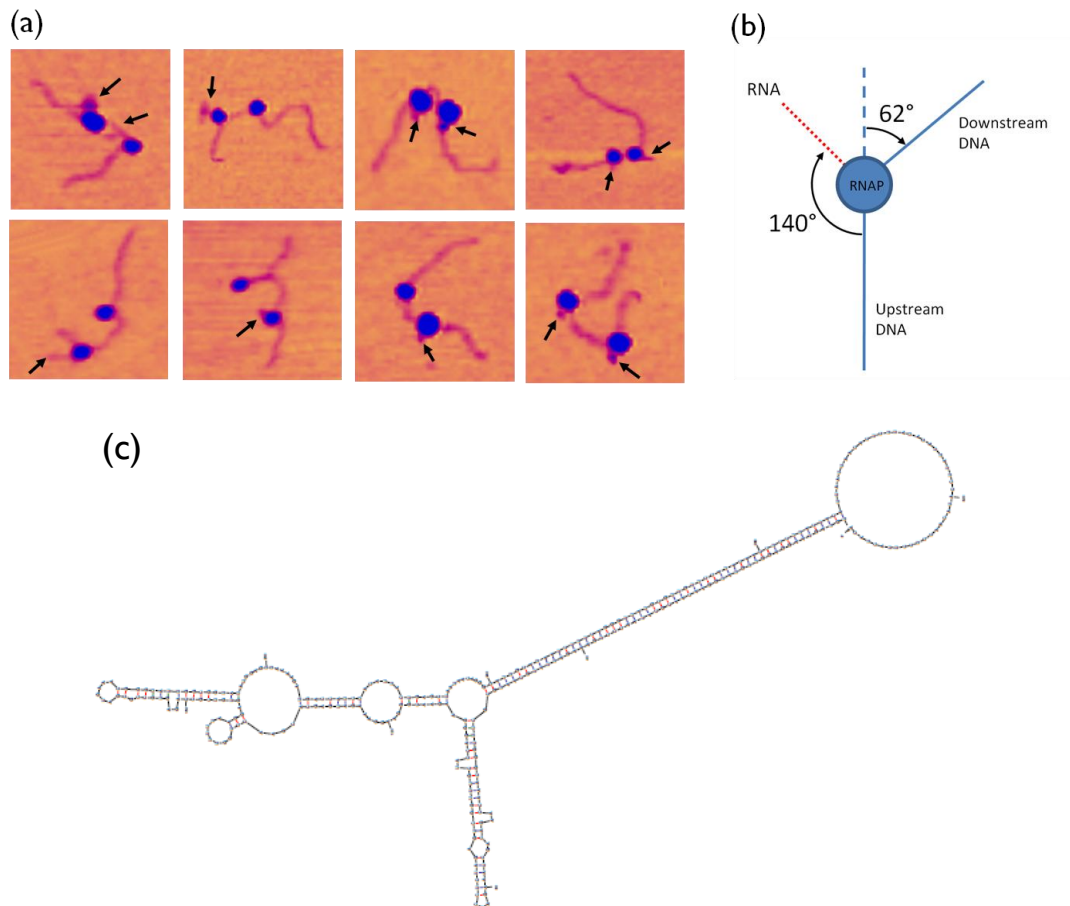


Figure 5-13. (a) Software zooms of DNA-RNAP complexes with putative RNA transcripts exiting the protein (indicated by arrows). (b) Model of the elongation complex. The transcript was typically observed on the opposite side to the upstream and downstream DNA arms. Images are all 300 nm × 300 nm. (c) Prediction of the secondary structure of RNA formed after transcription using pDSU as a template.

Another interesting class of complex was also observed, in which a transcript feature was hidden by a third RNAP which appeared bound in very close proximity to a template-bound protein (see Figure 5-14 for examples). Measurements of the angle between the centres of the two neighbouring

polymerases, one on the template and one off-axis, appeared similar to the relative position of the RNA exiting the transcribed RNAPs in the case described above ($128 \pm 5^\circ$, $n=22$). In addition to this, the additional RNAP was observed in the position expected of the transcript; i.e. on the opposite side to the smallest angle subtended by the two DNA arms. This is interpreted as free RNAP in solution binding to the RNA transcript once it is sufficiently large to be unrestricted by steric hindrance from the RNAP which has produced it.

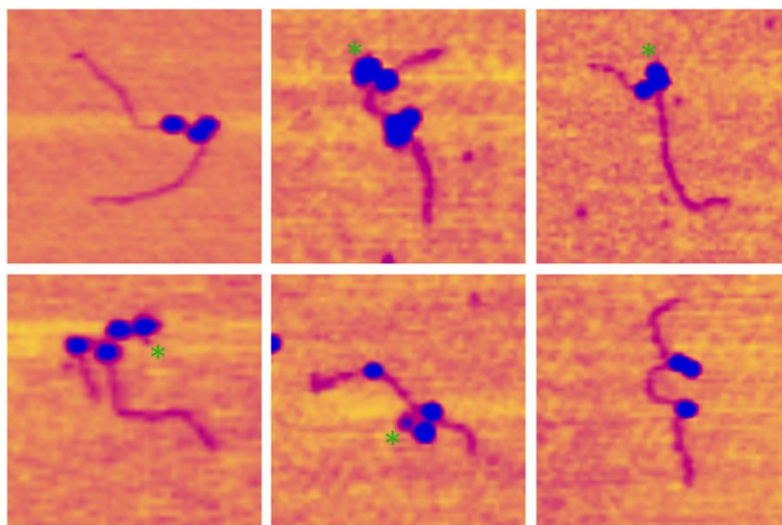


Figure 5-14. Examples of complexes that have an additional RNAP associated with a template-bound one. Measurements confirm that the position of the extra RNAP correlates with the position of the RNA, observed in other cases. On some complexes there even appears to be an apparent transcript exiting from behind the protein. These are marked by green asterixes. Images are all 300 nm × 300 nm.

5.3.6 Comparison of complexes

Complexes representing different stages of the convergent transcription reaction were formed and imaged using AFM. Contour length measurements of complexes with two RNAPs bound were performed, as were measurements of their relative positions on the chain. The average values of such measurements are displayed in Table 5-1. There are marked shifts in the values of RNAP separation for CCs compared to OPCs or SECs, demonstrating that the RNAPs are able to approach one another.

Table 5-1 Summary of measurements taken for pDSU template

Sample	Short-Arm (nm)	Inter-RNAP (nm)	Long-Arm (nm)	Contour Length (nm)	n
Open promoter complex	91 ± 3	113 ± 6	127 ± 3	330 ± 4	54
Stalled elongation complex	111 ± 2	103 ± 3	147 ± 2	362 ± 3	88
Collided complex	103 ± 5	69 ± 7	189 ± 7	361 ± 4	62

This method of analysis provides a “snap-shot” in time of what occurs at various events during the convergent transcription pathway. Table 1 only gives ensemble averages. To get a clearer idea of what occurs during these “collision” events, it becomes important to study individual members of the population. The AFM makes it possible to observe any sub-populations or intermediates formed in a given process. To gather information on what occurs during the collision, it becomes necessary to study how the distributions of the measurements as a whole change during the reaction. From length measurements along the DNA contour it is possible to infer how the position of each RNAP changes relative to the other RNAP. Figure 5-15 shows the distributions of the arm-length and inter-RNAP measurements for each kind of complex (OPC, SEC and CC (1-step)) to facilitate comparison.

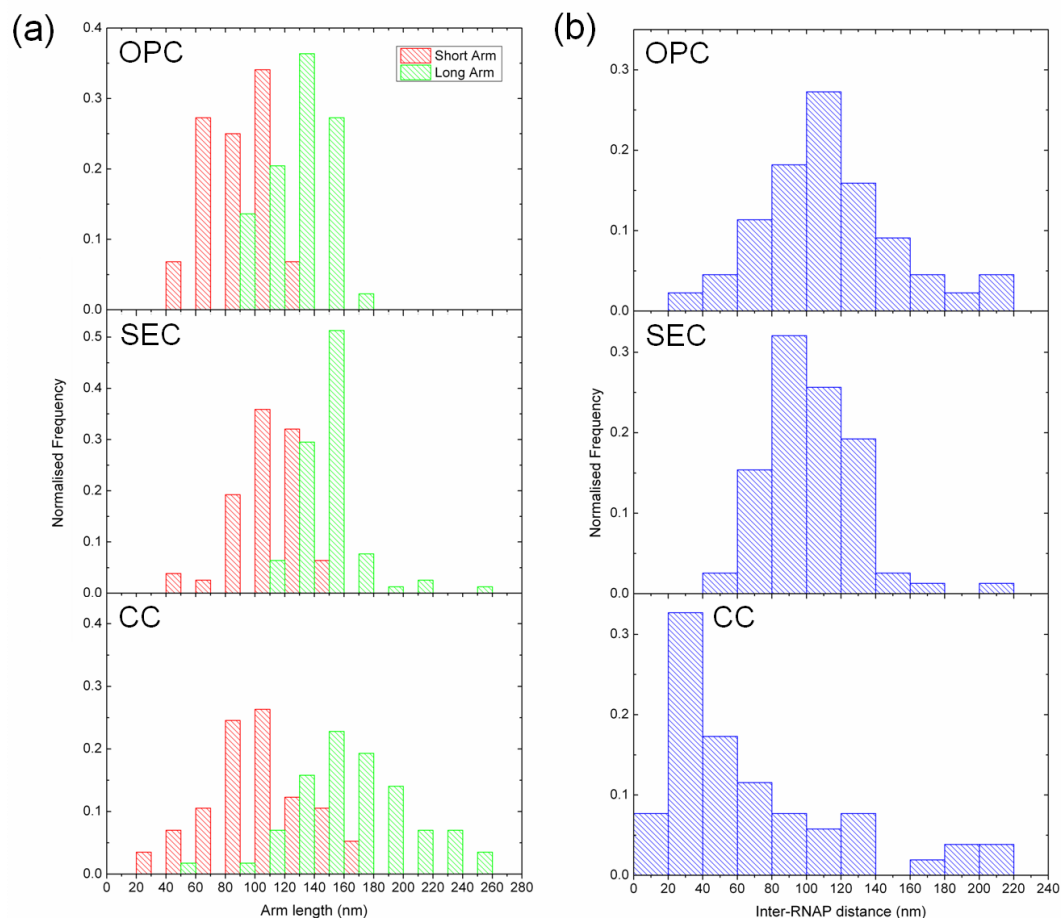


Figure 5-15. Evolution of RNAP position and separation in convergent transcription. (a) pDSU arm length contour distances for OPC (n = 54), SEC (n = 88) and CC (1-step reaction) (n = 62). (b) pDSU inter-RNAP distance contour measurements for the same complexes.

5.3.7 Inter-RNAP separation

The transition from OPC to SEC is accompanied by a small reduction of the inter-RNAP distance (compare separations in Figure 5-15b). This is indicative of the RNAPs being able to begin transcribing and moving away from their promoters, but then stalling when the first occurrence of the missing base is encountered. However there is a marked change on addition of all four NTPs, since the elongation stage can proceed as normal (CCs in Figure 5-15b). In this case, what becomes most apparent is that there is a major shift towards complexes that have much smaller separations. In Figure 5-15, the CC distribution represents collisions initiated via a 1-step approach. However, such shifts towards smaller separations were also seen in the 2-step case, together

with an additional peak (80-100 nm in Figure 5-12a), referring to complexes which had failed to re-initiate after stalled intermediates had been formed. The large shift towards a smaller separation on addition of all four NTPs, suggests that after a “collision” event, the two RNAPs move closer, a situation that appears to be indicative of stalling, induced by the presence of another RNAP.

5.3.8 Arm-lengths

Both arm lengths show an expected increase in magnitude once a subset of NTPs is added to the reaction mix, and the RNAPs transcribe towards each other (Figure 5-15a). However, once all four NTPs are available to participate, and presumably a collision occurs, there are significant shifts in both measurements in complementary directions. (Compare Figure 5-15a OPC and CC showing that the overall range of the measurements taken together increases). That is to say, whilst the long arm moves to higher values, the short arm begins to shift towards lower values. A significant amount of short-arm measurements are even smaller than the initial short-arm promoter length. Additionally, the distributions in the CC arm-lengths begin to overlap each other more, confirming that RNAPs are travelling towards each other. A lengthening of the long arm occurs for many complexes, sometimes to distances greater than the separation between the two promoters, and would suggest that one RNAP can transcribe across the full inter-promoter distance without being blocked by the other RNAP, i.e. one RNAP can travel continuously in the 5' → 3' direction. The reduction of one arm suggests that a head-on collision can lead to the backtracking of one RNAP by the other, and that this can occur even past its original starting point at the promoter. To summarise, the results from the arm-length distributions, coupled with significant decreases in the inter-RNAP distances puts forward a model in which collision results in the backtracking of one enzyme by the other, and the subsequent stalling of the forward moving RNAP.

5.4 Tandem pDSP template

A more common gene arrangement than the nested gene is to have genes overlapping on the same strand of DNA. To examine the differences between convergent transcription arising from oppositely aligned promoters, where collisions are head-on, and the more frequent arrangement, in which promoters are aligned in the same direction, a different but related template was used. The pDSP template contains two identical λ_{PR} promoters, similar to the pDSU template (see Figure 5-16), however, the promoter belonging to what is termed the short arm is flipped around. Hence, two RNAPs transcribing from the different promoters can travel in the same direction and use the same DNA strand as a template for RNA production.

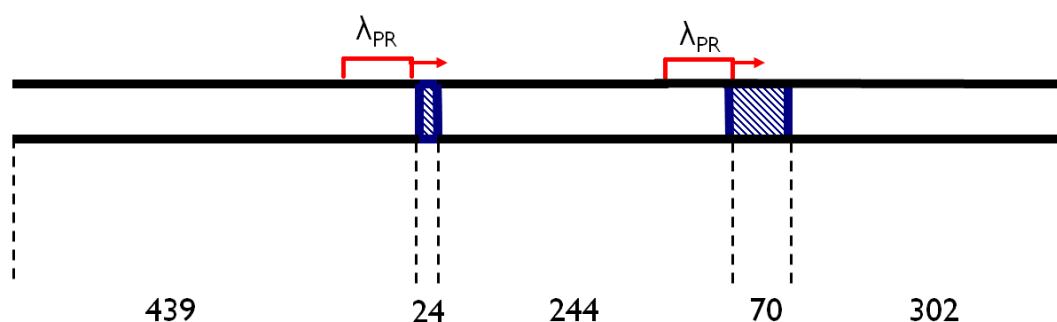


Figure 5-16. pDSP template. This contains two λ_{PR} promoters in a tandem arrangement. The blue-hatched region shows DNA regions deficient in cytosine.

Such an arrangement negates transcription interference (TI) from direct head-on collisions. Rather TI relating not only to promoter occlusion, but also rear-end collisions may become significant. This alternative arrangement allows us not only to compare two collision mechanisms (head-on vs. rear end), but also to test our previous models of backtracking. If both promoters were to “fire” during a short time interval, then there would be no collision and both enzymes are expected to run off the end of template.

5.4.1 Initial Imaging of pDSP complexes

DNA-RNAP complexes were formed in an identical manner to those in the convergent case: by varying the reagents available to participate in the reaction different stages of transcription could be imaged. A montage of single-molecule software zooms of pDSP complexes representing OPCs, SECs, and CCs is shown in Figure 5-17. Visually the main classes of complex look similar to their pDSU counterparts. Also, as before, different populations were observed containing 0, 1, 2 or multiple RNAPs bound. OPCs and SECs appeared outwardly similar with two RNAPs situated roughly centrally on the template and separated by a small distance. CCs were prepared via the one-step approach, and typically appeared with two RNAPs situated toward one end of the template, and with a much smaller separation.

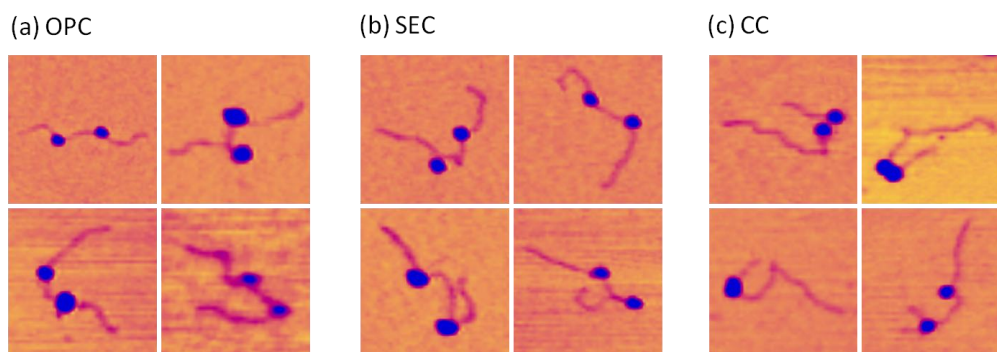


Figure 5-17. Montage of single-molecule software zooms of RNAPs bound to pDSP template representing (a) Promoter bound RNAPs (OPCs), (b) Stalled RNAPs (SECs), and (c) Collided RNAPs (CCs). The different stages appear outwardly similar to the convergent pDSU case. Images are 300 nm × 300 nm.

5.4.2 pDSP results

An analysis was performed as before on DNA molecules where two RNAP remained bound, measuring and comparing values for the short-arm, long-arm, and RNAP separation. Table 5-2 summarises the average values for each case of complexes. As expected, due to the promoter sites being in identical positions, measurements relating to the OPC class appeared quite similar to the pDSU case.

Table 5-2 Summary of measurements for pDSP template.

Sample	Short-Arm (nm)	Inter-RNAP (nm)	Long-Arm (nm)	Contour Length (nm)	n
OPC	93 ± 4	106 ± 5	138 ± 4	337 ± 4	48
SEC	91 ± 4	114 ± 5	154 ± 6	359 ± 7	54
CC	76 ± 5	74 ± 8	220 ± 8	368 ± 5	65

With regard to the pDSP template, the short-arm promoter has shifted position such that the overall arm lengths should remain the same as in the convergent pDSU case. The measurements of the OPC short arm show a general agreement with this; however, the long arm appeared slightly larger than expected. This appears to be partially taken up by a small increase in total contour length of around 7 nm. This suggests that the direction of the promoter, and therefore the extent of wrapping, may lead to different DNA structures or binding behaviour.

Once a sub-set of NTPs are added, both arm lengths shift in the directions expected. Taken with the fact that the inter-RNAP distance is increased (as expected by the positions of the stall sites), together with an overall increase in total contour length, these results suggest that the RNAPs are able to begin transcribing the pDSP template. However, there is some discrepancy in the magnitude of the changes. With stall distances of 24 and 70 nucleotides, corresponding to distances of around 8 nm and 23 nm, for the long and short arms respectively, one would expect the short arm to show the greatest shift. However the opposite is true. This could be explained by considering how the wrapping is partitioned between upstream and downstream DNA. If there was a change in this partition corresponding to structural changes between an initiation and an elongating complex, such that there was increased wrapping of downstream DNA relative to upstream DNA, then such results could be

explained. There is an increase in total contour length, indicating that the overall degree of wrapping had been altered on transition to an elongating complex.

All four NTPs were added to the reaction mix to form collided complexes via a one-step method. With both promoters being situated in the same direction one might expect the complexes to simply run off the end of the template. However we observed DNA templates that still had both RNAPs attached. These were often situated in close proximity, and towards one end of the template. The overall fraction of “naked” template, without an RNAP bound, was found to increase, from 29 % for the OPC mix (n = 234) to 42 % for CCs (n = 266). Although these figures may be dependent on sample preparation, it appears to indicate that there is a likelihood of both RNAPs running off the template, leading to a decrease of DNA with RNAPs bound, and an increase in bare DNA.

Due to the tandem promoter arrangement it is possible to study how transcribing RNAPs interact with each other, without some of the ambiguity in not knowing which chain end is which on an unlabelled template. It is expected that movement away from these two promoters would typically lead to an increase in the long-arm distance, and a decreasing short-arm. Any movement in the opposite direction is expected to be highly improbable, being against the 5' to 3' direction required for transcription. Any movements in this direction should relate to non-specifically bound complexes. After collision we do indeed observe the expected changes in magnitude of arm length, the averages in the short arm and long arm lengths changed to 76 ± 5 nm and 220 ± 8 nm, which represent decreases and increases of 19 nm and 82 nm respectively.

5.4.3 pDSP arm-length measurements

Looking at the distributions throughout the process can elucidate how the RNAPs behave at each stage, and what are the potential collision outcomes. Figure 5-18a shows how the distributions of the long and short arms change throughout the reaction. In the OPC case, there is a narrow distribution centred around the average value, consistent with RNAPs situated at their respective promoters. Once a sub-set of NTPs is added, both distributions shift their positions slightly, with a decrease in the short arm, and a more marked increase in the long arm. This is representative of RNAPs travelling to their stall sites. Once all four NTPs are available, then transcription can proceed unabated, and the arm-length distribution displays a corresponding shift. Similar to the stalled complexes, the short arm decreases whilst the long arm increases, however, these overall shifts are much greater. This is expected and coincides with the promoter arrangement, as both RNAPs transcribe toward the same template end.

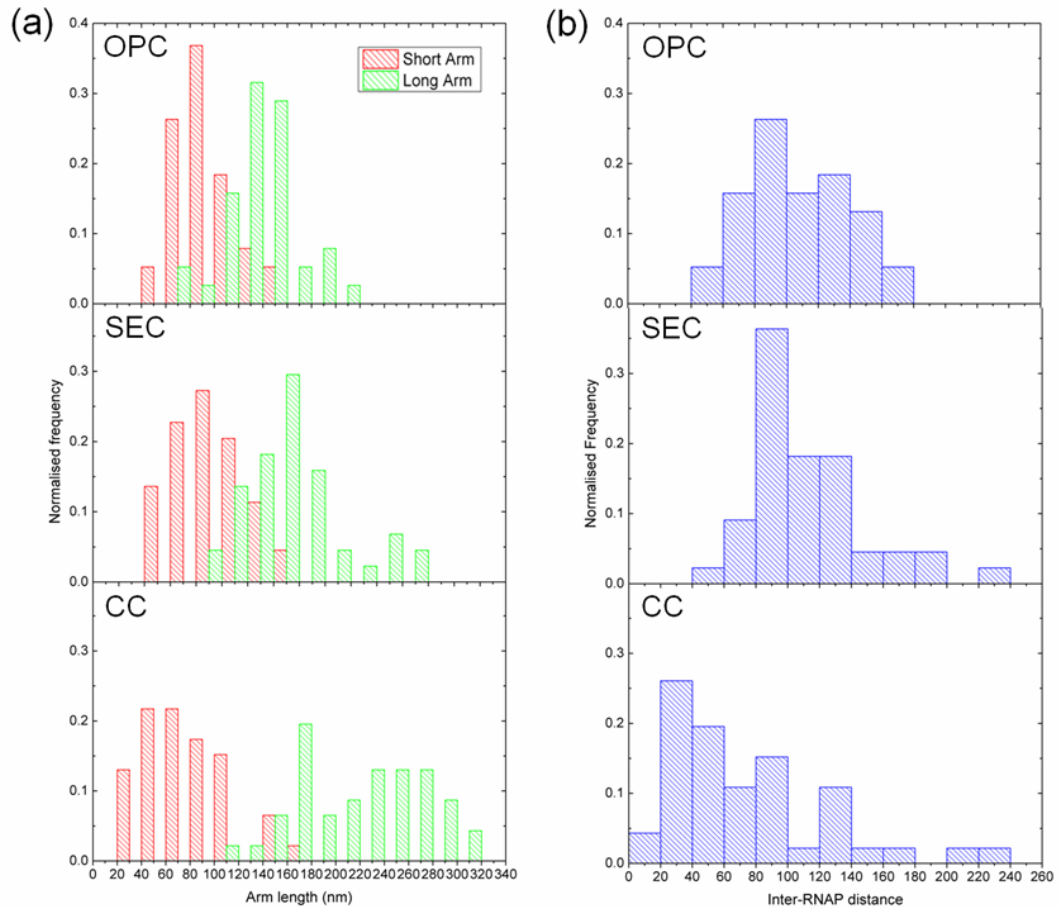


Figure 5-18. Evolution of RNAP position and separation in transcription originating from tandem promoters (a) pDSP arm length contour distances for OPC (n = 48), SEC (n = 54) and CC (n = 65) (1-step case). (b) pDSU inter-RNAP distance contour measurements for the same complexes.

Rather than being a simple normal distribution the long-arm distribution contains a number of different peaks, of which there are two main ones. These appear to correspond to different populations, and therefore potential collision outcomes. The peak around 160 nm in the CC (1-step case) long-arm distribution appears to coincide with a similarly large peak in the OPC case, suggesting that they may represent complexes that have failed to come out of the initiation stage. The peak around 260-270 nm in the long-arm almost corresponds exactly to a scenario whereby an RNAP has “fired” from the long-arm promoter, and has travelled a distance consistent with the initial inter-promoter distance to where the short-arm promoter is situated (a distance of 777 bases from the opposite chain end or ~ 251 nm). With regard to the short

arm there is a single peak centred around 50 nm indicating a simpler behaviour, but also showing that not all RNAPs transcribe off the end of the template.

Hence, the distribution for pDSP displayed different characteristics to the pDSU template. In the convergent case, the distributions increased in width but remained Gaussian. Additionally, whilst the two arm length distributions begin to overlap, approaching similar values in pDSU after collision, for pDSP the two began to spread apart. This is expected for this different arrangement, and confirms that our transcription reactions are proceeding as expected. The difference between the two average arm lengths for CCs in pDSP is 144 nm, compared to 86 nm in pDSU. However, it appears that the outcomes of each interaction are similar for each case, with two RNAPs remaining on the template in close proximity.

5.4.4 pDSP Inter-RNAP separation

The distributions of the RNAP separation for pDSP also showed differences to pDSU. From OPC to SEC, the distribution displays a small shift towards higher values, corresponding to the slight increase in separation once the sub-set of NTPs is added. Once all four NTPs are available to participate in the reaction to form CCs, there is a marked shift to lower separations. However, whilst in the convergent pDSU case there was one main peak at quite small separation, there were a couple of other noticeable peaks for pDSP. The largest peak (20-40 nm) is indicative of RNAPs that have collided with each other and subsequently stalled roughly next to each other.

Of the other peaks in pDSP, one is centred in the range 80-100 nm, which is roughly consistent with the initial inter-promoter distance. This class of complex most likely represents a sub-population in which both RNAPs have failed to enter the elongation stage. In a comparison with the other template, the fraction of complexes separated by a distance typical of the promoter separation was around 0.20 for pDSP, and 0.075 in pDSU. Hence, it is almost three times as likely to find RNAPs at the two promoter sites, in the tandem case as opposed to the convergent arrangement. This could confirm that with the

two promoters situated in the same direction the RNAPs have a greater chance of running off the end of the template. As such, in pDSP RNAPs failing to elongate become a greater proportion of the overall population of templates which have two polymerase bound, and are subsequently analysed.

A third peak in the range 120-140 nm is perhaps a result of RNAPs originating from the long-arm promoter stalling. Stalling appears to be a major factor in our *in vitro* approach. With this particular promoter arrangement if there was never any stalling, either due to collisions or intrinsic pausing, then we would expect both RNAPs to finish transcribing the template, and to only observe bare DNA templates. It is proposed that this increase in separation is a result of such stalling and transcriptional arrest. An increase in separation would require that the RNAP at the short arm was able to initiate and move away from its promoter. If the remaining RNAP situated at the remaining promoter either failed to initiate or stalled downstream of the transcription start site than an increase in separation would be observed.

Such a noticeable peak, at an increased separation, is not present in the same measurements for pDSU, a consequence of the convergent arrangement meaning that RNAPs should always be travelling towards each other. It appears that a stalled RNAP acts as a barrier to another elongating RNAP. There is a collision and the transcribing polymerase either stalls, or induces backtracking upstream in the other RNAP. It should be noted, that if the RNAP at the long-arm promoter of pDSP is stalled or arrested there will never be a collision with the other RNAP, owing to the promoter directions.

5.5 Discussion

5.5.1 Wrapping

When imaging templates with RNAPs bound it became apparent that the total contour length was reduced, relative to free DNA, by a characteristic amount. By comparing the contour lengths of DNA with one ($n = 50$), or two ($n = 48$) RNAPs bound, the reduction in contour length was found to be around 30 nm per enzyme. The values were also similar for pDSP (27 nm) and pDSU (29 nm), indicating that it is not template specific. These observations are consistent with a model whereby promoter DNA is wrapped around the RNAP. When just a single RNAP was associated the difference was reduced by about half, indicating that each RNAP can wrap a region of DNA independently. Rivetti *et. al* also performed a detailed analysis of the contour lengths of many DNA molecules and found that in the case where RNAPs were bound to the promoter in an open conformation, the length of the fragment was also reduced by ~ 30 nm (~ 90 bp) [168]. Crystallographic studies have shown that E. Coli RNAP has a globular structure with a circumference of ~ 32 nm [55]. As such, a reduction of such magnitude is consistent with DNA being wrapped once around the protein surface. Such a process is similar to how DNA is packaged by nucleosomes, by wrapping around a histone core. Such nucleosome-like structures, whereby OPC wrapping is particular obvious, are shown in Figure 5-19.

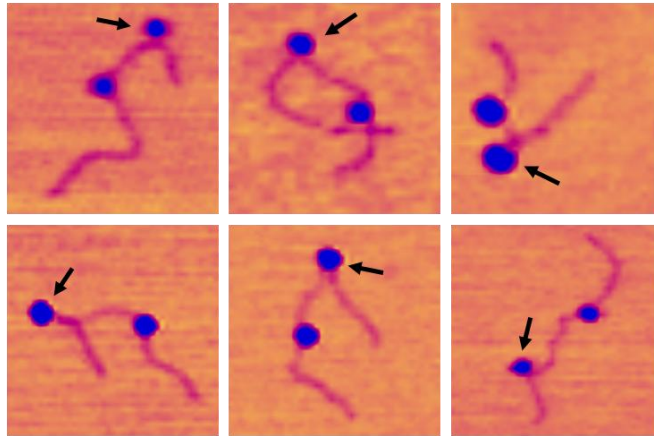


Figure 5-19. Montage of OPCs. The shape of these complexes suggests a structure in which DNA is completely wrapped around RNAP, in a similar way to a nucleosome (Indicated by arrows). Each image is 300 nm × 300 nm.

The wrapping of DNA around RNAP also leads to the bending of upstream and downstream DNA regions by characteristic values. The average bend angle values were centred in the range $60^\circ - 70^\circ$, for both pDSU and pDSP. Similar values were obtained for both promoters, and for templates with only one specifically-bound RNAP, again suggesting that the RNAPs do not influence each other at this stage. Rivetti measured bend angles that were centred from $55^\circ - 88^\circ$, and suggested that such a broad distribution reflected complexes with well-defined bends being adsorbed to the surface in slightly different orientations. The fact that the DNA is wrapped means that the overall DNA distortion is much larger, and the DNA is actually bent through 300° (see Figure 5-20a).

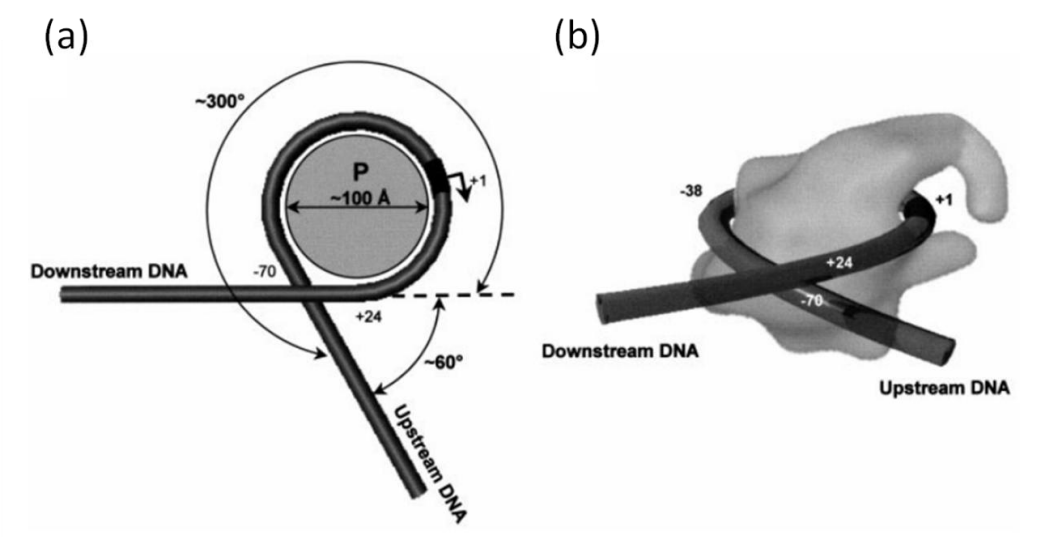


Figure 5-20. Model of the E.coli RNAP OPC proposed by Rivetti *et al.* [168] (a) In their study a bend angle of $\sim 60^\circ$ was measured by AFM and gel retardation. DNA wraps around the outside of the enzyme. (b) Three-dimensional representation drawn according to the crystallographic results of Darst *et al.* [55].

Similar results, pointing towards DNA wrapping around the RNAP core, have been obtained through DNA foot-printing experiments, in which it was observed that the length of protection of the DNA by RNAP was similar to the reduction in contour length observed by AFM [45, 177]. Rivetti used templates with an asymmetric promoter location, meaning it was possible to distinguish between upstream and downstream regions of DNA. By comparing the lengths of the two arms with the values expected for the template, they were able to conclude that two-thirds of the wrapped DNA was attributed to the upstream region, and one-third to the downstream. As such, the transcription start site is located near the cleft produced by the thumb-like structure in RNAP, a region that has been suggested to contain the active site of the enzyme (see Figure 5-20b) [55, 156]. We observed differing proportions of wrapped DNA upstream and downstream of the RNAP for each promoter. This may be a net effect of having two RNAPs wrapped around a single template. If one RNAP wraps its DNA first it may partially inhibit the second RNAP from being able to manipulate the DNA maximally, by providing a significant drag force.

DNA wrapping increases the extent of the RNAP-DNA interactions by maximising the contact area between the two. Such interactions are important

for complex stability. The initial RNAP state is known as a closed complex, and involves the wrapping of DNA around the protein in a left-handed fashion. This can propagate negative writhe in the template. It has been suggested that the negative writhe associated with the complex is converted into localised untwisting in the promoter region [6]. This results in unwinding of the two strands, a necessity for transcription initiation.

With regard to the OPCs, two main classes of conformation were observed. The conformation of the molecules may be affected through being bound to the surface. In the case of our template, where two RNAPs are situated at separate promoters, it is expected that the two arms exiting the RNAPs are likely to lie in different planes. During binding the complex effectively undergoes a transition from being a three-dimensional object to one that is flat, lying in a single 2D plane. The transition into a surface-bound complex was found to separate the molecules into two main conformation classes, U-shaped in which the arms lie in the same sense, and S-shaped when they are in the opposite sense (see Figure 5-21). It is interesting to note that in both the convergent pDSU case and the tandem pDSP case, the respective frequencies of each group is almost identical, with around a third being S-shaped and two-thirds being U-shaped.

These two shapes could represent different isomers, in which the DNA structure is changed in the inter-RNAP region in the transition to two dimensions. B-form DNA is right handed and as such DNA twisting in an anti-clockwise fashion would lead to the unwinding of the helix. As such the formation of U-shaped conformations would result in the under-winding of the helix, whilst S-conformations will over-wind the helix (Figure 5-21 inter-RNAP regions). This would translate to a small amount of negative or positive supercoiling respectively. The unwinding of the helix is likely to aid the formation of the transcription bubble and hence stabilise the OPC. This can explain why the U-shaped conformation was more common.

Additionally, the different promoter arrangements of pDSU and pDSP may affect how the DNA is able to wrap around the DNA (see schematics in Figure 5-21). In an OPC, DNA is wrapped around the protein in a left-handed fashion i.e. anti-

clockwise. In the pDSU template containing convergently aligned promoters, the two OPCs wrap the DNA in opposite fashions relative to each other; in which case, the S-shape will have more extensive wrapping, whilst wrapping would be reduced in the U-conformation. The opposite is true for the tandem pDSP case, where each promoter wraps in the same direction. Here, S-conformations have less extensive wrapping than U-conformations. The more DNA-protein interactions there are, the more stable the OPC is expected to be.

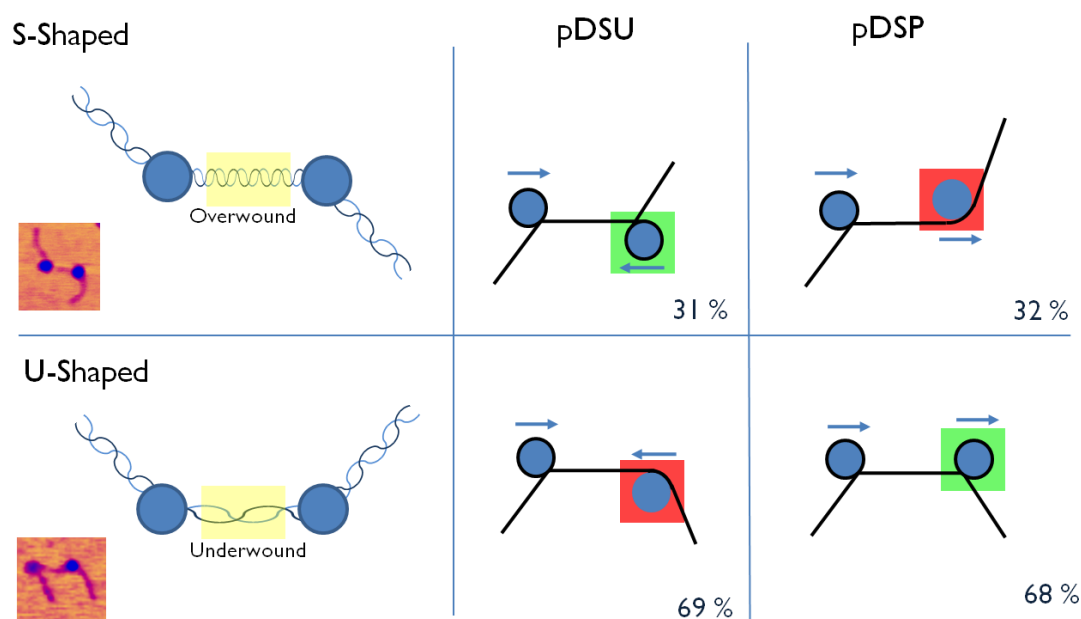


Figure 5-21. Examples of main conformations of OPC seen: S- and U-shaped. On transition into 2D upon surface-binding, the different conformations will affect the inter-promoter DNA differently. OPCs wrap DNA in a left-handed fashion, meaning the structures formed lead to either an increase or decrease of twist. The different promoter arrangements of pDSU and pDSP mean that S- and U-conformations affect how extensively DNA is wrapped differently for each template. Extensive wrapping is indicated by a green box, whilst a reduction is shown in red. Also shown are the relative occurrences of each conformation for each template.

Measurements of the contour length for each group in pDSU support the hypothesis that the extent of DNA wrapping is coupled to the conformation the complex takes up on the surface. For pDSU the U-Shaped conformation had an average contour length of 334 ± 5 nm, whilst the S-shaped conformation had an average of 323 ± 4 nm. The lower length would be an indication of the increased wrapping of the S-shape. However, the values for each group in pDSP were

similar, with average contour lengths of 337 ± 5 nm and 334 ± 4 nm for the U- and S-conformations respectively. The fact that both values are similar may be a result of insufficient imaging resolution.

The fact that the relative frequencies of each type of shape is almost identical for pDSU and pDSP suggests that the different stabilising and destabilising effects can compensate for each other. The structure that the inter-RNAP region takes up may be more significant than the wrapping for effective stabilisation of the complex.

On maturation into an actively transcribing elongation complex some of the initial wrapping is lost. SECs with two RNAPs bound show increases in total contour length of 32 and 22 nm's for pDSU and pDSP respectively, over their OPC equivalents. However, free DNA was still observed to have a higher total contour length than SECs, in which reductions of around 13-15 nm per RNAP were observed. Rivetti *et al.* observed a slightly higher compaction of 22 nm when studying SECs of similar template length with AFM [166]. Our results do agree better with DNA foot-printing experiments, which indicated that the interaction involved no more than 35 bp (~ 12 nm) [132, 134]. The reduced wrapping, when compared to initiation complexes, is attributed to the loss of contacts between RNAP and DNA. OPCs have major sequence-specific contacts upstream of the transcription start sites, which have been suggested to be lost upon transition into elongation.

Rivetti suggested that DNA wrapping is also important for the stability of the elongation complex [166]. Left-handed wrapping favours the unwinding of the DNA helix, and stabilises the transcription bubble within the active site. In the same study Rivetti *et al.* looked at visualising RNA extrusion in both *E. Coli* and eukaryotic RNAP III transcription elongation complexes using AFM, in which the enzyme had been stalled at various positions downstream of the promoter [166]. They often observed a transcript exiting the RNAP, providing that the polymerase was stalled far enough away from the promoter. The RNA chain was prevalently mapped on the opposite side of the protein relative to the smaller angle formed by the DNA arms. Measurements of the angle formed by the

upstream DNA and the transcript discovered a distribution centered on 140° and 110°, for bacterial and yeast RNAPs respectively. Our observations agree with their results for *E. Coli* RNAP. They surmised that the location of the RNA exit site may help in keeping the transcript far away from the DNA arms, and could contribute to unhindered growth of the RNA chain. DNA wrapping may prevent RNAP from rotating around the DNA, meaning that the chain would have to rotate on its axis as it passes through the protein. With RNA being situated on the opposite side of the RNAP relative to the DNA this would prevent the transcript from getting entangled with the DNA.

They also observed the binding of secondary RNAPs to the nascent RNA chain, and explained that the existence of an RNA binding site on the enzyme, could act to increase the local concentration of polymerase in the vicinity of a gene, and thus function as an allosteric transcription regulator.

5.5.2 Inter-RNAP separation

By adding NTPs into a reaction mixture containing RNAP and template DNA, and observing the change in RNAP separation, we were able to show that transcription began to occur in the direction expected for their promoter arrangements. For example, after adding only a subset of NTPs, RNAPs were unable to transcribe past their stall sites.

Once all four NTPs are available to participate in the reaction, a significant number of RNAPs were observed in close proximity to each other, and often situated toward one end of the template. This was true of both pDSU and pDSP. Such a change was confirmed by the average values of inter-RNAP distance decreasing by 44 and 32 nm's relative to the initial separations seen in the OPC scenario, for the convergent and tandem promoter arrangements respectively. What is particularly striking however, is that there were large spikes in both pDSU and pDSP distributions, representing very low separations (< 40 nm) (CC distributions in Figure 5-15b and Figure 5-18b).

Such complexes, displaying such a small separation, were not observed prior to NTP addition. These collided complexes, therefore, represent the aftermath of a

transcription event, in which the two RNAPs displayed transcriptional interference. Such small separations were rare when OPCs were imaged. RNAPs binding to DNA prior to transcription initiation utilise a 1-dimensional diffusion mechanism to locate their promoter region [81, 82, 206]. If more than one RNAP was occupied with locating their promoters one might expect a similar collision to occur on the template. No such event was observed, and as such stalling is a direct consequence of having two distinct promoters situated on a single DNA template, and subsequent interference between their two RNAPs.

A large decrease in the inter-RNAP distance suggests that one or both RNAPs begin to transcribe towards each other (in the convergent case), until they begin to become a hindrance to each other and stall against each other. These findings are supported by the work of Crampton *et al.* using a similar template [47]. Of the templates containing two polymerases, a significant number were observed in very close proximity to each other. This was interpreted as the molecules colliding with each other and subsequently stalling.

The method of analysis used in this study is limited to studying only the movements of RNAPs on the template, during which both RNAPs remain on the chain after interacting. As such, the complexes under investigation in this study represent one potential outcome to the collision. It is possible that one or both RNAPs may become detached after interacting. Sneppen *et al.* modelled the interaction of RNAPs originating from convergent promoters, and discovered that the *in vivo* results into transcription interference of Callen *et al.* are best explained if one RNAP is knocked off in the collision [30, 190]. However, other *in vitro* studies showed that colliding RNAPs in a co-aligned case can act together to remove obstacles such as protein roadblocks, with an elongating RNAP being able to rescue a stalled complex by “shunting” it from the rear. In these cases both RNAPs were observed to remain on the template [63, 64].

It appears that the direction of the promoter has no effect on the RNAP separation after transcription has occurred. Both head-on (pDSU) and rear-end (pDSP) collisions result in similar distributions, implying that a collision event always resulted in the eventual stalling of the complexes.

When a collision in pDSU was initiated by a two-step process, via the formation of a stalled intermediate, the inter-RNAP measurements suggested that a significant number of RNAPs remained stuck at their stall sites. In fact, the distribution indicated that almost half of complexes displayed a separation that was typical of the expected distance between the two stall sites (see Figure 5-12a two-step case). Of CCs formed in this way 45 % displayed a separation applicable to the distance between the two stall sites. Kommisaorova and Kashlev observed similar elongation complexes that remained particularly sensitive to certain stall sites and became what they termed arrested-complexes [110]. If the probability of both RNAPs remaining irreversibly stalled is 45 %, then the chance of either RNAP re-initiating (P) can be worked out by solving $(1 - P)(1 - P) = 0.45$, which gives value for P of 0.33. Hence, the probability of both RNAPs being able to continue transcribing the template is thus $P^2 = 0.11$, and the chance of only one re-initiating is $2(P \times (1 - P)) = 0.44$. Thus, in complexes that were able to move away from their stall sites and undergo a collision, the interaction occurs predominantly between a transcribing RNAP and one that remains stationary ($0.44/0.55 = 0.80$), where 0.55 represents the fraction of complexes where at least one RNAP is not irreversibly stalled. In the case of the tandem pDSP arrangement it is expected that SD-EC interactions will be the only form of TI possible. If neither RNAP were to stall then we would expect to observe bare templates alone, as they should both travel along the template in the same direction and run of the end. We can surmise that collisions are caused by the leading RNAP either failing to move away from its promoter or stalling downstream of the site, thus acting as a SD. The second RNAP then travels along and hits it, causing both RNAPs to stall in close proximity to each other.

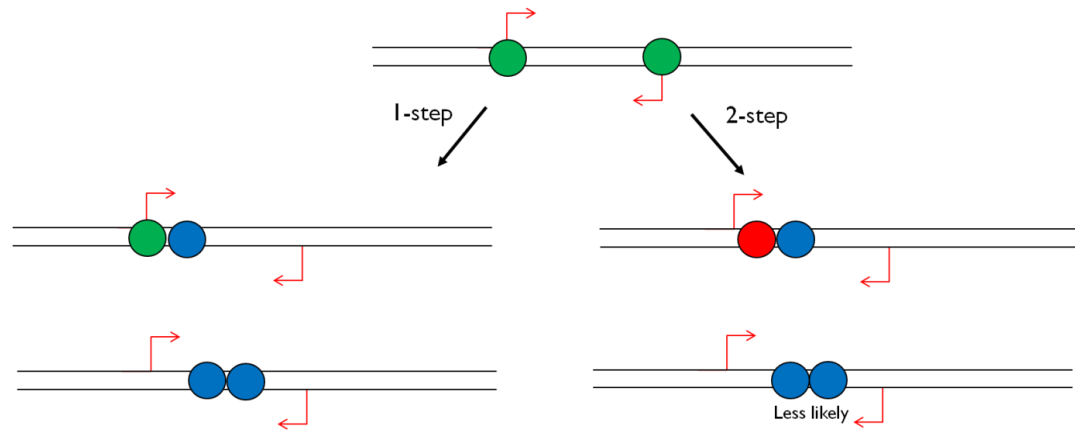


Figure 5-22. Summary of collisions between RNAPs firing from two convergent promoters. RNAPs at the promoter are shown in green, elongating RNAPs in blue and stalled RNAPs in red. In the 1-step reaction collisions can occur between an elongating complex and an RNAP that is still engaged with initiation at the promoter. Collisions between two elongating complexes are also possible. In the 2-step reaction most collisions occur between an RNAP that is permanently stuck at a stall site and a single elongating RNAP. Collisions between two elongating complexes are rarer.

In the 1-step process it is also possible for collisions to occur between an RNAP that is situated at the promoter and a transcribing RNAP. Such complexes were termed sitting ducks (SDs) by Callen *et al.* [30]. A sitting duck is prone to being hit and dislodged by the arrival of an elongation complex "fired" from a second promoter. It should also be likely that there will be more collisions between two elongating RNAPs than in the two-step process, as a large proportion of stalled complexes fail to re-fire, and thus function as a sitting duck. In the 1-step case RNAPs do not show a propensity to stall at particular sequences, and are observed across the inter-promoter region. A summary of types of collision is shown in Figure 5-22.

5.5.3 Arm-Lengths

Considering first the measurements of the two arm lengths for pDSU only (convergent case), an interesting observation is that once a collision occurred, the long arm was seen to move to higher values, whilst perhaps more significantly the short arm shifted to lower values. These observations must be connected. The long arm observations are consistent with the RNAP moving in the direction expected of transcription. However the movement observed in the short arm is at odds with the direction needed to form a transcript. These observations allow us to elucidate an outcome of collision induced backtracking (See Figure 5-23).

These are similar to the results observed by Crampton *et al.* [47]. In a two-step process whereby transcribing complexes were initially stalled by nucleotide omission and subsequently elongated by adding the missing NTP, contour length distributions taken before and after the addition of the missing nucleotide displayed different distributions over time. The short arm became progressively shorter over the time period studied, whilst the long arm length distribution became broader, with a greater number exhibiting ever longer arm lengths. The interpretation of this data was that one polymerase transcribes into the other and pushes it backwards along the template.

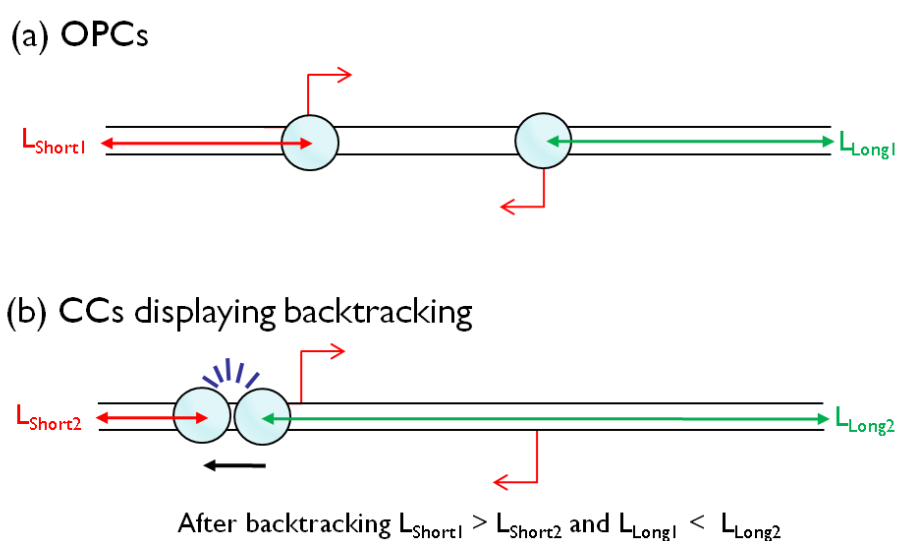


Figure 5-23. Schematic diagram of how backtracked CCs appeared. Collision induced backtracking of one RNAP results in a reduction in the short arm and a lengthening of the long arm.

Backtracking, or the reverse sliding of the elongation complex, is commonly seen *in vitro* after nucleotide mis-incorporation, and as a response to transcriptional pausing or arrest [110]. During backtracking, the RNA transcript is reverse-threaded, resulting in the loss of the 3' RNA terminus from the enzymes catalytic centre. Reactivation requires special factors, which induce the cleavage of the 3' end of the nascent transcript, thus generating a new terminal in the catalytic site. GreA and GreB are examples of such factors in bacteria, whilst TFIIS assumes this role in eukaryotes [65, 99, 195]. The lack of any such factors in our experiment might mean that any backtracked complex would be unable to re-initiate. It has also been suggested that backtracking is

reversible, if the enzyme can oscillate between an inactive and active state [109]. However spontaneous backtracking due to such events as nucleotide misincorporation are much smaller (around five bases) than the amounts we observe [181]. As the backtracking occurs by a different mechanism (induced by a collision), and the two RNAPs are often observed towards one end of the template it is likely to be irreversible.

For the tandem pDSP template, a reduction of short arm length was also observed. Rather than being indicative of backtracking, as in the convergent case, it is actually the RNAP associated with the short arm promoter moving in the direction expected for transcription. The shapes of the distributions are similar to pDSU and roughly monomodal. There is however a spike in the long-arm distribution. This is in a region near to the long-arm promoter, and likely represents complexes that had failed to fire. Such stalled RNAPs would never experience interference from another RNAP, as their relative separations should increase. No such spikes in the distribution are observed in the convergent case, because if one RNAP is stationary it is very likely to be hit by another RNAP that is active and travelling in the opposite direction, before being pushed backwards away from its promoter. The spike seen in the tandem arrangement is representative of RNAPs that have not travelled far enough to influence the other RNAP. As such, for two RNAPs to still be bound they would have to both be in an arrested state. Stalling appears possible, and the stationary RNAP could act as a barrier to other transcribing complexes.

Most of the previous studies into collisions between RNAPs have taken place on DNA templates from which multiple RNAPs initiate transcription from a single promoter [39, 63, 64, 170, 202]. Hence, collisions arise through an arrested complex being hit from behind by an elongating RNAP. It has been shown that RNAPs transcribing along a DNA chain can act cooperatively to rescue stalled complexes or remove obstacles (road-blocks) such as transcription factors [63, 64]. Hence, steric interactions between RNAPs are closely linked to the efficiency of gene expression. In a head-on collision, as in the pDSU experiment, one might expect different end results. Rather than having a cooperative effective the “shunting” of one RNAP would lead to the loss of gene expression.

This appears to happen in complexes that were observed to backtrack, as one RNAP is still able to continue moving in the 5'→3' direction necessary for RNA production.

In a previous study it was suggested that T7 and T3 RNAP approaching one another on the same DNA are able to pass each other [124]. This could be achieved by temporarily releasing their non-template strand, whilst maintaining an association with the template strand. If this mechanism was to occur *in vivo* in nested genes then it would allow expression of both host and nested genes. However, such single-subunit RNAPs are much smaller than the large multi-subunit *E. Coli* RNAP used in the present study (99 kDa vs. 378 kDa), and could behave rather differently. The common occurrence of backtracking and the large bias in location of RNAPs towards the ends of the chain, coupled with the small separation, suggests that passing of *E. Coli* RNAP in our experiments is highly unlikely, if not impossible.

5.5.4 Numerical Analysis

The average total contour length of the DNA templates displayed an increase during the development from OPC to an actively transcribing and colliding complex (e.g. for pDSU from 330 nm to 382 nm). The reduced wrapping can be attributed to the loss of contacts between the DNA and polymerase during the transition from initiation to elongation [168].

In order to directly compare the differences in arm length on maturation to an actively transcribing complex it becomes necessary to take into account the differences in DNA wrapping. To this end, the arm measurements were converted into fractions of the total contour length. This allows us to better investigate the phenomena of collision induced backtracking. The average fractional arm lengths observed for the OPCs were taken as a starting point for each RNAP, and CCs were classified as having backtracked if the fractional values for their short-arm lengths were less than the OPC case (taking errors into account). Backtracking appeared to be fairly common for pDSU with around 47 % of complexes displaying the phenomena. This rises to 81 % for

pDSP but is not actually backtracking, but rather the shortening of the short arm that would hand-in-hand with the occurrence of transcription. The average distance that the RNAP travelled towards the template end, from the short-arm promoter was found to be 23 nm for pDSU (backtracking in convergent transcription) and 38 nm for pDSP (movement of the front RNAP in the tandem case).

The positions of the RNAP on the template (from arm length measurements) were compared with the inter-RNAP distance. This revealed three main classes of collided complex: (1) those that were situated in the inter-promoter zone in the middle of the template, (2) those that displayed backtracking of one RNAP and had two RNAPs in close proximity to each other, (3) those that displayed apparent backtracking of one RNAP but whose neighbouring polymerase was situated much further back on the template, with a separation greater than the original distance between the promoters. These are summarised for both templates in Figure 5-24.

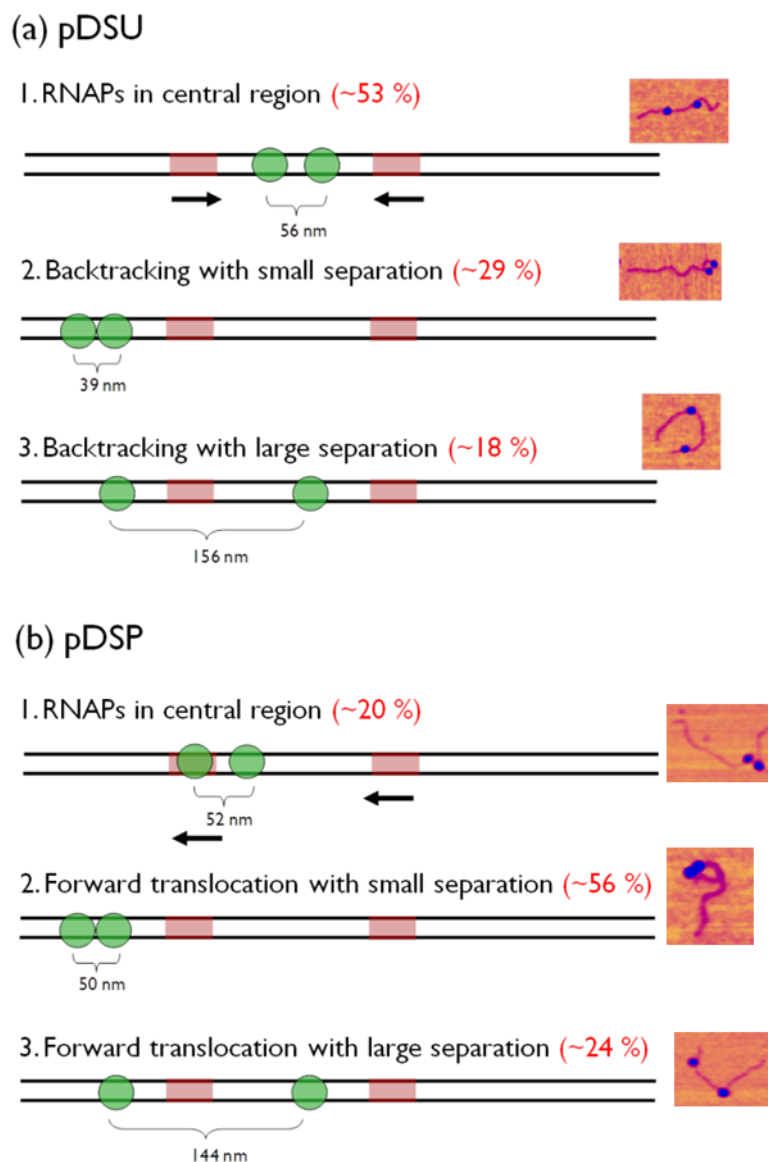


Figure 5-24. After collision, three main classes of CCs were observed for (a) pDSU and (b) pDSP. These results are summarised in schematic diagrams, together with the relative frequencies of each, and their average separations. Next to the each case are software zooms of individual collided complexes typical of each group.

Those situated in the central region of the template (i.e. not backtracked) represented around 53 % of the total number of collided complexes for pDSU and 20 % for pDSP. The average separation of the two polymerases was 56 ± 8 nm for pDSU and 52 ± 6 nm for pDSP, raising the possibility that a collision did not occur in these cases and that the RNAP have both stalled at a distance. Another possibility is that a collision event occurs and that these are the average recoil positions of two RNAPs. For pDSP, this class represent RNAPs

that are situated at a short-arm promoter that have been collided into from behind by another RNAP. In such a collision the other RNAP appears unable to dislodge the promoter-bound polymerase. This may indicate that RNAP is more readily pushed backwards than forwards.

The second class of collided complex (those displaying backtracking with small separations) represented around 29 % of those complexes imaged for pDSU. For pDSP 56 % of complexes displayed a small separation and a similar movement towards the short arm end. The average separations were found to be 39 ± 3 nm, and 50 ± 2 nm respectively. This is very similar to the first group, suggesting that both classes of complex represent the aftermath of colliding complexes.

There is an interesting third class that was not uncommon (18 %) for pDSU. These complexes displayed apparent backtracking but also had a very large inter-RNAP distance. Their average separation was 156 ± 7 nm, which is large when compared not only with the more common class of backtracking (39 nm), but also the initial distance observed in the OPCs (90 nm). This class was also observed in pDSP with a frequency of 24 %, and a separation of 144 ± 8 nm. This was relatively common and manifested itself as a peak in the inter-RNAP distribution (see Figure 5-18 CC). With regard to the tandem promoter arrangement this class is easier to explain, and appears to be the result of RNAPs from the long arm either failing to fire or stalling, while the RNAP from the short arm promoter transcribes a certain distance away from the promoter.

For the convergent pDSU template, it is unclear whether these represent an actual collision outcome, or were simply non-specifically bound RNAPs. If a collision was responsible it would appear that such an outcome leads to one RNAP becoming stalled at the point of impact, whilst the other receives a major push back or recoil along the template, leading to a large separation. In studies into simultaneous transcription by multiple RNAP II molecules, it was observed that the stalling of the leading RNAP and subsequent rear-end collision can lead to substantial backtracking of the trailing molecule ($\sim 20 - 25$ nt) [170].

It is possible that these three scenarios represent different snap-shots in time of the same event. They may also represent the aftermaths of different kinds of collision. For pDSU it seems sensible to suggest that those RNAPs stalled in the central region represent a collision event between two complexes engaged in active elongations. The fact that so many of the collided complexes are situated toward one end of the template could be indicative of collisions between an EC and a SD still bound at the promoter.

5.6 Conclusions

This chapter studies the outcome of two RNAPs actively transcribing the same region of DNA. Both tandem and convergent promoter arrangements were investigated, and it was shown that collisions between RNAPs led to similar outcomes in both cases i.e. RNAP stalling. It appears that such transcriptional interference can lead to the halting of both RNAPs. As such, gene expression levels would be fundamentally altered. It is conceivable that *in vivo* there would be certain mechanisms that would act to re-initiate these arrested complexes.

Measurements, as to how the positions of the RNAPs on the template developed during the reaction, helped to confirm that in the convergent promoter arrangement one RNAP could cause the other to backtrack. Such observations demonstrate that nested genes and the resulting convergent transcription would have a fundamental role in gene expression. Chatterjee *et al.* postulated that convergent, over-lapping genes provide a unique mechanism of gene regulation [36]. By using modelling and experimental approaches they were able to show that convergent transcription in the prgX/prgQ operon was regulated by a combination of collision-induced TI, and anti-sense regulation through RNA:RNA interactions between complementary transcripts.

With regard to the tandem promoter arrangement we also observed movements towards one end of the template. As backtracking in pDSU demonstrates that one RNAP can push another along the template, it seems feasible to suggest that a transcribing RNAP originating from upstream can push on another RNAP. Such shunting has been observed previously [63, 64]. As

a significant number remain near the short arm promoter, it may be possible that the increased wrapping associated with an OPC can prevent it from being dislodged as easily as an elongation complex that is stalled.

A number of scenarios exist for the collision between the two RNAPs [183]. For example, there could be collisions between an actively transcribing elongation complex (EC) and an RNAP that is still at the promoter in the initiation stage. We observed three main classes of CC for both pDSU and pDSP. It is possible that they represent different collision scenarios, e.g. “sitting duck” TI or collisions between two ECs.

This current analysis is limited to polymerases that remain attached to the template after collision. However, it is possible that other scenarios exist. It is feasible that the collision could be of a force large enough to knock one or both RNAPs off the template. However, overall a significant number remain template bound after the event, suggesting that these outcomes are fundamental to the mechanics of transcriptional interference, *in vitro* at least.

Chapter 6

6 Discrimination of collision outcomes in convergent transcription using single-stranded DNA loop end-labelled DNA templates

6.1 Introduction to DNA labelling

We have been able to identify a number of collision outcomes between RNAPs travelling in opposite directions as in convergent transcription, however the approach taken in the previous chapter does have a major limitation. Whilst it is possible to identify promoter-bound RNAPs fairly easily, due to the initial template asymmetry and a small spread of position, once the RNAPs fire from their respective promoters and begin to interact with one another, it becomes increasingly difficult to identify which protein belonged to which promoter. As such, measurements of RNAP position identified as being of the short arm may actually belong to the promoter on the long arm and vice versa, meaning that certain collision events may have been obscured.

We aimed to further understand collisions between convergently aligned promoters using a similar approach to before, with the notable exception that we used a DNA template that was labelled at a single end. This allows unequivocal identification of RNAPs belonging to their respective promoters, with a proviso that assumes that the two RNAPs are unable to pass each other, which is the most likely scenario from the work on the unlabelled templates in Chapter 5, whereby RNAPs were observed stalled against each other. As AFM typically provides a visual or topographical representation of a molecule if recognition imaging methods are not used, any label must be identifiable by the difference in size of its feature relative to dsDNA.

Linear DNA has previously been end-labelled by incorporating biotinylated nucleotide tri-phosphates (NTPs) into the chain, which can then be complexed

with the protein streptavidin [142, 146, 180], or a streptavidin-gold conjugate [182]. A drawback of using these techniques with AFM arises because these labels can affect the binding of the DNA complex to the surface. If non-specific protein-protein interactions occur between the end-label protein and the enzyme of interest, they will influence the conformation of the DNA rendering determination of the relative positions of multiple proteins through morphological identification alone difficult. Protein-protein interactions are affected by salt concentrations [185] and deposition of DNA-protein complexes for AFM is often carried out in low ionic strength buffers where non-specific interactions may be more dominant.

A nucleic acid based label is more desirable since it has the same surface chemistry as the DNA template and therefore does not influence the binding of DNA-protein complexes to the surface. One such approach for labelling sequence-specific regions in the interior of a DNA plasmid has utilised DNA stem-loops, where the loop hybridises with a specific sequence in the dsDNA to form a DNA triplex [66, 74]. Different sized stems, one of 200 bps and one of 500 bps were utilised as the labels protruding from the plasmid to determine polarity. This method was optimised for electron microscopy rather than atomic force microscopy and is ideal for circular DNA since triplex formation can be stabilised by DNA supercoiling [129]. However, the sensitivity of triplex formation to supercoiling makes this approach unsuitable for linear DNA molecules.

In this case, we utilise DNA stem-loops to end-label the template. We describe different methods of achieving such an arrangement. By end-labelling dsDNA templates with a single-stranded DNA loop, the polarity of the molecules can be established in the AFM. This approach was applied to convergently transcribing RNAPs as a basic *in vitro* model of nested genes, in similar fashion to the pDSU work in Chapter 5. It allowed much better discrimination between outcomes of collision events on single DNA molecules, and importantly it enabled a quantitative comparison of the relative frequencies of the outcomes and distributions of inter-RNAP separation after the collisions.

6.2 Ligation-based labelling

6.2.1 Introduction to approach and structures used

The first approach we took towards end-labelling of the template involved the ligating of small oligonucleotides to the end of the pDSU linear DNA fragment (containing the two convergently aligned promoters on opposite DNA strands). The oligonucleotides have been designed in such a way that if they are denatured, certain regions of the single strands will form base-pair interactions with themselves, and the molecule will then fold up to form secondary structures (see Figure 6-1). Forming a stem loop requires the presence of inverted repeats, where a given sequence is followed soon after by the complementary sequence in opposite orientation. It was expected that these structures will have a different morphology to the standard DNA backbone when visualised with AFM, thus allowing the identification of which end they had been attached to.

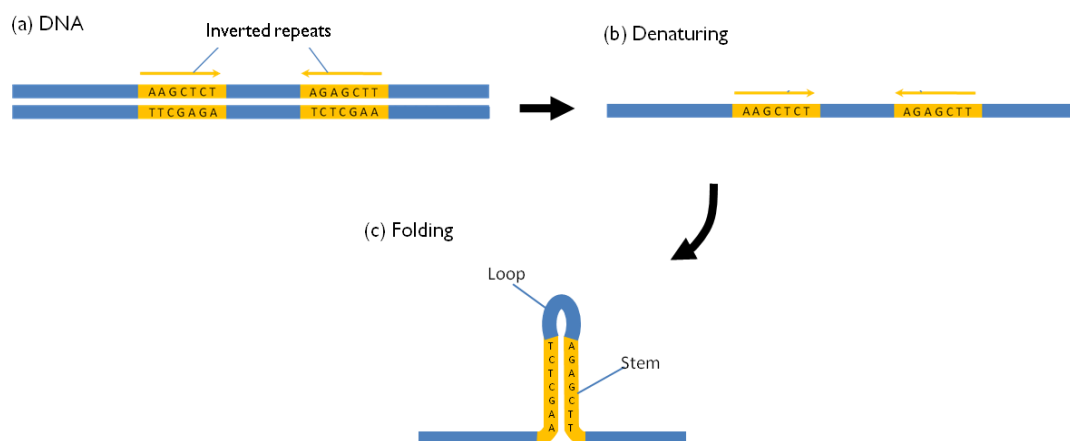


Figure 6-1. Example of how DNA stem-loops can form. (a) This requires the presence of inverted repeats on the same DNA strand. (b) If the double-stranded structure is denatured then single-stranded DNA becomes more flexible. (c) The complementary regions will base pair forming the stem of the structure, whilst the region in between the inverted repeats becomes a single-stranded loop.

This approach requires the use of DNA ligases. These are enzymes that can repair breaks in DNA chains. They are able to take two ends of DNA and covalently connect the phosphate-sugar backbone to form a continuous chain via a phosphodiester linkage. Two types of chain end can be ligated: blunt, and sticky, in addition to the repair of single-stranded nicks in the interior of the DNA chain. *In vivo*, DNA ligase is involved in DNA repair, in addition to replication, where it acts to join fragments that were formed after incorrectly inserted nucleotides are removed by exonucleases. They are also used extensively for creating recombinant DNA sequences.

Sticky end ligations are particularly efficient because the complementary ends will preferentially fit together, since the over-lapping complementary sequences base-pair through hydrogen bonding. Blunt end ligations are also possible, but are more difficult and require different reaction conditions. However, they can be useful as they allow the recombining of any DNA fragments. DNA ligase can also fix nicks in the chain backbone where only a single strand is broken. We used two different types of ligase: T4 DNA ligase which is able to join both blunt and sticky ends, and *E. Coli* DNA ligase which can only join sticky ends.

We initially designed three different oligonucleotides, with a view for use as labels (see Figure 6-2 for structures). Two were designed to fold up into stem-loops, whilst the other was expected to form a cruciform. All three structures were designed such that distal regions of complementary bases can form double stranded regions (inverted repeats). This leaves regions in the middle that cannot base-pair. These single-stranded regions are expected to have much greater flexibility forming loop structures. The complementary regions mainly consist of G and C bases in order to impart greater stability into the structure (G and C pair through three hydrogen bonds). These are designed not simply as strings of one G or C, but slightly more irregular with some interruptions by A or T bases. This should make it more likely that the molecule will fold into a defined structure, and prevent misaligning of certain regions. The flexible, single-strand parts consist of strings of A bases only. Both stem-loop forming oligonucleotides were 54 bp in length, but differed in the size of their loops. One

consisted of 12 nucleotides whilst the other contained 26, with each also having a corresponding difference in the size of the stem. Cruciforms also arise as a consequence of intra-strand base pairing but consist of two or more stem-loop structures. The cruciform we designed was 70 bp in size, and consisted of three loop structures. Each stem was made up of 7 bp, and the loop part consisted of four A bases.

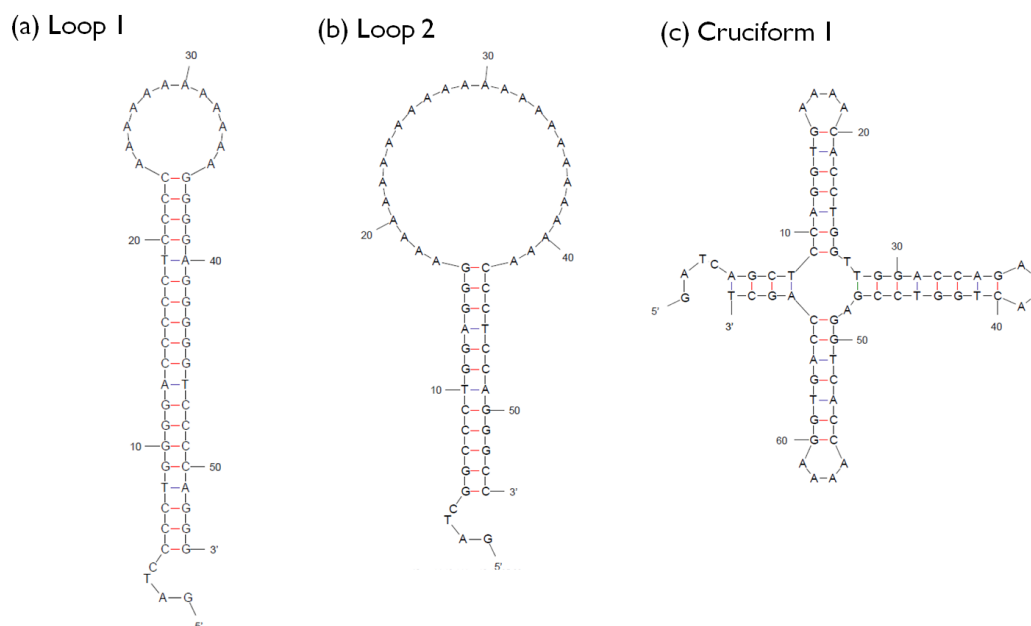


Figure 6-2. The three structures designed to end-label the template, as predicted by DNA secondary structure prediction tools. **(a) Loop 1:** Stem-loop consisting of a 12 string of A bases. **(b) Loop 2:** Stem-loop containing a larger loop consisting of 26 A bases. **(c) Cruciform 1:** DNA cruciform consisting of stems of 7 base pairs and loops of four A bases. At the 5' end of each chain is the sequence GATC that will allow the loops to be attached to the main template by ligation.

6.2.2 AFM of small oligonucleotides

To investigate the structure of the end structures we attempted to visualise their morphology with AFM. This involved an initial stage of secondary structure formation, in which each oligonucleotide was heated to 95 °C to denature the double-stranded structure, before the mix was then cooled slowly to encourage the chain to fold up. DNA solutions of around 1 nM concentration were prepared for imaging. All three structures appeared globular in the AFM (Figure 6-3). The general absence of any linear forms appears to indicate that

the chains were able to fold up. It was difficult to assess the structures by AFM because they were all globular, suggesting that not only had the stems formed, but other higher order interactions were occurring, i.e. tertiary or quaternary interactions. Since the oligonucleotides are relatively small in length, folding up into 3D structures can make the molecules more prone to being affected by tip convolution effects. However, other fairly small dsDNA samples (200 bp) imaged in previous chapters were easily identified as having a linear form.

We checked the possibility that the features observed were impurities by depositing the imaging buffer alone onto a mica substrate, without addition of DNA. In such case, the images were featureless and appeared identical to a clean background mica surface. Thus we can conclude that the globular features observed were of the oligonucleotides.

We compared the three different structures: the stem-loops (Figure 6-3a and Figure 6-3b) appeared slightly more elongated than the cruciform, which appeared almost circular (Figure 6-3c). Slightly larger structures were observed in the stem-loop samples, where it appeared that a number of molecules had aggregated. The longer stem structures in these molecules may enable multiple chains to form base-pairing interactions with each other. The CTAG 5' overhang of each structure means that individual ends of the molecules are complementary, and that they can base pair together with each other to form dimers. Evidence of this dimerisation is shown in Figure 6-3d.

The diameter across the longest side of each of the features was also measured, with Loop 1 having an average diameter of 14.1 ± 0.5 nm and Loop 2 being 24 ± 1 nm. Molecules that had appeared to have formed dimers or larger aggregates were discounted from the analysis. As might be expected, the loop with the largest single-stranded A region appeared larger. The cruciform appeared slightly smaller than the loops with an average diameter of 10.3 ± 0.4 nm. These measurements suggest that the size of the observed feature is governed by the length of the unpaired region, and hence how flexible the molecule is. The fact that the two loop structures had different diameters indicates that the observed features under AFM are those of the folded secondary structure. If they existed

as dsDNA of 54 bp long then the size of each fragment would be similar in each case (~18 nm).

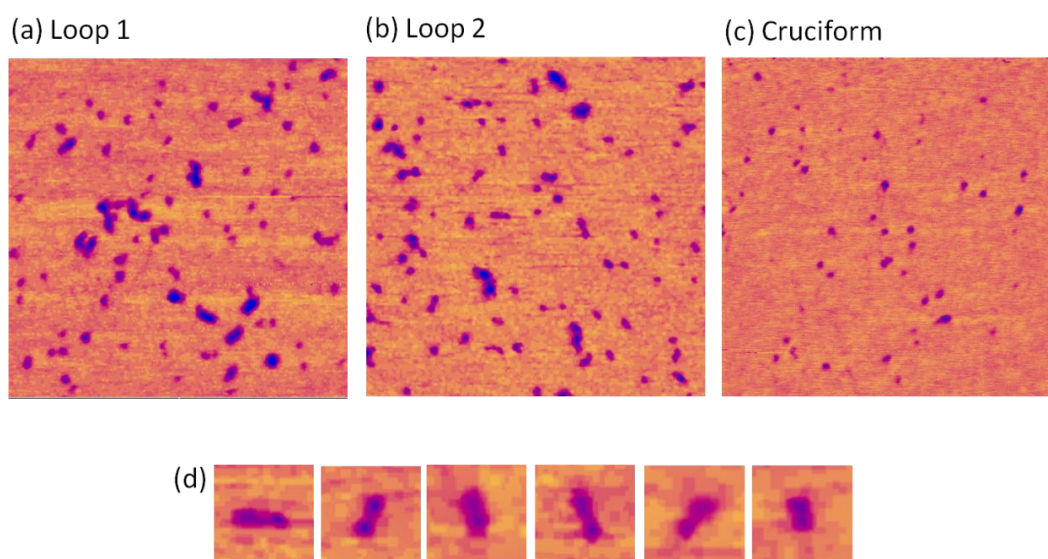


Figure 6-3. AFM images of the three different structures proposed for use in end-labelling. (a) Loop 1: stem with a loop of 12 nucleotides. (b) Loop 2: Stem with a loop of 26 nucleotides. (c) DNA cruciform. Each image is 500 nm × 500 nm. (d) Individual software zooms of Loop 2. The molecule has a morphology comprised of two distinct regions, suggesting that the structures are able to form dimers.

6.2.3 Strategy for ligation-based labelling

To be effective as end-labels, there must be an efficient strategy to attach these DNA secondary structures to the template strand. To this end we used a ligation-based approach. As sticky end ligations are more efficient than blunt end ones, we looked at ways to produce single-stranded tails at the chain ends of the convergent transcription template that would be complementary to the single-stranded over-hang on the labels. This was achieved through a two-step process involving PCR and restriction nuclease digestion. The purpose of the PCR step is two-fold. It allows the template region to be amplified to provide a large amount of starting material, but it also enabled additional sequences to be added to the end of the template to aid functionalisation. Restriction digestion of the PCR generated fragment can then leave a single-strand extension at one end. The labelling loop structures were designed to contain a single-stranded

tail, complementary to the overhang on the DNA template. The two molecules then underwent the ligation process. Figure 6-4 summarises the entire process.

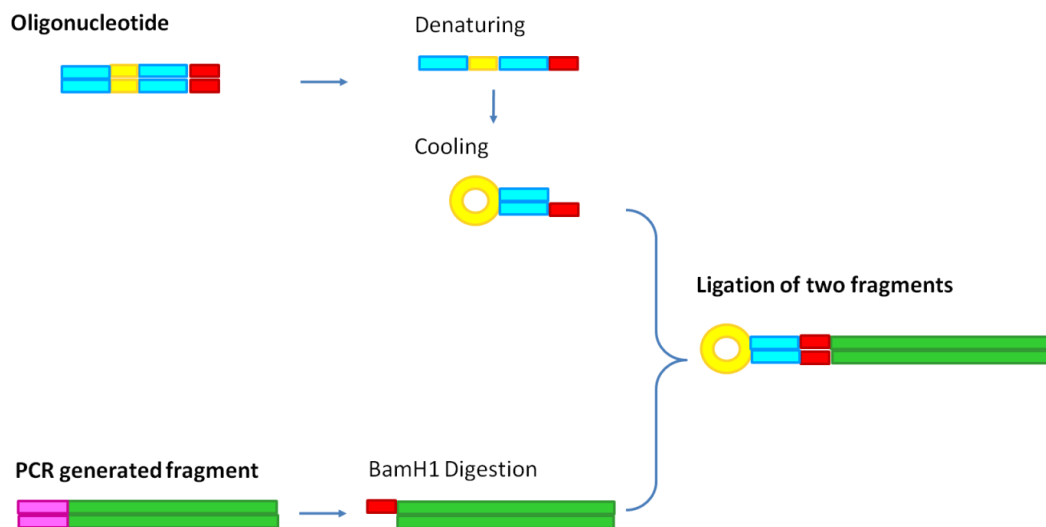


Figure 6-4. Schematic diagram showing how labelling was performed using an oligonucleotide and a PCR-generated fragment as starting material. In short, the two different fragments are ligated together through complementary sequences at the end of each fragment, leading to a loop structure at just a single end. Complementary sequences are indicated by regions of the same colour.

The PCR was performed on the pDSU plasmid using forward and reverse primers that were complementary to sequences flanking the two promoters. These were designed such that the region to be amplified would be identical to the linear fragment produced from direct enzymatic digestion of the plasmid used in the previous chapter. In addition to sequences complementary to the plasmid, the forward primer was designed with additional bases that subsequently get incorporated into the very end of the amplified fragment. The additional sequence comprised of **GGTGGATCC**. The bases in bold type represent the base sequence recognised by the restriction nuclease BamHI. If the PCR generated fragment is subjected to digestion by this nuclease then the result would be a single-stranded overhang of GATC on the 5' DNA end. The three additional bases (GGT) at the beginning of the primer were added as BamHI needs some flanking DNA to recognise its target sequence. It can be difficult to achieve effective digestion of DNA if the target sequence is near the end of linear DNA, and the extra three bases make the reaction more efficient.

Practically the PCR reaction was performed using the GoTaq Hot Start Polymerase kit (Promega) by making up a 50 μ l solution containing the heat resistant Taq polymerase (1.0 U), template DNA (3 ng), dNTP mix (0.2 mM of each dNTP), and $MgCl_2$ (4.0 mM). Also added to this were 10 ng of the template pDSU plasmid and 0.4 μ M of forward and reverse primers. The reaction mix was then placed in a thermal cycler and 30 PCR cycles of denaturing (95 °C for 30s), annealing (55 °C for 60 s), and extension (72 °C for 60 s) were performed. The DNA fragments were purified using the QIAquick PCR purification Kit (Qiagen, Valencia, CA), as was described previously.

To digest the end of the PCR generated fragment, thus leaving a single-stranded tail, restriction endonuclease BamHI (Roche) was used as per the manufacturer's instruction. Around 1 μ g of DNA was digested at a time, and the reaction mix was incubated at 37 °C. To overcome the potential difficulties of digesting a site close to the chain terminus, the incubation time was lengthened to 10 hours. The digested DNA was subsequently purified using the QiaQuick PCR purification kit (Qiagen), in order to remove BamH1 enzyme and other impurities.

To facilitate effective ligation of the two fragments, an extra sequence of bases (GATC) was added to the 5' end of each loop structure. As such, both fragments have overhangs that are complementary to each other (sticky ends). The loop structures were encouraged to form using the following process. The solution was heated at 95 °C for 2 minutes to break the hydrogen bonds in the double-helix, leading to denaturing of any double-stranded oligonucleotides. The solution was then cooled down slowly at a rate of 1 °C per minute to below the DNA melting temperature. This encourages the single-stranded molecules to form base pairs between distal sites, and ultimately fold up into the expected secondary structures that we intend to use for labelling.

Initially *E. Coli* DNA ligase (New England Biolabs) was used to attempt to attach the loop structures to the digested template. An excess of the loop structure (6:1) was added to the template in an attempt to get the labelling reaction to dominate over dimerisation of the loops. The overall concentration of DNA used

was 5-10 $\mu\text{g ml}^{-1}$. The reaction mix was incubated at 16° C for optimal ligation overnight.

Another method of attaching the oligonucleotides to the pDSU template was attempted using T4 DNA ligase (Promega). This was thought to be quicker and more efficient. Again, a 6:1 molar ratio of labelling oligonucleotide to template was used, and the reactions were incubated at room temperature for 4 hours according to the manufacturer's protocol. Although T4 ligase can join blunt-ends together, this kind of reaction is generally more efficient at temperatures between 15-20 °C, in addition to needing a longer incubation time. As such, it is expected that sticky-end ligations should be more likely in the reaction conditions used here.

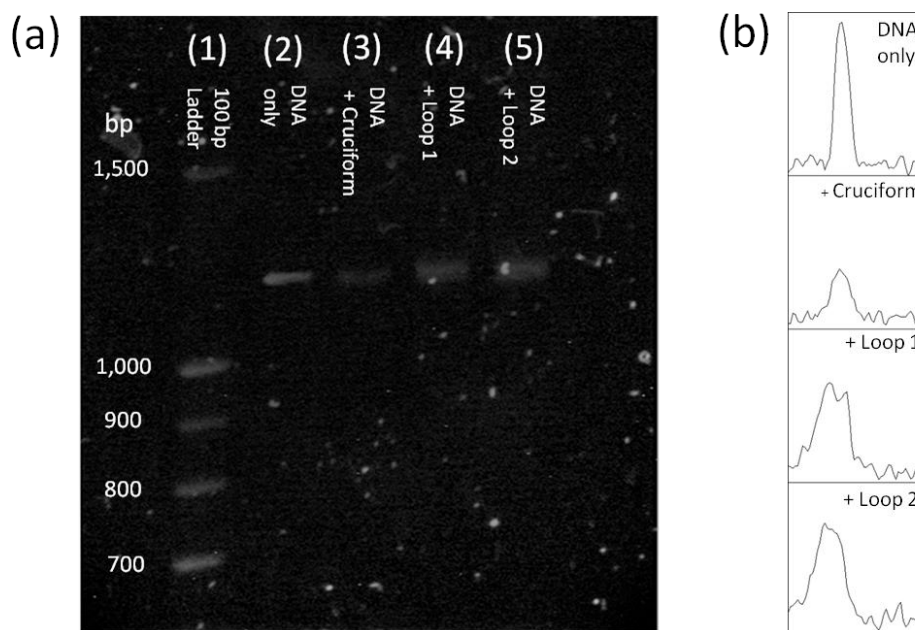


Figure 6-5. (a) Gel showing aftermath of ligation reactions. (1): 100 bp ladder. (2): Control DNA 1149 bp that hadn't been used in a ligation reaction. (3) Ligation of DNA and cruciform. The band appears similar to the control. (4) Ligation of DNA and loop 1. (5) Ligation of DNA and loop 2. These bands appear broader than the control, suggesting they may contain multiple components (labelled and unlabelled DNA). (b) Cross-sections through the bands were taken and plotted as intensities.

All ligation reactions were subsequently investigated using agarose gel electrophoresis (see Figure 6-5a for example). For all ligation reactions, a band around 1100 bps was observed. It was hoped that it would be possible to

distinguish between fragments representing successful ligation reactions and those that remained unlabelled. The additional loop structure might mean that this species would run slower through the gel and that the two components might separate into two bands. It appears though that the difference in size between the two components is not sufficient to be resolved on an agarose gel. However, when the components were run on a large gel for an extended period of time (~ 6 hours), the bands relating to the two stem-loop labelled fragments were observed to broaden slightly, perhaps indicating that the two fragments (labelled and unlabelled) were beginning to separate. Such an effect was not observed in the cruciform labelled fragment. Cross-sections through the gel bands were taken and plotted using the ImageJ image analysis software (Figure 6-5b). The full widths at half maximum were taken from these plots, and found to be equal to 33.33 for the bare fragment, 51.23 for the DNA plus the smaller loop, 53.36 for the DNA plus the large loop, and 32.01 for DNA plus the cruciform (units are arbitrary units of area used by the program). From these measurements it would appear plausible that the increased width of the two loop bands could be explained by the fact that it consists of two components.

We also observed a fainter band that was around twice the size of the convergent transcription template. This occurs through dimerisation of the template chains. The BamH1 digestion leaves an overhang of GATC. As such the ends of two strands can act as inverted repeats. The ligase would then act to join these complementary sequences of the PCR generated fragments. Although these dimers indicate that the ligation reaction was working correctly, such superfluous fragments reduce the efficiency of the labelling reaction.

6.2.4 AFM of ligated fragments

The band around 1100 bp was excised from the gel and purified using the gel extraction kit. The resulting complexes were diluted into standard imaging buffer containing Mg(II) ions. Each sample was imaged, and a number of interesting features were observed at a single chain end for the two stem-loop containing samples. These globular features appeared at just a single chain end

and were wider than the chain backbone. No such features were observed for the cruciform-labelled fragment, suggesting that either the ligation did not work in this case, or that the AFM could not detect these features. Examples of each kind of labelled fragment are shown in Figure 6-6. The longitudinal diameter of these features along the chain backbone was measured and the average of which was found to be 14.1 ± 0.5 nm and 23 ± 1 nm for the small loop (12 bases) and large loop (26 bases) respectively. Clear and consistent differences between the two sizes and the correspondence of these DNA end-features to the size of the loop structures when imaged separately (Section 6.2.2), demonstrate that these labelling strategies were successful, ruling out other anomalies, such as protein impurities.

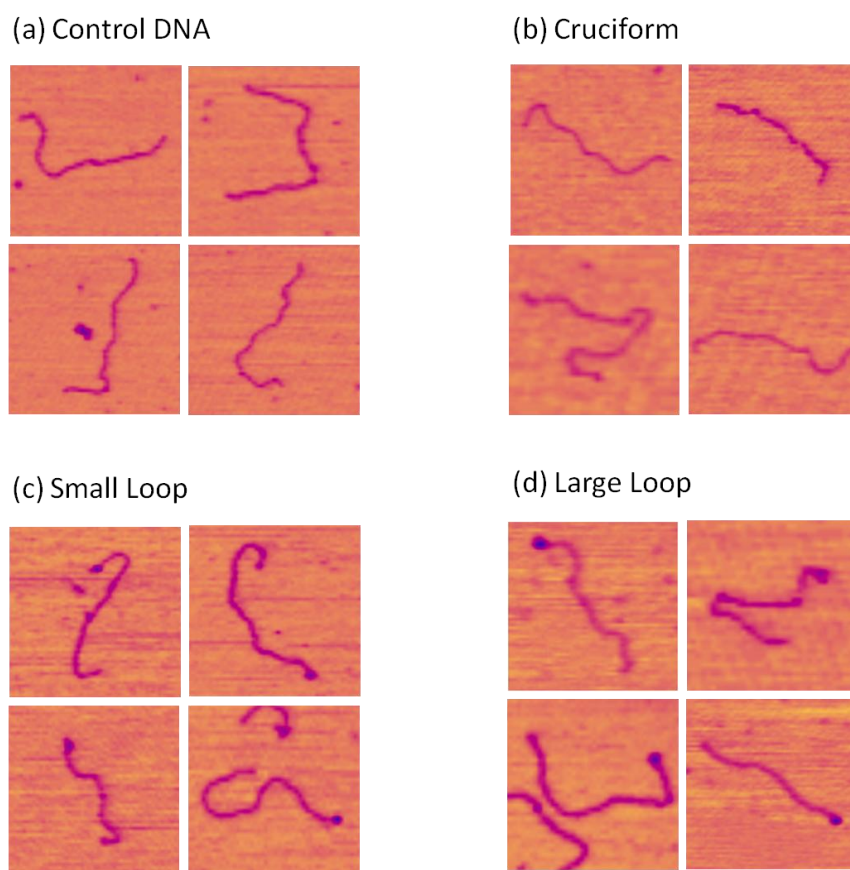


Figure 6-6. Montage showing individual software zooms of different ligation reactions. (a) Example of control DNA. (b) Cruciform: These appeared outwardly similar to the control DNA. (c) Ligation of the small loop: On some molecules some globular features appeared on a single chain end. (d) Large loop: Similar to the previous case some molecules contained an end feature. These were larger in size, corresponding to an increase in the number of nucleotides that are able to form the loop. Images are 300 nm².

Although we observed some fragments where it appeared that the end-labelling was successful for the two stem-loops, such molecules were in the minority. The prevalence of DNA with a feature at one end was around 23 % for the small loop and 22 % for the larger loop. Additionally, it did not appear possible to ligate a cruciform to the convergent transcription fragment. A bulky structure containing a number of different “arms” may disrupt the activity of the ligase enzyme. The actual labelling process also requires three main stages: digestion of the PCR-generated fragment, formation of the stem-loop or cruciform, and the eventual ligation, together with purification after each step. After each stage we would expect a reduction in the number of molecules which had successfully completed the reaction, and thus still able to participate in effective labelling. The relatively low yield of labelled molecules via this method of labelling meant that it was not pursued further.

6.3 Attaching stem-loop via a single-cycle PCR approach

6.3.1 Protocol for labelling

With a view to increasing the yield of labelled molecules we attempted a new labelling strategy, in which again we incorporated a DNA stem-loop into the chain at a single end. In particular we aimed to simplify the manner in which the loop structure could be attached. Here, we performed a single cycle of linear PCR using a single primer designed to anneal at one end of the convergent transcription template. The primer also contained additional nucleotides designed to fold up to form a DNA stem-loop once cooled. Primer extension was then performed to create a double stranded template with a loop feature at one end. This method has the advantage that the new structure can be attached via a single process, being both quicker and limiting the reduction in effective yield of the multiple-step ligation process.

A single oligonucleotide stem-loop forming primer was used for single-cycle linear PCR, having the sequence: GGCCCTGGAGGGAAAAAAAAAAAAAAAAAAAAA **CCCTCCAGGGCCAAGCTTAGGTGAGAACATCC**. The bold bases represent the sequence that will anneal to the HindIII digested fragment of pDSU acting as

primers for PCR, underlined sequences show inverted repeats that will hybridise to each other to form a stem, whilst the italics show the unpaired loop region. The oligonucleotide sequence was inputted into the MFold DNA secondary structure prediction tool to check that it would fold into the structure expected [224] (see Figure 6-7 for a schematic diagram of the end-structure).

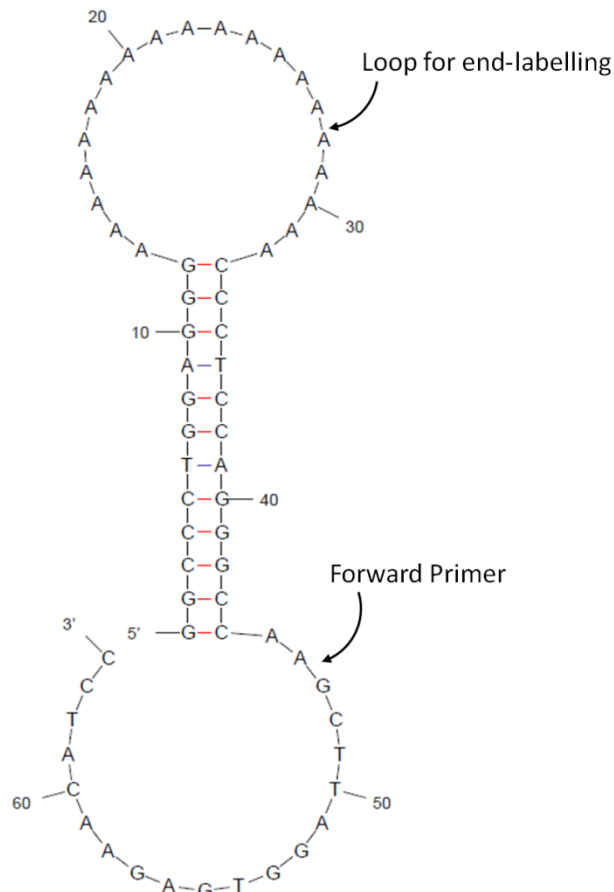


Figure 6-7. Structure used for PCR-based end-labelling. After denaturing and cooling the DNA folds to form a stem-loop. A single stranded tail region is complementary to the template DNA and acts as a primer allowing the end-structure to be incorporated into the main convergent transcription template.

The labelling process is summarised in Figure 6-8. The single-cycle PCR was performed using the *GoTaq* Hot Start Polymerase kit (Promega) by making up a 50 μ l solution containing the heat resistant *Taq* polymerase (2.5 U), template DNA (300 fmol), dNTP mix (1.6 mM of each dNTP), and $MgCl_2$ (4.0 mM). An excess of primer to template was used to encourage the labelling reaction to

dominate over re-annealing. We experimented with a 10, 100 and 1000 molar excess of primer over template.

A denaturing step of 98 °C for 10 minutes was initially performed, using all the reagents except the NTPs and *Taq.* polymerase, to prevent them from being damaged by the high heat. The higher than typical temperature and longer denaturing time period to more conventional PCR reactions should guarantee that the majority of the target was denatured, and also that the single stranded regions diffuse away from their partner strands to discourage re-annealing. The reaction mix was then immediately put on ice to encourage the stem-loop structure to form and to further prevent re-annealing of the strands. At this stage *Taq* polymerase and the NTP mix was added. Following this, annealing was performed at 60 °C for 10 minutes, and extension at 72 °C for 15 minutes. The PCR-generated fragments were purified using the QiaQuick PCR Purification Kit (Qiagen, Valencia, CA), as per the manufactures instructions.

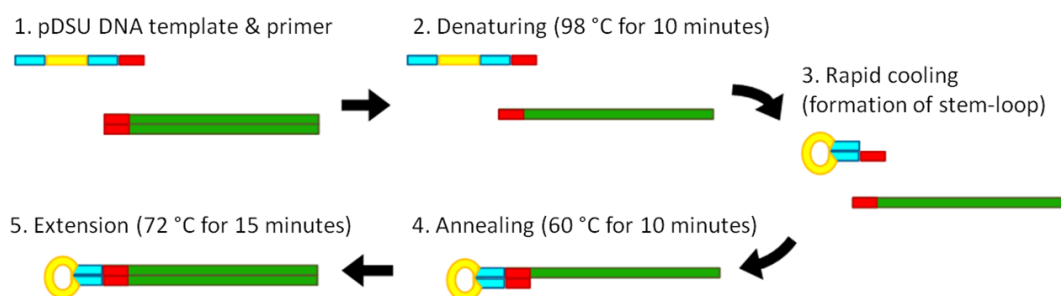


Figure 6-8. Summary of method used to end-label DNA, utilising a PCR based approach. Complementary base sequences are indicated by the same colour. The template is shown in green, whilst the stem and loop are highlighted in blue and yellow respectively. The region acting as a primer is shown in red.

The PCR reaction is intended to attach the stem-loop at a specific chain end. With the promoter arrangement of the template DNA being asymmetric (Figure 6-9a) we can still define the distance from a promoter-bound RNAP to the chain end as the long and short arms, the main difference being that there should be a stem-loop at a single end. The stem-loop forming oligonucleotide has been designed to attach at the end that was originally termed the long-arm in the unlabelled reactions in Chapter 5 (Figure 6-9b).

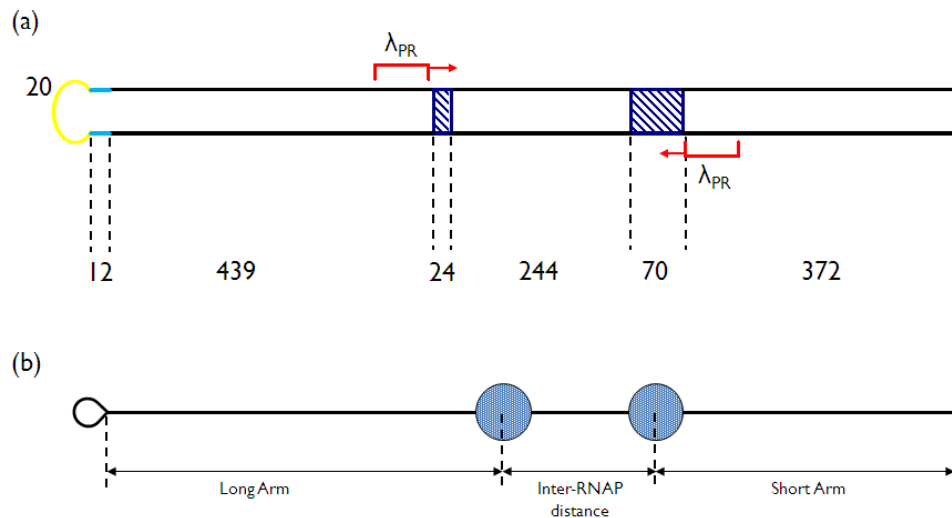


Figure 6-9. Schematic representation of the labelled template used in this study. It contains two λ_{PR} promoters separated by 338 bases. The numbers underneath display the number of base pairs present in certain regions of the template. Regions deficient in C-bases are shown by hatching. The stem and loop are made up of 12 base-pairs and 20 unpaired nucleotides respectively. (b) Schematic representation of the measurements taken during the analysis. The loop is situated on what is termed the long arm. By studying how these distances change relative to the RNAPs during the reaction, it is possible to infer what occurs during the collision.

6.3.2 Initial imaging of complexes

Attempts were made to attach a stem-loop structure to the convergent transcription template, with the aim of end-labelling the chain and allowing each RNAP to be better identified. Different molar ratios (1000:, 100:, and 10:1) of primer to template DNA were used in order to investigate what might be most efficient for the labelling reaction. Visualisation of the stem-loop end-labelled DNA templates in the AFM, revealed a class of molecules that contained a feature at one end of template (see Figure 6-10 for a montage).

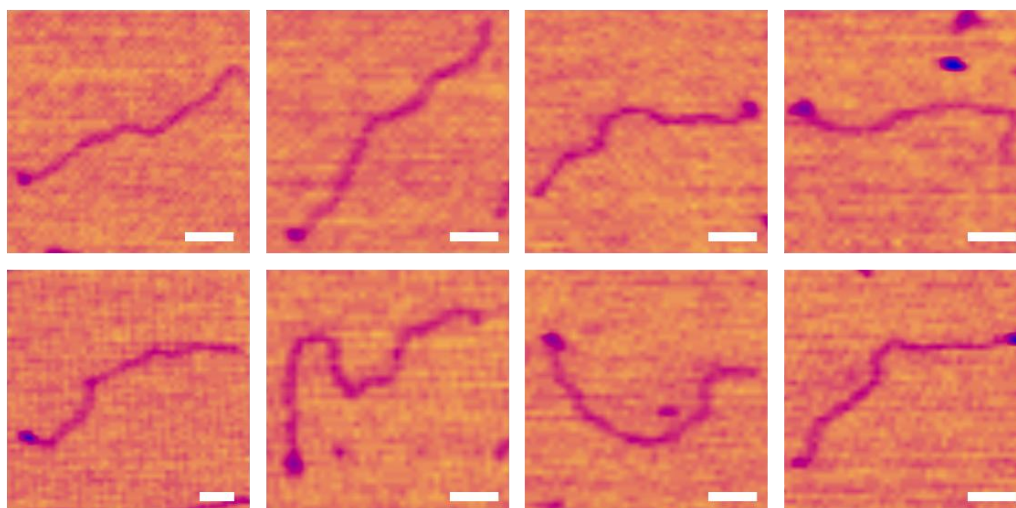
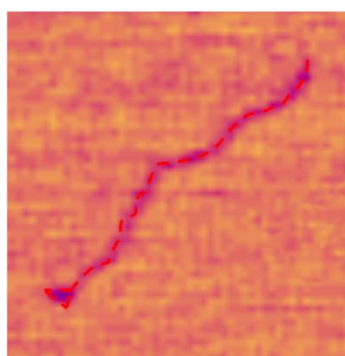


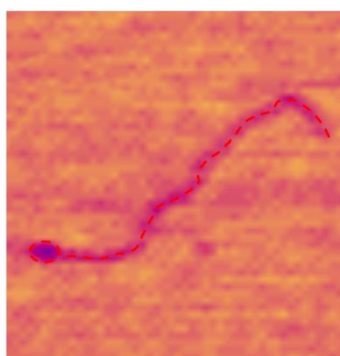
Figure 6-10. Montage showing software zooms of individual molecules with a globular feature at one end: the stem-loop attached through the PCR reaction. Scale bars: 50 nm long.

These appeared wider and often higher than the bulk of the DNA molecule and appeared in various forms with a triangular shape being the most common ($P = 0.55$). Spherical ($P = 0.37$), and irregular or branched structures were also observed ($P = 0.08$) ($n = 84$) (see Figure 6-11 for examples of classifications). This shows that the loop structure has a certain degree of flexibility and can fold into different morphologies. The appearance was similar to the end-features we observed for some molecules in the ligation approach used previously.

(a) Triangular ($P = 0.55$)



(a) Spherical ($P = 0.37$)



(c) Branched ($P = 0.08$)

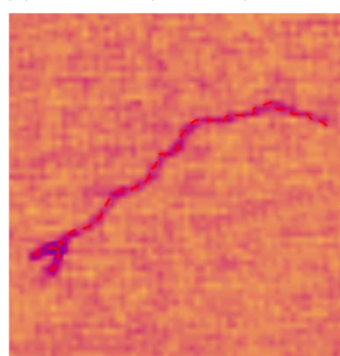
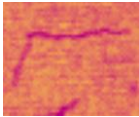




Figure 6-11. Examples of the different morphologies of end-loop structures, together with the relative occurrence of each class. Each molecule is traced out (red dashed line) to better illustrate the difference in end structure. All images are 300 nm × 300 nm.

Attempts were made to qualify the success of the labelling reaction by categorising complexes into three different groups: those with a globular

feature at one end, those with no end feature, and those with two end features. Table 6-1 shows the relative proportions of each class for the three PCR reaction mixes containing different molar excesses of primer to template and a control that had not undergone labelling.

Table 6-1. Comparison of different classes of DNA molecules observed for each preparation mix and a control that had not undergone an end-labelling reaction. Under the column headings is a software zoom of a representative molecule for each class.

Sample	Fraction with no end feature	Fraction with one end feature	Fraction with two end features	n
				
Control DNA	0.92	0.08	0	188
× 10 PCR	0.86	0.13	0.01	175
× 100 PCR	0.66	0.32	0.02	196
× 1000 PCR	0.50	0.48	0.02	319

If the fraction of molecules displaying a single feature at one end is a sign of how successful the labelling reaction had been, the table above shows that as the ratio of primer to template increases so does the efficiency of the process. With an increase in primer concentration, there is a greater likelihood of binding at the target sequence before the single-stranded regions are able to re-anneal. All subsequent experiments were performed using a × 1000 molar excess of primer in the reaction mix. It is possible that increasing the primer concentration further would reduce the labelling efficiency as the loop structures become unable to locate their target sequence.

Each AFM image contained a mixture of DNA that had a feature at the end of the chain and those that appeared to be unlabelled DNA (see Figure 6-12). Using the protocol described the number of end-labelled molecules observed in the × 1000 reaction mix was similar to those displaying no features ($P = 0.48$ versus P

= 0.50, n = 319). Templates displaying two distinct end features proved to be rare ($P = 0.02$), and as such, unspecific labelling at the wrong end, i.e. the short arm, was not a problem. As a control, pDSU template that had not undergone end-labelling, had a much lower prevalence of molecules with an end feature ($P = 0.08$). One might assume that the end feature would take the form of a ring when visualised with AFM. The lack of an inner hole could be for a number of reasons; insufficient imaging resolution due to the finite size of the AFM tip, preferential retention of salts from the imaging buffer within the loop during the sample preparation procedure, or non-specific interactions between the single strands of DNA in the loop when confined on a 2D surface.

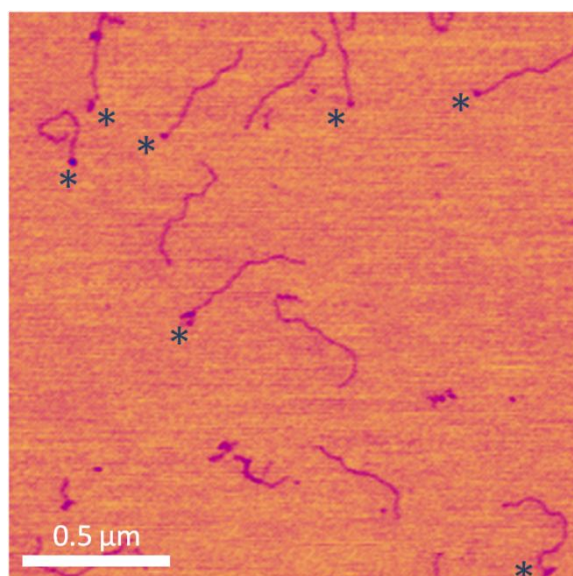


Figure 6-12. Example of a typical scan of the end-labelled DNA. Each image contained a mix of labelled and unlabelled complexes. Molecules that have a feature at a single end are marked by an asterisk.

To gain further proof that the end-labelling reaction had worked, and that the observed features were not simply an effect of surface binding, contour length measurements were taken for the individual classes from the same scans. The average contour lengths for molecules that had no end-feature, and those that had one at a single end, were 386 ± 3 nm and 407 ± 3 nm respectively. The value for the unlabelled 1149 bp fragment (386 ± 3 nm) correlates to a base pair repeat of 0.336 nm, a value typical of B-form DNA. These observations indicate that those templates with end-features represent the results of a successful labelling reaction. Such labelling is characterised by an increase in

contour length (see distributions in Figure 6-13). Interestingly, the control DNA that appeared to have a feature at one end, but which had not gone through the labelling reaction had an average contour length of 344 ± 5 nm, significantly shorter than the expected length. In this case the blob feature on the end of the template is representative of truncated molecules, perhaps caused by some form of DNA condensation. Such effects were prevalent for shorter linear DNA fragments, especially under conditions of increasing humidity, as seen in Chapter 4.

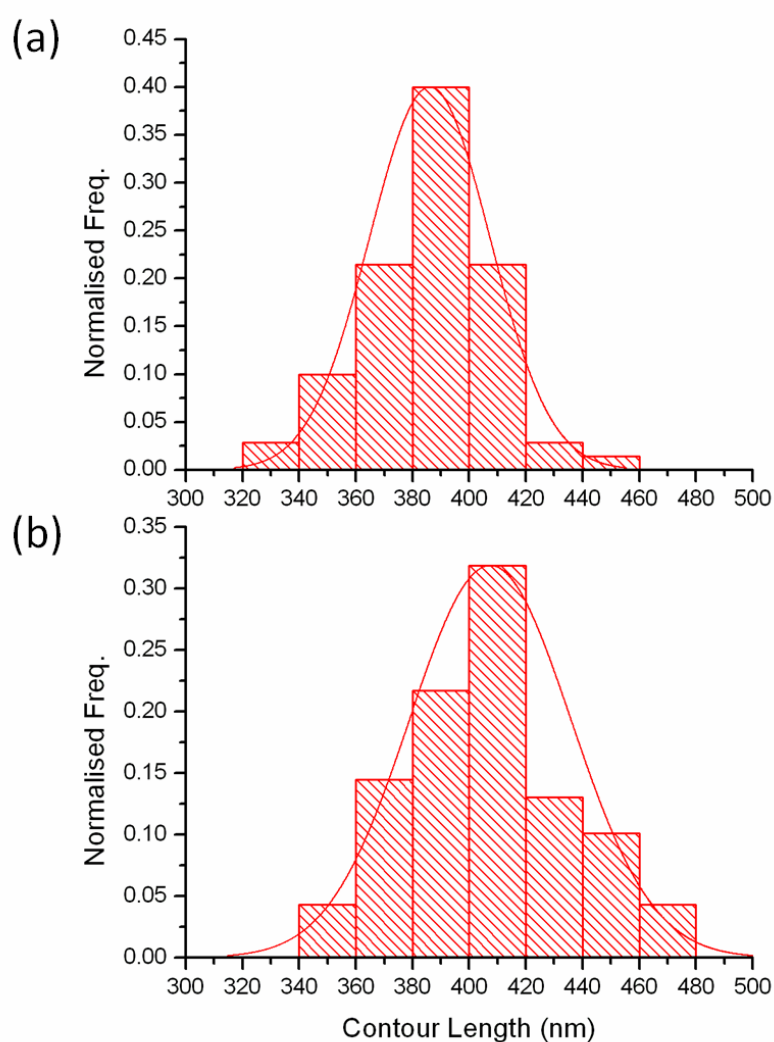


Figure 6-13. Contour length distributions for (a) bare DNA (Average = 386 ± 3 nm, $n = 70$), and (b) DNA with an end feature (Average = 407 ± 3 nm, $n = 74$). It appears that the labelling of the complex is characterised by an increase in total contour length, as indicated by shifts in the distribution.

The longitudinal distance measured separately across the end feature was 20.5 ± 0.5 nm (Figure 6-14) and explains the extra contour length measured for templates with a loop feature. The loops appear larger than might be expected as the stem is only 12 base-pairs long which one might expect to increase the size by only a few nm. Additionally, if it is assumed that the loop forms a roughly circular structure, consisting of 20 bases each separated by 0.33 nm, then this ring-structure would have a diameter of around 2 nm. However the single-stranded loop has a greater degree of flexibility than double-stranded regions, and as a result may be able to form varying secondary structures, which coupled with tip-convolution effects lead to a larger than expected end-feature.

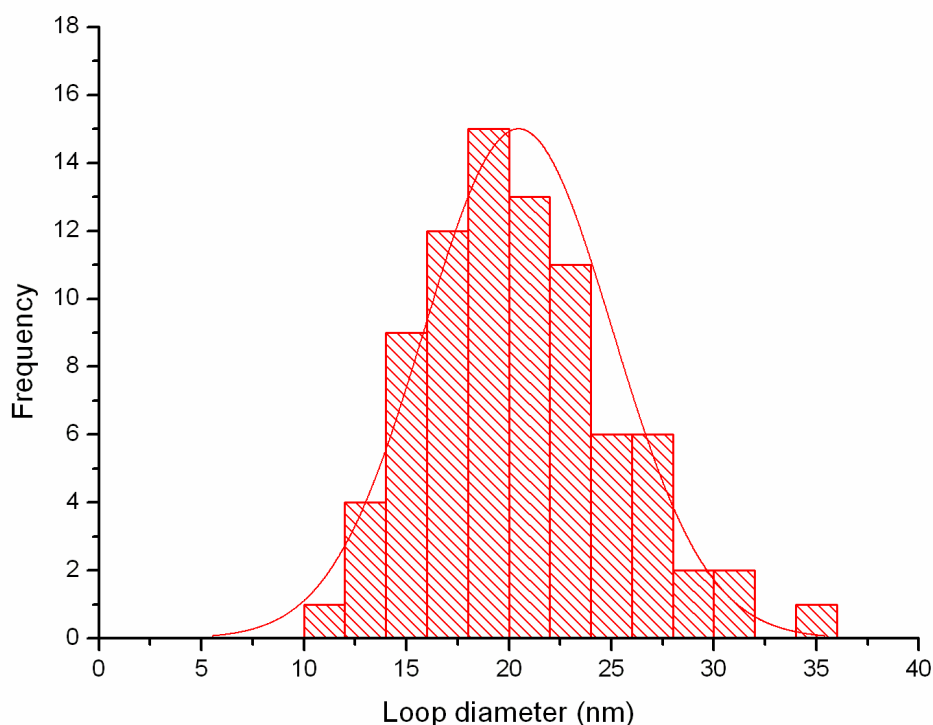


Figure 6-14. Distribution of diameter of the end-feature. The values are fitted to a Gaussian distribution, corresponding to an average of 20.5 ± 0.5 nm. This measurement corresponds well with the extra contour length observed in molecules containing a feature at one end.

Interestingly, the size of the PCR derived end-feature (20.5 ± 0.5 nm) is in between the results observed in the ligation-labelling approach, being 14.1 ± 0.5 nm for the small loop and 23 ± 1 nm for the larger. With the single-stranded

region (the sequence expected to form the loop and enable labelling) being constructed from 20 A bases in the present strategy and 12 and 26 bases in the previous approach, the increasing measured size of the end-feature with the number of bases is further proof that the labelling strategy has been successful. It is suggested that as the single-stranded region increases in size so does the area over which the loop is able to flex and therefore the larger the globular feature at the chain end.

6.3.3 Convergent Transcription

Once it became apparent that it was possible to attach the loop feature, thus labelling the molecule, the same labelled template was used to study convergent transcription. The template contained two λ_{PR} promoters situated on opposite strands and separated by 338 bp. DNA-RNAP complexes were formed that represented different stages of the transcription cycle, by performing *in-vitro* transcription reactions that used altered reagents in each case. Reactions were performed using the same protocol as in the previous chapter. Complexes were then bound to mica and imaged in ambient conditions with AFM.

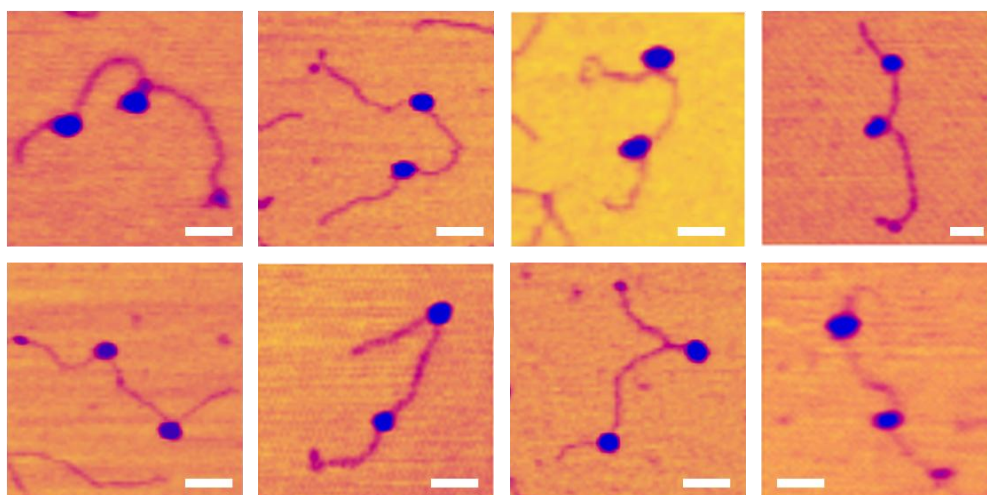


Figure 6-15. Montage of single molecule software zooms depicting OPCs. The RNAPs are identified as two globular features, much wider than the chain width, and separated by a distance consistent with the inter-promoter length. Also visible is the loop feature at one end of the template. Scale bars: 50 nm.

In the first instance OPCs were formed, whereby DNA molecules containing none, one, two or multiple (a result of unspecific binding) RNAPs were observed under AFM. Of the complexes containing two RNAPs, the holoenzymes were identified as being two equally sized globular structures situated in the central region of the template. The loop was identified as being a much smaller globular feature at a single chain-end. A montage of labelled DNA molecules containing two RNAPs, characteristic of OPCs is shown in Figure 6-15. These contain RNAPs bound in the position expected of the promoter.

As a consequence of the asymmetry of the template, it is possible to identify the distinct promoters without the necessity of end-labelling. RNAPs are expected to be situated in a narrow region around each promoter site. This allowed another control experiment concerning the end-labelling to be carried out. As the stem-loop is situated on the long-arm of the template, the other two measurements (the short-arm and inter-RNAP distance) should be unaffected by the presence of the stem-loop. In addition to this, of the end-labelled complexes, the distance from the loop to its nearest RNAP should be the larger of the two chain end to enzyme measurements. Table 6-2 summarises the average values for the three main measurements, plus the contour length, for both labelled and unlabelled complexes. As expected there is little difference in the short-arm and inter-RNAP distances. The difference in the long-arm is 24 nm, which compares favourably with the measurements of the loop-feature diameter (20.5 ± 0.5 nm). Additionally, the loop feature was identified as being on the long-arm in 86 % of complexes with two RNAPs bound. The remaining complexes probably represent non-specific protein binding.

Relative to free DNA, the total contour length is reduced to 360 ± 6 nm (a reduction of 47 nm). This represents a reduction of 23.5 nm per bound RNAP. This is consistent with a model of DNA wrapping observed in similar AFM studies into *E. Coli* RNAP, and was similar to what was observed in the previous chapter [168].

Table 6-2. Comparison of length measurements for labelled and un-labelled OPCs

OPC (nm)	Sample	Short-Arm (nm)	Inter-RNAP (nm)	Long-Arm (Loop) (nm)	Contour Length (nm)
Unlabelled		91 ± 3	113 ± 6	127 ± 3	331 ± 4
Labelled		94 ± 4	115 ± 7	151 ± 5	360 ± 6

Collisions between RNAPs were initiated by addition of NTPs into a solution containing OPCs. The one-step approach of collision inducement was used with the labelled fragment. This should enable the RNAPs to begin elongating, and for complexes with two RNAPs bound to approach each other until they become an obstacle to transcription. These collided complexes (CCs) were deposited onto mica and imaged under AFM. A montage of such complexes is shown in Figure 6-16. Qualitatively they looked different to the OPC case with the RNAPs being in much closer proximity to each other. Such effects were also observed in the unlabelled case. The number of molecules with an inter-RNAP distance typical of OPCs was only around 10 %, suggesting that the majority of complexes moved away from the initiation stage of transcription to begin elongating.

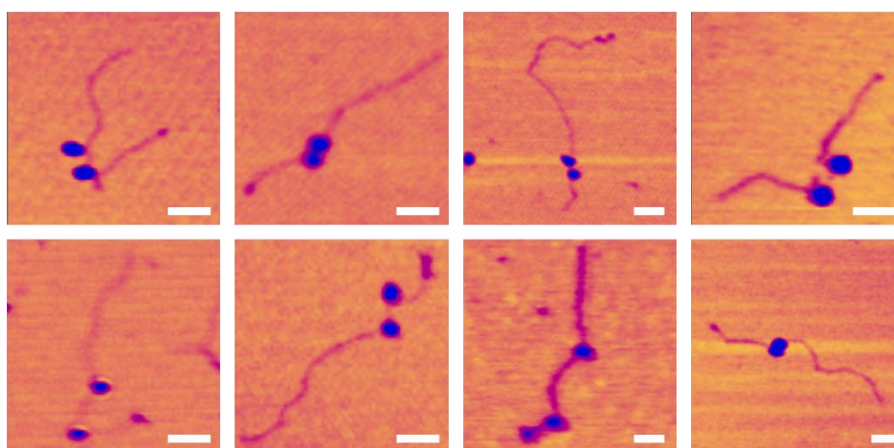


Figure 6-16. Montage of single molecule software zooms depicting CCs. Here the RNAPs appear much closer to each other. RNAPs in such close proximity were not observed when imaging OPCs. The label can still be observed at one end. Scale bars: 50 nm.

6.3.4 Analysis of contour lengths

The contour length along the DNA template between the centres of the two RNAPs was measured for both OPCs and CCs. For the OPC case, the Gaussian distribution is centred around 110 nm and is mono-modal (Figure 6-17a). This suggests a single class of complex, where both RNAPs observed on the template are always situated at the respective promoters. Once the NTPs are added and the RNAPs begin to migrate towards each other, the distribution shifts to much smaller separations (Figure 6-17b). There is a small shoulder in the distribution (around 100-120 nm) that corresponds to complexes that failed to “fire” and remained situated in the promoter region.

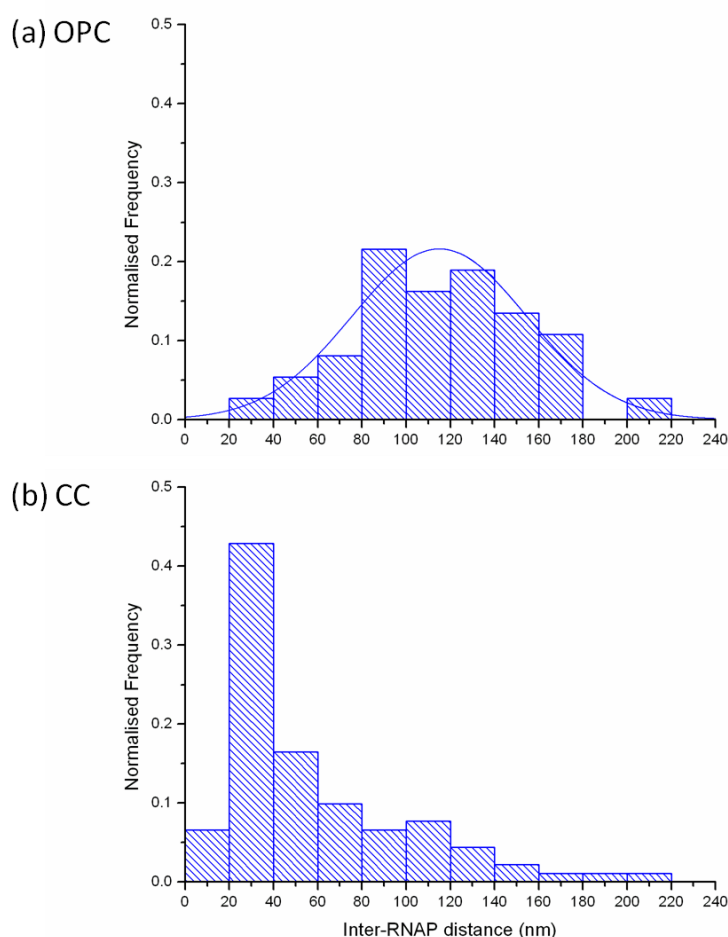


Figure 6-17. Inter-RNAP measurements from (a) OPCs (n = 58) and (b) CCs (n =91). The distribution for the OPC case is normal, and centred around 110 nm. This correlates with the inter-promoter distance expected. After NTPs have been added, the distribution shifts to much lower values, suggesting that collisions result in stalling.

The distribution of inter RNAP distance was very similar to that observed when the same *in-vitro* transcription experiments were performed using an unlabelled template in the previous chapter.

One potential drawback of attaching a stem-loop to one end is that an RNAP travelling towards the loop may be able to travel around the entire loop and to begin transcribing in the opposite direction from which it started. This could lead to a situation whereby multiple collision events could occur. The fact that the inter-RNAP distributions are the same for the labelled and unlabelled cases appears to confirm that reflection is not an issue, and those complexes that undergo a collision event and remain bound to the template are unable to reach the end of the template.

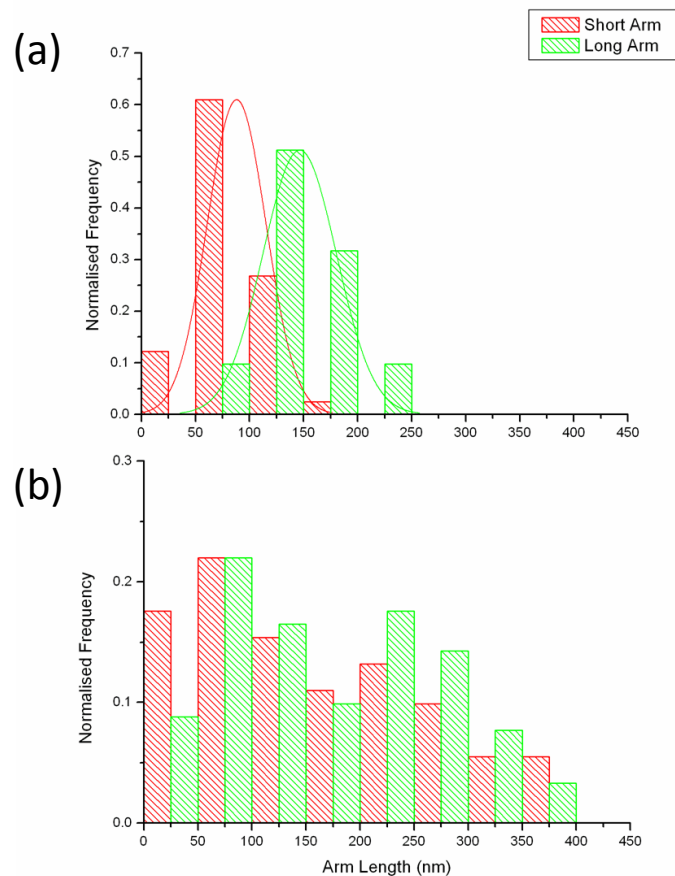


Figure 6-18. Distribution of arms lengths for (a) OPCs (n = 58) and (b) CCs (n = 91). The short arm is shown in red and the longer loop arm in green. The distributions change significantly from a narrow mono-modal shape in the OPC case, to a much broader shape with bimodal distribution. Both arms display both an increase and a decrease in values.

To compare the OPC and CC populations, the lengths of both chain ends to the nearest RNAP were measured (Figure 6-18). Using this method, it is possible to infer the results of the TI, effectively providing a snapshot in time of the collision. There is a marked change in the distributions of arm lengths for CCs (Figure 6-18b) compared with OPCs (Figure 6-18a). In the CC case both the loop-arm and short-arm length distributions are broader, with each arm length displaying shifts to both longer and shorter values, suggesting a correlation between measurements on individual complexes. In addition, both arm lengths shift from a mono-modal distribution to one that is bi-modal. A mono-modal peak suggests that each RNAP exists in one particular state, whilst the shift to two peaks suggests a change in behaviour for each RNAP. The switch to a bi-modal distribution suggests that after a collision, the RNAP can exist in one of two different positions, representing two possible outcomes. In one outcome, the RNAP is able to increase its respective arm length, whilst a different event causes the arm lengths to be reduced. These changes would correspond to transcription and backtracking respectively, and are a definite effect of the addition of NTPs, and hence interactions between two RNAPs on the template.

Numerical Analysis

The average total contour length of the DNA templates displayed an increase during the development from OPC to an actively transcribing and colliding complex (from 360 nm to 382 nm). Similar to the analysis performed in the previous chapter, the arm measurements were converted into fractions of the total contour length, allowing back-tracking to be assessed. The average fractional arm lengths observed for the OPCs were taken as a starting point and complexes were classified as having back-tracked if the fractional values for their arm lengths were less than the OPC case (taking errors into account). Back-tracking was found to be a common result of the collision with an 85 % probability (n=91). Back-tracking was also fairly evenly split between RNAPs initiated from the two promoters. Of the back-tracked molecules, 45 % back-tracked past the promoter situated at the short arm, whilst 55 % back-tracked past the long arm promoter with the stem-loop. Since both promoters are

identical this is not unexpected, but the small difference could indicate slight modulations arising from sequence dependent effects.

The fractional amount of the total contour length that each RNAP back-tracked from each promoter was 0.12 for the promoter located on the short-arm and 0.19 for the promoter located on the long-arm containing the loop. It is interesting to note that the amount of back-tracking is greater at the longest arm. These values mean that the distance from each chain end to the closest RNAP is similar in each case (being 244 bp for the short arm and 242 bp for the long arm), implying that the terminus of the template may have some impact on the stalling of complexes. The DNA is not anchored so the chain is free to bend in space. Such flexing at the chain ends may provide a large amount of viscous drag that may stall RNAPs at a certain point.

Only one complex was observed where both RNAPs moved backwards from their respective promoters and appeared not to have interacted with each other. This is most likely the result of unspecific binding. This is strong evidence that the back-tracking of one RNAP is the result of the other one transcribing into it, forcing it backwards.

With an end-labelled template, one can identify the specific promoter that an RNAP originated from, thus enabling greater understanding of how one RNAP is directly influenced by its neighbour. To this end, the positions of the RNAP on the template (from arm length measurements) were compared with the inter-RNAP distance. This revealed three main classes of collided complex: (1) those that were situated in the inter-promoter zone in the middle of the template, (2) those that displayed back-tracking and had two RNAPs in close proximity to each other, (3) those that displayed back-tracking but whose neighbouring polymerase was situated much further back on the template, with a separation greater than the original distance between the promoters. These are summarised in Figure 6-19. These three classes are identical to what was observed previously, but knowledge of where each promoter is allows the relative group frequency to be accurately defined.

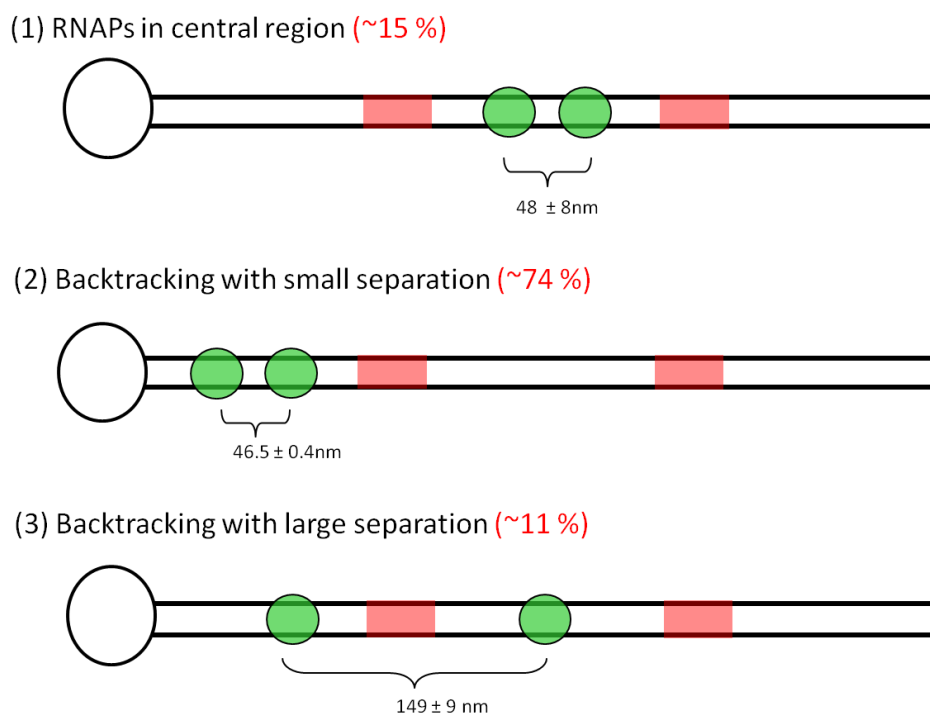


Figure 6-19. Numerical analysis of the positions of one RNAP relative to the other revealed three main classes of complex. These are represented as cartoons roughly depicting the RNAP positions. Also shown is the frequency of each class of complex and their average separation. In the two cases where back-tracking occurs (2 and 3), the two RNAPs are depicted moving past the promoter closest to the loop, but back-tracking was seen to be roughly evenly split between each promoter. The RNAPs are shown as green spheres, and approximate locations of the promoter are shown by red shading.

Those situated in the central region of the template, represented around 15 % of the total number of collided complexes. The average separation of the two polymerases was 48 ± 8 nm, a value similar to the same group in the unlabelled sample, indicating that the behaviour of RNAPs are similar for each case. The final average movement away from the promoter was slightly different for each promoter: 29 nm for the short-arm promoter and 40 nm for the loop-arm promoter. This equates to around 90 and 120 nucleotides respectively, showing that it is possible for RNAPs at either promoter to “fire” and become actively elongating complexes. The relative frequencies of the positions of the RNAP in the inter-promoter region showed that no particular promoter was biased, or “fired” preferentially, as we may expect for identical promoters.

Of the complexes situated in the central region, which RNAP travelled furthest along the template before the expected point of collision was found to be independent of which promoter it originated from. Out of these particular complexes, in 53 % of cases the RNAP originating the loop arm promoter had travelled the furthest before colliding with the other elongating complexes, whilst in 47 % of cases the opposite was true and the short-arm promoter RNAP travelled further, having presumably entered elongation first.

The second class of collided complex (those displaying backtracking with small separations) represented the most widespread group; around 74 % of those complexes imaged. The average separation was found to be 46.5 ± 0.4 nm. This is very similar to the first group, suggesting that both classes of complex represent the aftermath of colliding complexes. Back-tracking past one or other promoters was observed, with the average amount of back-tracking being 59 ± 4 nm (~ 180 bp), upstream from either promoter. In the unlabelled case the occurrence of this group of collided complex was only 29 %. This demonstrates the advantages of labelling the complex in terms of better quantifying the results of the collision. This effect is explained further in Section 6.4.3.

The third class represents complexes which displayed apparent back-tracking of one RNAP but with a very large final inter-RNAP distance (11 %). Their average separation was 149 ± 9 nm, which is large when compared not only with the more common class of back-tracking (46.5 ± 0.4 nm), but also the initial distance observed in the OPCs (94 ± 4 nm). The average movement of the non-back-tracked RNAP away from its own promoter was around 50 ± 4 nm, whilst the other RNAP was found to be backtracked by 60 ± 9 nm. This value is comparable with the previous case (backtracking with a small separation). Additionally, the back-tracking was found to be split evenly between each promoter.

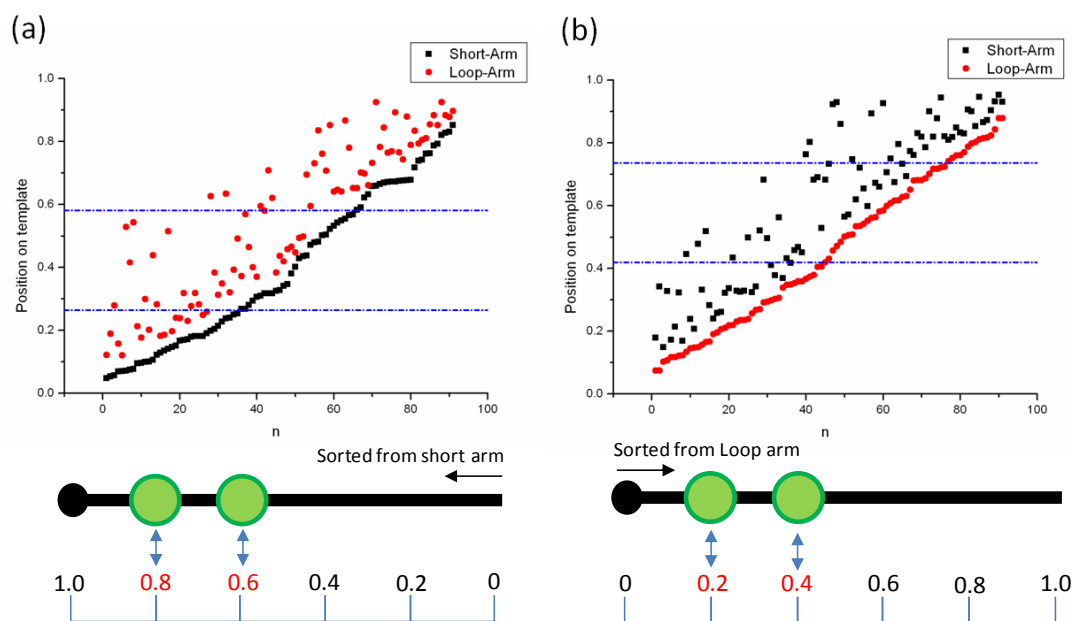


Figure 6-20. Plots depicting the data for each pair of arm-lengths sorted in numerical order of (a) the short-arm and (b) the loop arm. Also plotted is the position of the remaining RNAP belonging to the other promoter. The position of the two promoters is indicated by the two dashed lines. Underneath each graph is a schematic diagram of how the positions of two RNAPs would appear and the values plotted from (a) the short arm terminus and (b) the loop arm terminus.

To visualise better the relationship between one RNAP and its neighbour, the length measurement on individual templates were collated and displayed graphically. The positions of the RNAPs on the template (converted into fractions of the total length) were sorted into numerical order of each arm in turn, such that a value of 0 is the end closest to that particular promoter, and a value of 1 is chain-end furthest away. Also plotted on the same axis are the corresponding positions of the remaining partner RNAP. Figure 6-20 shows the data for this particular analysis, sorted in order of (a) the short-arm and (b) the loop arm. In both cases there seems to be a cluster of points that closely follows the data of the other RNAP that has been sorted. This shows that the position of one RNAP is closely linked to the behaviour of the other. There is a group of RNAPs that have a greater separation, representing the third group (backtracking with a large separation). The majority of the points belong to the two groups, in which RNAPs remain in close proximity to each other after stalling.

6.4 Discussion

6.4.1 Labelling

We pursued a technique for end-labelling DNA molecules for AFM using stem-loop forming oligonucleotides. Through measurements of the DNA contour length, and the size of the end-feature, a distinct population of molecules was identified. This group represented molecules for which the PCR-derived labelling reaction was successful, and thus had the stem-loop at the expected end. Ohta *et al.* studied the *Staphylococcus aureus* HSP70 operon and its interaction with the regulatory protein ORF37 using AFM [148]. The promoter of this operon contains a palindromic nucleotide sequence that can form a stem-loop structure. When imaging this promoter sequence with AFM they observed a small feature at one end of the chain, and concluded that this was direct evidence of promoter based stem-loops. The features imaged in their study appeared very similar to the features we observed at the end of the DNA template. Measurements of the diameters of the two different features also agree. They observed a diameter of around 10 nm for a loop consisting of 11 nucleotides, whilst we observed a diameter of 20.5 ± 0.5 nm, for the structure consisting of 20 nucleotides. In their study, the interesting secondary structure was only observed in 10 % of complexes, suggesting the reason for this low yield being that higher-ordered DNA structures were inherently unstable. This may explain why we only observed loop features on around half of complexes.

Yoshimura *et al.* aimed to study how telomere repeat binding factors (TRF) help to maintain telomeric DNA, by observing how the DNA complexed with the protein using AFM [217]. They used DNA carrying a stretch of the sequence $(TTAGGG)_n$, and a long 3' overhang, to represent telomeric DNA. They observed a number of higher-order DNA structures at one end, which were formed by the single-stranded over-hang. These took the form of not only globular structures (a common sight in our labelled molecules), but also branched structures, a structure that we occasionally observed. The fact that we observed similar end-

structures suggests that the structures seen were the result of single-stranded regions forming more complex secondary structures.

6.4.2 RNAP separation on labelled template

After addition of NTPs, a significant number of pairs of RNAPs were observed in close proximity to each other. Such complexes, displaying such a small separation, were not observed prior to NTP addition. As such, these collided complexes represent the aftermath of a convergent transcription event, in which the two RNAPs displayed transcription interference. No such small separation was observed when OPCs were imaged. The distribution shift from OPC to CC was almost identical to that of the unlabelled case. It is worth noting that labelling either end of the template does not affect measurements of the RNAP separation. The earlier study with unlabelled molecules observed a similar distribution with a large shift to a smaller separation after the elongation stage.

Very few RNAPs appear to be in direct contact after collision, while the modal group for inter-RNAP distance is in the range of 20-40 nm. Measurements taken from the AFM images found the RNAP diameter to be 19 ± 2 nm. Crampton *et al.* suggested that a general lack of hard-sphere contact results from the ability of two RNAPs to sense each other through action at a distance, either through electrostatic interactions, or through the mediation of topological stress within the intervening DNA template [47]. A model has been put forward suggesting that a transcribing RNAP, moving via unidirectional Brownian motion and rotating around DNA, provides a large viscous drag component, which leads to the build-up of positive super-coiling ahead of the enzymatic core, and negative super-coiling upstream of this region [121]. In the case of convergent transcription, this would lead to a large build-up in positive supercoiling in the inter-RNAP region. This region of high local strain could potentially inhibit elongating polymerases.

6.4.3 Back-tracking

The labelling of one end of the template becomes extremely useful once NTPs are added and RNAPs begin to move away from their promoters. It allows us to uncover populations of data that may have been previously hidden, in order to better understand the interplay between two RNAPs travelling in opposite directions. By observing the changes of arm length from initiation, to after elongation and interference, we were able to elucidate outcomes of collision induced back-tracking. This is similar to the results obtained in the previous chapter, however there are limitations in the previous approach. Once RNAPs begin to move along the template it became impossible to precisely say how the RNAPs are interacting. In the unlabelled case if an RNAP travelled past the long arm promoter in the 3' direction, but did not travel backwards more than the difference between the two arms, then a back-tracking event would not be recognised. Due to our knowledge of the DNA template polarity we were able to verify that back-tracking is more common than previously thought, with around 85 % of complexes being observed in a back-tracked state. Back-tracking in the unlabelled case was observed in only 47 % of CCs. The increased occurrence of back-tracking is related to the difference in length between the short and long arms (67 nt).

Perhaps the most marked difference between the labelled and unlabelled cases was in the arm length measurements. Our results on the labelled complexes reveal a different distribution of arm lengths after collision. In the previous chapter, the distributions of arm length for pDSU were essentially mono-modal, with shifts in the short and long arms to smaller and larger values respectively being indicative of back-tracking. However, the data we obtained for the labelled sample was bi-modal, with each arm exhibiting a similar shift to both smaller and larger values respectively (see Figure 6-18b). This is a result of the inability to completely discriminate between the two RNAPs in the earlier study. An RNAP originating from the short arm promoter, inducing back-tracking in the RNAP at the other promoter would be mistakenly identified as

the long-arm if there was no knowledge of which end belonged to a particular promoter.

A number of scenarios exist for the collision between the two RNAPs. The most common collision outcome (74 %) was back-tracking with the two RNAPs situated in close proximity. The back-tracking was found to be split roughly equally between each promoter. The range and relative frequencies of both arm-length distributions was also roughly equal. As such, it seems feasible to suggest that these complexes represent collisions between an elongating complex (EC) and a “sitting duck” (SD). The fact that both RNAPs are situated reasonably close to either promoter suggests that the SD represents a promoter-bound complex that has not yet become transcriptionally active.

Both promoter sequences are identical, and as such the time spent in the initiation stage should on average be similar. Hence, on average which promoter fires first is dependent only on which promoter an RNAP locates first. Promoter clearance is expected to be the rate-limiting step in the transcription reaction. Transcription elongation rates of *E.coli* RNAP are fast *in vivo*, with rates of 50 bp/s being measured [12]. It seems likely therefore that one RNAP is able to traverse the inter-promoter distance (338 bp), whilst another is stationary initiating at the other promoter. As such, the bimodal distribution in CCs represents two potential outcomes for an RNAP and its partner promoter. If it is the first to bind and fire then it is able to displace the other RNAP, leading to an increase in its effective arm-length. However if it is a late arriver at its promoter then it will experience a collision from an already elongating complex, back-tracking it from its position, and leading to a shortening of the arm-length. These outcomes are described in Figure 6-21. This transcription interference is indiscriminate of promoter.

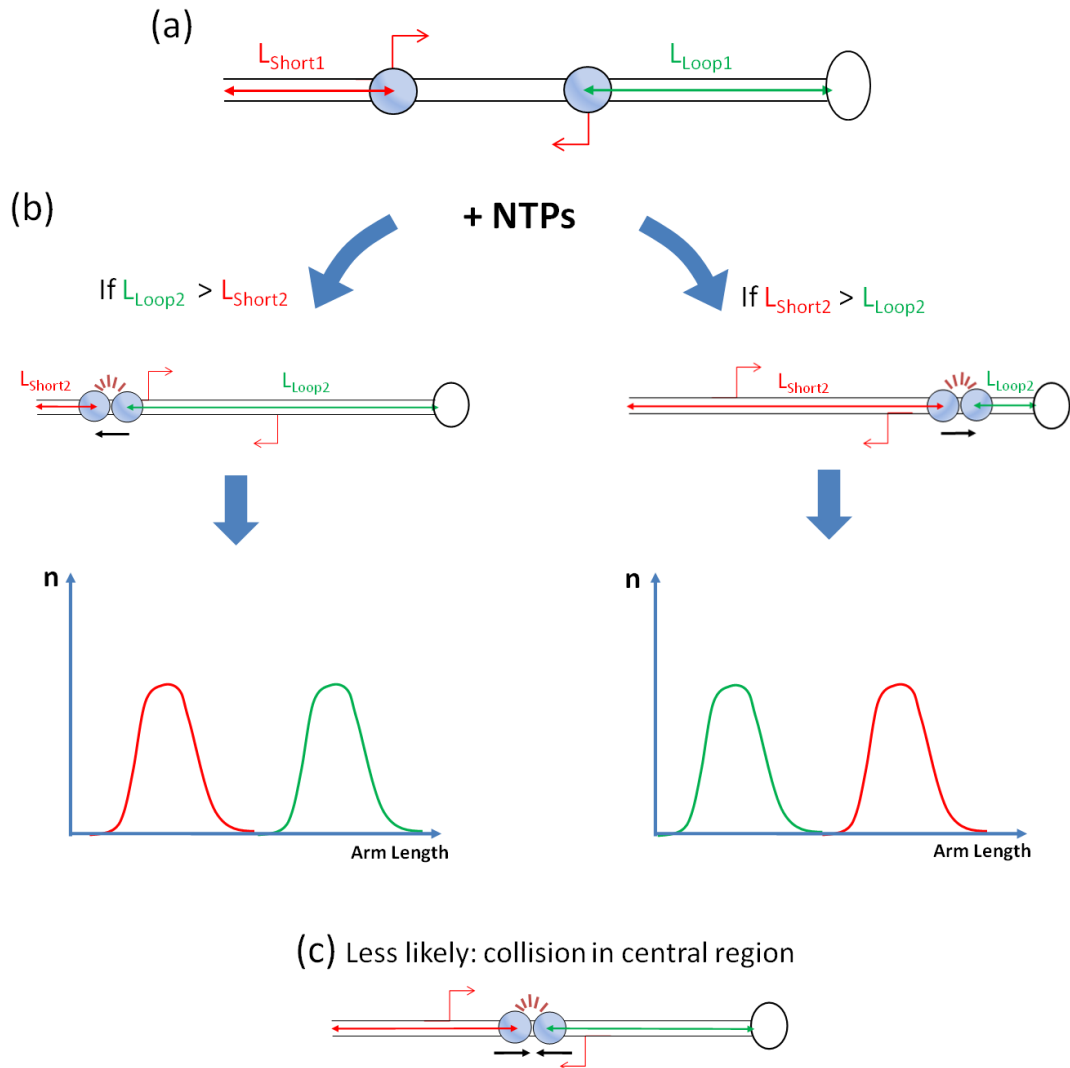


Figure 6-21. Collision outcomes observed in labelled complexes. (a) The process begins with RNAPs bound at the two promoters. (b) Once NTPs are added transcription elongation can proceed. Collisions between active complexes and those that are still promoter-situated dominate. These lead to back-tracking in one of two directions depending on which RNAP leaves its promoter first. Afterwards the RNAPs stall in close proximity to each other. This will also affect the CC arm-length distribution. An example distribution is shown below the schematic diagram, with what were originally the short-arm being shown in red, and the loop arm in green. These separate distributions combine to form a bimodal distribution for each arm. (c) Less likely are collisions between two RNAPs that have both “fired” and interact in the central region of the template. Later the RNAPs stall close to each other.

A second class of complex is situated in the inter-promoter region, which also had a small separation (15 %). The lack of any back-tracking past a promoter suggest that in this case both RNAPs were able to move away from their

respective promoters, and hence this scenario represents collisions between two ECs. In this case, it appears that transcription is irreversibly stopped in both directions. In the back-tracked case, one RNAP is still able to continue moving in its preferred direction, at least initially. The small separations suggest that the RNAPs cannot pass each other, but remain against each other in a stalled state. Within the arm-length distribution a clear peak relating to this class is not observed because it is obscured by the two larger peaks of the more common first class, and is a consequence of the relatively broad distribution widths. With an increasing separation between the two promoters it is expected that both RNAPs will have more chance to initiate, and thus a greater occurrence of collisions of this type (between two elongating complexes) would be expected.

A third less common state was observed that exhibited back-tracking, but with a large separation. Again, it is unclear whether these represent an actual collision outcome, or were simply non-specifically bound RNAPs. The collision itself could lead to a loss of the RNA-DNA hybrid which stabilises the complex. It is possible that after collision the RNAP becomes re-engaged in 1-dimensional diffusion in an attempt to locate the promoter, resulting in a larger than expected separation. Due to the loss of the sigma factor σ^{70} on transition into the elongation stage the enzyme will be unable to form the necessary contacts with a promoter region to form another OPC and re-initiate transcription [25].

However, as a potential drawback of these experiments, we cannot completely discount the effects of unspecific binding. It may also be possible to have asynchronous binding of RNAPs. For example after OPC formation it may still be possible for RNAPs that are free in solution to bind to a promoter site when a RNAP bound previously begins transcription once NTPs are available in the reaction mix. In this case the first RNAP could begin transcribing the template, and if there are no other RNAPs in its path it may pass the other promoter site unhindered before stalling towards the template end. An RNAP arriving at a later time point may bind to the promoter that the initial RNAP had just passed, and begin travelling in the opposite direction. Neither of these RNAPs would have interacted with each other, and the end result would be the appearance of

back-positioning of one RNAP on the template with another RNAP that is situated much further away (the third class of CC). One improvement in the experimental system could be to add heparin to the reaction mix after OPC formation. Heparin is a polyanion that has been shown to prevent unspecific binding of RNAP to DNA. This would have the effect of “mopping” up any free RNAP in solution before NTPs are available for transcription to begin, and may reduce the chances of asynchronous binding occurring.

Backtracking was also observed in the unlabelled sample. However, the results on the labelled sample show that either RNAP can exert a steric hindrance on the other. A strong cooperation, or coupling of the two RNAPs, was observed in these experiments. When either of the two arm measurements were plotted in position order together with the corresponding position of the remaining arm (Figure 6-20) the unsorted data generally mirrored the sorted data. This indicated that an RNAP's behaviour was closely linked to its neighbour.

Whilst the same three groups were also observed in the unlabelled case, the relative frequencies of each class were different. Of the two main groups, the occurrence of back-tracking with a small separation changed from 29 % previously to 74 % in the present system, whilst the apparent incidence of RNAPs in the central inter-promoter region went down from 53 % to 15 %. These changes roughly correlate with the increase in observed frequency of back-tracking (38 %). This allows us to better refine our model of what happens during the collision. If these two groups represent two different scenarios, namely SD-EC collisions and EC-EC collisions it would appear that the first case strongly dominates over collisions between two RNAPs both engaged in transcription.

6.5 Conclusions

These results show that transcription interference arising from the physical collision of interacting polymerases on the same DNA template could have important regulatory roles, and that the nested gene arrangement has an ability to control gene expression at a fundamental level. In this study we examined

how RNAPs transcribing from convergently aligned promoters could interfere with each other. By labelling the DNA template we were able to observe elongating RNAPs colliding with another RNAP situated near to its promoter, and subsequently inducing back-tracking. The two RNAPs eventually stalled close to each other. Presumably the back-tracked RNAP is unable to continue with transcription, whilst the remaining enzyme may still be engaged in catalysing RNA chain formation.

In our study the promoters were of equal strength, such that interference from each promoter was roughly equal. However, if one promoter was stronger than the other, either intrinsically or through availability of cofactors, then there would be a clear bias in which gene would be expressed. This study demonstrates that single molecule techniques such as AFM, with non-invasive labelling methods, can be used to gain a deeper insight in the mechanics of collision events that may be associated with nested genes.

In conclusion, end-labelling of DNA using nucleic acid structures does not interfere with the activity of the two RNAPs and can be used as a generic approach to studying interactions of multiple proteins on DNA templates at the single molecule level.

Chapter 7

7 Future work

While we were able to observe and quantify a number of outcomes of convergent transcription from two oppositely aligned promoters, including stalling and backtracking, the current time-lapse strategy is limited to studying transcription interference whereby both RNAPs remain on the template following transcription. Even though our results and those of other studies into colliding RNAPs indicate that after collision RNAPs can remain on a template, and thus affect the efficiency of transcription, we cannot discount either one or both RNAPs being knocked off the chain after collision.

Whilst it is possible to look at the relative amounts of RNAP bound to the template before and after NTP addition, with a reduction in the number bound indicating that collision may result in unbinding, such measurements may be subject to sample variability and are dependent on the stability of the complex. In our time-lapse approach we look at the aftermath of a collision, before working backwards to infer what happened during the interaction. This means that we may miss certain sub-groups of the population (e.g. template dissociation). An improved method would be to use an *in situ* approach whereby the entire reaction from promoter recognition to eventual collision could be followed directly for individual molecules. Such dynamic studies require imaging under bulk aqueous liquid, preferably with faster scanning AFMs

Traditionally AFMs slow scan speed, and problems adhering molecules to substrates well enough to be imaged clearly, has made studying dynamic processes such as transcription difficult. As such, studies into transcription under liquid have been performed using reduced concentrations of NTPs [82, 102]. Hence, NTP binding becomes the rate-limiting step in elongation, and the rate of transcription is low enough to allow movements of RNAP to be scanned

successively. However, if studying collisions between RNAP, such as those occurring in convergent transcription, the low speed of transcription necessary may lead to different outcomes. For example an RNAP may not be travelling quickly enough to exert a force on a stationary RNAP that is sufficient to induce backtracking.

Current developments into fast-scan AFM offers exciting possibilities for studying dynamic processes occurring at speeds similar to those seen *in vivo* [49, 51]. Transcriptionally active complexes can be formed and imaged *in situ* by adding RNAPs and NTPs into the imaging buffer containing the template, and in principal the entire process could be imaged from initiation to elongation, and the eventual collision. There are however some technical limitations to overcome to be able to study multiple proteins interacting on a single template. Previous real-time studies on transcription have shown DNA being pulled through a stationary surface bound RNAP as it undergoes transcription [102]. This method would not be suitable for studying convergent transcription, where both RNAPs must be free to rotate and approach unhindered.

To image transcription processes routinely will require tuning of the RNAP-mica interaction, such that the binding-affinity can be temporarily reduced allowing both RNAPs to detach and transcribe the strand. The binding of DNA in liquid has shown to be reversible where different buffers can be exchanged into the imaging cell, to either dissociate or promote binding of the molecules [197]. We would need to look at conditions where the affinity of RNAP for the surface is reduced. Such exchange would have to be quick so as to allow the entire process of colliding RNAPs to be followed using fast-scan AFM at a physiologically relevant time-resolution. If the RNAP remains bound to the surface it will act as an anchor to the attached DNA strand. If the RNAP is temporarily released from the surface, it may be possible that the entire complex is swept away by the motion of the tip. As such it may be necessary to anchor one end of the template to the substrate using a chemical tether.

Although an eventual aim of our studies would be to move away from predominantly simple studies on static systems to dynamic systems containing a number of interacting proteins, further development of the instrumentation and a greater understanding of how adhering biomolecules to a substrate affects biological activity will be required. It may be possible to gain further insights into the convergent transcription process, whilst still imaging samples in ambient conditions.

We observed two main outcomes of forming transcriptionally active complexes on templates containing two promoters on opposite DNA strands. The most common interaction was between a stationary RNAP and one engaged in elongation, while transcription interference between two elongating RNAPs was seen to be rarer. The remainder of complexes were undetermined. These scenarios were inferred by studying the relative positions of the RNAPs on the template after collision. To study whether the RNAPs behave in a similar manner after collision in each mode it would be useful to separate the two scenarios. Currently we have no control over when and from which promoter the RNAP will fire, and hence the type of collision. Such a filtering of the results would require a redesign of the template.

Currently the two promoters used are identical, and on average which RNAP enters elongation first will depend on which arrives at the promoter first on average. If each RNAP arrives at the same time then collisions between two moving RNAPs are expected to be more probable. To only allow collisions between an elongating complex and a sitting duck requires a situation where firing from one promoter is prohibited. This could be done by replacing one of the λ_{PR} promoters with a different promoter sequence, and using an equimolar mix of *E. coli* RNAPs with two different sigma factors. By using the σ^{70} factor we have a relatively simple system that does not require any additional protein factors for the RNAP to be transcriptionally active. This could be mixed with RNAP $\cdot \sigma^{54}$ holoenzyme, a sigma factor which allows for promoter recognition, but requires additional proteins to begin transcribing the DNA [179]. If we have two different promoters that are specific to each sigma factor, then its associated RNAP will only form specific contacts with one location on the chain.

On addition of NTPs the σ^{70} promoter will be able to begin RNA synthesis, whereas by omitting the necessary protein factors the σ^{54} promoter will remain stationary, permanently arrested. Hence, any sort of transcription interference will always be between an elongating complex and a sitting duck.

It is perhaps more difficult to only induce collisions between two RNAPs both engaged in transcription. On average, the time spent stationary at the promoter is greater than it takes for the RNAP to transcribe across the inter-promoter region. One way of making EC-EC interactions more common would be to experiment with increasing the inter-promoter distance of the template. The increased length will mean a greater time will elapse before an RNAP experiences an interaction with the other. This may mean that a RNAP arriving later at its promoter will still have time to finish transcription initiation, and also begin transcribing the template.

Other future work could centre on optimising the labelling reaction. Currently, the method of stem-loop attachment we have developed led to around half of the molecules imaged under AFM having an end-feature. Each DNA sample contains a mixture of unlabelled and labelled components, meaning that it may be necessary to discard DNA·RNAP complexes from any analysis if there is no end-feature on the chain. Also, as the labelling reaction uses a single cycle of primer extension there is no amplification of the template, meaning a large amount of starting material is required, and at best this represents the maximum possible yield of labelled template.

Traditional PCR protocols, whereby a small amount of template is used as starting point for amplification, could not be used for the attachment of the stem-loop label. For exponential amplification of a template, two primers (one forward and one reverse) must be used in the reaction mix. However, to achieve labelling we must use just a single primer, a consequence of *Taq.* polymerase possessing 5' to 3' exonuclease activity, which cleaves nucleotides from the end (exo) of a chain one at a time.

In a conventional PCR reaction, the first cycle of primer extension will lead to the attachment of the loop structure attached to the forward primer. The next

cycle will involve another round of chain denaturation. After this stage the extra end sequence attached in the first cycle should be able to refold to form a stem-loop once again. Annealing of a reverse primer, and subsequent chain extension will lead to the polymerase moving towards the end of the chain where the stem of the label is attached. Due to the exonuclease activity of *Taq*, the phosphodiester bonds linking the stem to the other strand will begin to be degraded, meaning that the stem-loop forming sequence will exist entirely as a single-stranded region at the chain end. This will now act as a template for the polymerase to catalyse the addition of dNTPs, with the final end result being a longer than original dsDNA molecule, with no end-label.

There are however other *Taq* polymerases for use with PCR that have been modified and are deficient in the 5' to 3' exonuclease ability [114]. In such a case, once an extending polymerase originating from a reverse primer meets the end of the attached stem region of the label rather than beginning to cleave the chain, it will simply drop off leaving a complete double-stranded region, together with the intact label. There are two main advantages to such an approach. As the process now involves direct amplification of the template, you need only a small amount of starting material to get a large amount of product. Perhaps more important is that due to the specificity of the two primers, nearly all the DNA present at the end of the reaction will contain the forward primer sequence together with the end-label. Unlabelled molecules will become a tiny proportion of the overall population, meaning that the efficiency of labelling will be improved. As such, molecules seen under AFM should be entirely labelled. Such PCR experiments will require investigations into the best conditions with which to attach the primers whilst allowing the stem-loops to form before extension.

Chapter 8

8 Conclusions

The arrangement of genes in a genome follows a complex spatial arrangement. Consequentially, interactions between different genes during expression are widespread and complex. Transcriptional interference (TI) occurs when one transcriptional process or RNAP has a negative impact on another. One particular example where TI will have significant consequences, with regard to effective expression, is for nested genes. A nested gene is defined as a gene that is totally located within the boundaries of a larger host gene. Such arrangements have an ability to control gene expression at a fundamental level, and pose interesting biological questions such as whether they are expressed simultaneously, co-regulated, or whether the proteins encoded may share a related function.

The work in this thesis has primarily focused on studying the interactions involved in DNA transcription. Traditional biochemical ensemble techniques are able to gain useful information about a large number of molecules, e.g. whether a gene may be up- or down-regulated in certain conditions. However, important sub-groups of the population may be missed as the measurements represent ensemble averages, and are better suited to studying large homogenous populations. Processes that contain a number of distinct stages and intermediates may be better studied using single-molecule techniques. We have used linear DNA templates containing multiple promoter sites to drive RNAPs towards each other, and observed the outcomes using atomic force microscopy (AFM). This approach can be seen as a simple experimental model for nested genes found *in vivo*. AFM provides topographical images of samples with molecular resolution, and is a powerful tool for visualising transcription from nested genes.

Initially, the imaging of DNA samples bound to mica substrates was studied to gain an understanding of how the surface-bound conformations observed represent the population in solution. The deposition and binding of DNA to mica was investigated, in which it was confirmed that the transport of molecules was governed by a diffusion mechanism. The binding of DNA to mica can be controlled through ion-exchanging the mica with divalent cations. Two extremes of binding were observed; measurements of the end-to-end distance of linear DNA molecules discriminate whether the binding mechanism occurs through 2D surface equilibration or kinetic trapping. A range of linear fragments were used to investigate length dependences of binding. Mica, ion exchanged with Ni(II) usually gives rise to kinetically-trapped DNA molecules, however short linear fragments (< 800bp) are seen to deviate from the expected behaviour. This indicated that ion-exchanged mica is heterogeneous, and contains patches or domains, separating different ionic species, and potentially allowing the binding of DNA to be controlled more readily.

A systematic study on the effects of changing the relative humidity in the imaging chamber of the AFM while scanning over DNA was performed. Here, we also investigated cation mediated binding further. A range of different DNA test samples were imaged using two methods of binding. This study illustrated that local molecular motion of DNA can be induced when imaging at high humidity on hydrophilic surfaces. The extent of the conformational changes that arose depended on the amount of superhelical stress stored by the DNA molecules prior to binding. The outcomes were also modulated by the surface binding mode (surface equilibration vs kinetic trapping). Controlled humidity imaging may be useful for studying DNA structure and dynamics, as it is possible to distinguish between different topoisomers. DNA supercoiling is important in a number of biological processes including transcription.

Two linear templates were used for initial investigations into transcription, differing only in the arrangement of the two promoters: pDSU contained convergently aligned promoters, and pDSP had its promoters in a tandem arrangement. These modelled different classes of nested gene, the more common orientation with two promoters on opposite strands, and another with

the promoters in the same direction on one strand. A “snap-shot” approach was taken, by which complexes representing different stages of the reaction were formed. By altering the NTP subset available, RNAPs are allowed to collide and were imaged before and after collision with AFM. Both templates behaved similarly with RNAPs being unable to pass each other after collision, and remaining stalled next to each other. Measurements, as to how the positions of the RNAPs on the template developed during the reaction, helped to confirm that in the convergent promoter arrangement one RNAP could cause the other to back-track. Such observations demonstrate that nested genes and the resulting convergent transcription may have a fundamental role in gene expression.

Whilst it is possible to identify promoter-bound RNAPs fairly easily, once they fire from their respective promoters and begin to interact with one another, it becomes increasingly difficult to identify which protein belonged to which promoter. As such, measurements may be misidentified, particularly in the *ex-situ* measurements made herein. We looked into ways of end-labelling the template, such that its polarity is known, and providing that the RNAPs do not pass each other, the promoter from which it originated from can be identified. A nucleic acid based approach to end-labelling is desirable because it does not compromise the sample preparation procedures for biomolecular AFM that protein labels may cause. We developed a method involving oligonucleotide loop-primed synthesis for the end labelling. The stem-loop appeared as a globular structure at the chain-end, and allowed much better discrimination between outcomes of the collision events. We discovered that collision-induced backtracking is a much more common outcome of convergent transcription. By looking at the positions of RNAPs on the template we were able to conclude that most collisions occur between an elongating complex and one that is stationary at the promoter engaged in initiation, termed a sitting duck. Collisions between two RNAPs actively engaged in transcription were much rarer. End-labelling using nucleic acid structures does not interfere with AFM sample preparation or transcription and could be used as a generic approach to studying interactions of multiple proteins on DNA templates at the single molecule level.

In short, we were able to expand on previous studies of DNA-protein interactions using AFM. In particular we were able to demonstrate the effectiveness of single-molecule techniques for investigating distinct members of a population, and showed that the presence of nested genes could have a number of effects on the efficiency of expression of the larger host gene. Further developments in real-time AFM techniques may allow much greater insights into processes such as convergent transcription to be obtained.

References

1. http://mol-biol4masters.masters.grkraj.org/html/Deoxy_Ribonucleic_Acid2-Structure.htm.
2. Abrescia, N.G.A., T. Huynh-Dinh, and J.A. Subirana, Nickel-guanine interactions in DNA: crystal structure of nickel-d[CGTGTACACG](2). *Journal of Biological Inorganic Chemistry*, 2002. **7**(1-2): p. 195-199.
3. Adelman, K., et al., Single molecule analysis of RNA polymerase elongation reveals uniform kinetic behavior. *Proceedings of the National Academy of Sciences*, 2002. **99**(21): p. 13538-13543.
4. Albrecht, T.R., et al., Frequency modulation detection using high-Q cantilevers for enhanced force microscope sensitivity. *Journal of Applied Physics*, 1991. **69**(2): p. 668-673.
5. Alessandrini, A. and P. Facci, AFM: a versatile tool in biophysics. *Measurement Science And Technologies*, 2005. **16**: p. R65-R92.
6. Amouyal, M. and H. Buc, Topological unwinding of strong and weak promoters by RNA polymerase : A comparison between the lac wild-type and the UV5 sites of Escherichia coli. *Journal of Molecular Biology*, 1987. **195**(4): p. 795-808.
7. Ando, T., et al., A high-speed atomic force microscope for studying biological macromolecules. *Proceedings of the National Academy of Sciences of the United States of America*, 2001. **98**(22): p. 12468-12472.
8. Argaman, M., et al., Phase imaging of moving DNA molecules and DNA molecules replicated in the atomic force microscope. *Nucleic Acids Research*, 1997. **25**(21): p. 4379-4384.
9. 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.
10. Assis, R., et al., Nested genes and increasing organizational complexity of metazoan genomes. *Trends in Genetics*, 2008. **24**(10): p. 475-478.
11. Avery, O.T., C.M. MacLeod, and M. McCarty, Studies on the chemical nature of the substance inducing transformation of pneumococcal types. *The Journal of Experimental Medicine*, 1944. **79**(2): p. 137-158.
12. Bai, L., T.J. Santangelo, and M.D. Wang, Single-molecule analysis of RNA polymerase transcription. *Annual Review of Biophysics and Biomolecular Structure*, 2006. **35**: p. 343-360.
13. Bar-Nahum, G., et al., A Ratchet Mechanism of Transcription Elongation and Its Control. *Cell*, 2005. **120**(2): p. 183-193.
14. Bates, A.D. and A. Maxwell, DNA Topology. 1993: IRL Press.
15. Batey, R.T., R.P. Rambo, and J.A. Doudna, Tertiary motifs in RNA structure and folding. *Angewandte Chemie-International Edition*, 1999. **38**(16): p. 2327-2343.
16. Beadle, G.W. and E.L. Tatum, Genetic Control of Biochemical Reactions in Neurospora. *Proceedings of the National Academy of Sciences*, 1941. **27**(11): p. 499-506.

17. Beaglehole, D. and H.K. Christenson, Vapor adsorption on mica and silicon - entropy effects, layering, and surface forces. *Journal of Physical Chemistry*, 1992. **96**(8): p. 3395-3403.
18. Behrend, O.P., et al., Intermittent contact: tapping or hammering? *Applied Physics A*, 1998. **66**: p. S219-S221.
19. Berg, J. and G.A.D. Briggs, Nonlinear dynamics of intermittent-contact mode atomic force microscopy. *Physical Review B*, 1997. **55**(22): p. 14899.
20. Berg, O.G., R.B. Winter, and P.H. von Hippel, Diffusion-driven mechanisms of protein translocation on nucleic-acids .1. Models and theory. *Biochemistry*, 1981. **20**(24): p. 6929-6948.
21. Binnig, G., C.F. Quate, and C. Gerber, Atomic Force Microscope. *Physical Review Letters*, 1986. **56**: p. 930-933.
22. Binnig, G. and H. Rohrer, Scanning Tunneling Microscopy. *Surface Science*, 1985. **152/153**: p. 17-26.
23. Blackman, G.S., C.M. Mate, and M.R. Philpott, Interaction forces of a sharp tungsten tip with molecular films on silicon surfaces. *Physical Review Letters*, 1990. **65**(18): p. 2270.
24. Bloomfield, V.A., DNA condensation by multivalent cations. *Biopolymers*, 1997. **44**(3): p. 269-282.
25. Burgess, R.R., et al., Factor Stimulating Transcription by RNA Polymerase. *Nature*, 1969. **221**(5175): p. 43-46.
26. 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.
27. Bustamante, C., et al., Entropic elasticity of lambda-phage DNA. *Science*, 1994. **265**(5178): p. 1599-1600.
28. Bustamante, C., et al., Circular DNA-molecules imaged in air by scanning force microscopy. *Biochemistry*, 1992. **31**(1): p. 22-26.
29. Butt, A.L.W.a.M.K.a.S.K.a.V.G.a.H.-J., Deformation and height anomaly of soft surfaces studied with an AFM. *Nanotechnology*, 1993. **4**(2): p. 106.
30. 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.
31. Carpousis, A.J. and J.D. Gralla, Cycling of ribonucleic-acid polymerase to produce oligonucleotides during initiation in vitro at the lac UV5 promoter. *Biochemistry*, 1980. **19**(14): p. 3245-3253.
32. 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.
33. Chargaff, E., et al., The composition of the desoxyribonucleic acid of salmon sperm. *Journal of Biological Chemistry*, 1951. **192**(1): p. 223-230.
34. Chargaff, E., S. Zamenhof, and C. Green, Composition of human desoxypentose nucleic acid. *Nature*, 1950. **165**(4202): p. 756-757.
35. Chastain, M. and I. Tinoco, Structural elements in RNA. *Progress in Nucleic Acid Research and Molecular Biology*, 1991. **41**: p. 131-177.

36. Chatterjee, A.J., C. M.; Shu, C. C., Convergent transcription confers a bistable switch in *Enterococcus faecalis* conjugation. *Proceedings of the National Academy of Sciences of the United States of America*, 2011. **108**(23): p. 5.
37. Cherny, D.I. and T.M. Jovin, Electron and scanning force microscopy studies of alterations in supercoiled DNA tertiary structure. *Journal of Molecular Biology*, 2001. **313**(2): p. 295-307.
38. Chien, A., D.B. Edgar, and J.M. Trela, Deoxyribonucleic-acid polymerase from extreme thermophile *Thermus-aquaticus*. *Journal of Bacteriology*, 1976. **127**(3): p. 1550-1557.
39. Chowdhury, D., et al., Intra-cellular traffic: bio-molecular motors on filamentary tracks. *Eur. Phys. J. B*, 2008. **64**(3-4): p. 593-600.
40. Christenson, H.K., Adhesion and surface-energy of mica in air and water. *Journal of Physical Chemistry*, 1993. **97**(46): p. 12034-12041.
41. Clemmer, C.R. and T.P. Beepe Jr., Graphite: A Mimic for DNA and other Biomolecules in Scanning Probe Microscopy Studies. *Science*, 1991. **251**: p. 640-642.
42. Colchero, J., et al., Observation of liquid neck formation with scanning force microscopy techniques. *Langmuir*, 1998. **14**(9): p. 2230-2234.
43. Corden, J., et al., Promoter sequences of eukaryotic protein-coding genes. *Science*, 1980. **209**(4463): p. 1406-1414.
44. Coury, J.E., et al., Scanning force microscopy of small ligand nucleic acid complexes: Tris(o-phenanthroline)ruthenium(II) as a test for a new assay. *Journal of the American Chemical Society*, 1997. **119**(16): p. 3792-3796.
45. Craig, M.L., W.C. Suh, and M.T. Record, HO.bul. and DNase-I probing of E-sigma(70) RNA polymerase-lambda-PR promoter open complexes - Mg²⁺ binding and its structural consequences at the transcription start site. *Biochemistry*, 1995. **34**(48): p. 15624-15632.
46. Cramer, P., D.A. Bushnell, and R.D. Kornberg, Structural Basis of Transcription: RNA Polymerase II at 2.8 angstrom Resolution. *Science*, 2001. **292**(5523): p. 1863-1876.
47. 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.
48. Crampton, N., et al., Formation of Aminosilane-Functionalized Mica for Atomic Force Microscopy Imaging of DNA. *Langmuir*, 2005. **21**(17): p. 7884-7891.
49. Crampton, N., et al., DNA looping and translocation provide an optimal cleavage mechanism for the type III restriction enzymes. *Embo Journal*, 2007. **26**(16): p. 3815-3825.
50. 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**: p. 133-138.
51. Crampton, N., et al., Fast-scan atomic force microscopy reveals that the type III restriction enzyme EcoP151 is capable of DNA translocation and looping. *Proceedings of the National Academy of Sciences of the United States of America*, 2007. **104**(31): p. 12755-12760.
52. Crick, F.H., On protein synthesis. *Symp Soc Exp Biol*, 1958. **12**: p. 138-63.

53. Crick, F.H.C. and J.D. Watson, The complementary structure of deoxyribonucleic acid. *Proceedings of the Royal Society of London Series A-Mathematical and Physical Sciences*, 1954. **223**(1152): p. 80-&.
54. Dame, R.T., C. Wyman, and N. Goosen, Insights into the regulation of transcription by scanning force microscopy. *Journal of Microscopy-Oxford*, 2003. **212**: p. 244-253.
55. Darst, S.A., E.W. Kubalek, and R.D. Kornberg, Three-dimensional structure of Escherichia coli RNA polymerase holoenzyme determined by electron crystallography. *Nature*, 1989. **340**(6236): p. 730-732.
56. DeHaseth, P.L., M.L. Zupancic, and M.T. Record, RNA polymerase-promoter interactions: the comings and goings of RNA polymerase. *Journal of Bacteriology*, 1998. **180**(12): p. 3019-3025.
57. Drew, H., et al., High-salt d(CpGpCpG), a left-handed Z' DNA double helix. *Nature*, 1980. **286**(5773): p. 567-573.
58. Driscoll, R.J., M.G. Youngquist, and J.D. Baldeschwieler, Atomic-scale imaging of DNA using scanning tunnelling microscopy. *Nature*, 1990. **346**: p. 294-296.
59. Duguid, J., et al., Raman-spectroscopy of DNA-metal complexes .1. Interactions and conformational effects of the divalent-cations - Mg, Ca, Sr, Ba, Mn, Co, Ni, Cu, Pd, and Cd. *Biophysical Journal*, 1993. **65**(5): p. 1916-1928.
60. 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.
61. 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.
62. Epshtein, V., et al., An allosteric mechanism of Rho-dependent transcription termination. *Nature*, 2010. **463**(7278): p. 245-249.
63. Epshtein, V. and E. Nudler, Cooperation Between RNA Polymerase Molecules in Transcription Elongation. *Science*, 2003. **300**(5620): p. 801-805.
64. Epshtein, V., et al., Transcription through the roadblocks: the role of RNA polymerase cooperation. *EMBO Journal*, 2003. **22**(18): p. 4719-4727.
65. Erie, D.A., et al., Multiple RNA polymerase conformations and GreA: control of the fidelity of transcription. *Science (New York, N.Y.)*, 1993. **262**(5135): p. 867-73.
66. Escude, C., et al., Multiple topological labeling for imaging single plasmids. *Analytical Biochemistry*, 2007. **362**(1): p. 55-62.
67. Falk, M., K.A. Hartman, and R.C. Lord, Hydration of deoxyribonucleic acid .3. A spectroscopic study of effect of hydration on structure of deoxyribonucleic acid. *Journal of the American Chemical Society*, 1963. **85**(4): p. 391-&.
68. Fang, Y., T.S. Spisz, and J.H. Hoh, Ethanol-induced structural transitions of DNA on mica. *Nucleic Acids Research*, 1999. **27**(8): p. 1943-1949.
69. Feng, X.Z., et al., Conformational transition in DNA on a cold surface. *Nucleic Acids Research*, 2000. **28**(2): p. 593-596.
70. Franklin, R.E. and R.G. Gosling, Molecular Configuration in Sodium Thymonucleate. *Nature*, 1953. **171**(4356): p. 740-741.

71. Fujimoto, B.S. and J.M. Schurr, Monte Carlo simulations of supercoiled DNAs confined to a plane. *Biophysical Journal*, 2002. **82**(2): p. 944-962.
72. Fukuma, T. and S.P. Jarvis, Development of liquid-environment frequency modulation atomic force microscope with low noise deflection sensor for cantilevers of various dimensions. *Review of Scientific Instruments*, 2006. **77**(4).
73. Garcia, R. and R. Perez, Dynamic atomic force microscopy methods. *Surface Science Reports*, 2002. **47**(6-8): p. 197-301.
74. Geron-Landre, B., T. Roulon, and C. Escude, Stem-loop oligonucleotides as tools for labelling double-stranded DNA. *The FEBS journal*, 2005. **272**(20): p. 5343-52.
75. Gibson, C.W., et al., Nested genes: Biological implications and use of AFM for analysis. *Gene*, 2005. **350**(1): p. 15-23.
76. Giessibl, F.J., Atomic Resolution of the Silicon (111)-(7x7) Surface by Atomic Force Microscopy. *Science*, 1995. **267**(5194): p. 68-71.
77. Gilmore, J.L., et al., Single-Molecule Dynamics of the DNA-EcoRII Protein Complexes Revealed with High-Speed Atomic Force Microscopy. *Biochemistry*, 2009. **48**(44): p. 10492-10498.
78. Gnatt, A.L., et al., Structural Basis of Transcription: An RNA Polymerase II Elongation Complex at 3.3 Å... Resolution. *Science*, 2001. **292**(5523): p. 1876-1882.
79. Gromer, A., M. Rawiso, and M. Maaloum, Visualization of hydrophobic polyelectrolytes using atomic force microscopy in solution. *Langmuir*, 2008. **24**(16): p. 8950-8953.
80. Guajardo, R. and R. Sousa, A model for the mechanism of polymerase translocation. *Journal of Molecular Biology*, 1997. **265**(1): p. 8-19.
81. 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 of the United States of America*, 1994. **91**(26): p. 12927-12931.
82. 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.
83. Hansma, H.G., et al., Polymerase activities and RNA structures in the atomic force microscope. *Journal of Structural Biology*, 1999. **127**(3): p. 240-247.
84. 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. 1933-1939.
85. Hansma, H.G., et al. *Progress in sequencing deoxyribonucleic-acid with an atomic force microscope*. in *5th International Conf on Scanning Tunneling Microscopy/Spectroscopy (Stm 90) / 1st International Conf on Nanometer Scale Science and Technology (Nano 1)*. 1990. Baltimore, Md: Amer Inst Physics.
86. Hansma, P.K., et al., Tapping mode atomic force microscopy in liquids. *Applied Physics Letters*, 1994. **64**: p. 1738-1740
87. Henikoff, S., et al., Gene within a gene - nested drosophila genes encode unrelated proteins on opposite DNA strands. *Cell*, 1986. **44**(1): p. 33-42.

88. 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. 149-176.
89. 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.
90. Herskovi, T.T. and J.P. Harringt, Solution studies of nucleic-acid bases and related model compounds - solubility in aqueous alcohol and glycol solutions. *Biochemistry*, 1972. **11**(25): p. 4800-&.
91. Hobbs, J.K., C. Vasilev, and A.D.L. Humphris, Real time observation of crystallization in polyethylene oxide with video rate atomic force microscopy. *Polymer*, 2005. **46**(23): p. 10226-10236.
92. Hsu, L.M., Promoter clearance and escape in prokaryotes. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, 2002. **1577**(2): p. 191-207.
93. Hu, J., et al., The structure of molecularly thin films of water on mica in humid environments. *Surface Science*, 1995. **344**(3): p. 221-236.
94. Hu, J., et al., Artificial DNA patterns by mechanical nanomanipulation. *Nano Letters*, 2002. **2**(1): p. 55-57.
95. Humphris, A.D.L., J.K. Hobbs, and M.J. Miles, Ultra high-speed scanning probe microscopy capable of over 100 frames per second. *Scanning Tunneling Microscopy/Spectroscopy and Related Techniques*, 2003. **696**: p. 79-85.
96. Hurwitz, J., The Discovery of RNA Polymerase. *J. Biol. Chem.*, 2005. **280**(52): p. 42477-42485.
97. Israelachvili, J.N., Intermolecular and Surface Forces Second ed. 1991: Academic Press.
98. Ivanov, V.I., D.Y. Krylov, and M.J.L.a.J.E.D. David, [6] A-DNA in solution as studied by diverse approaches, in *Methods in Enzymology*. 1992, Academic Press. p. 111-127.
99. Jeon, C. and K. Agarwal, Fidelity of RNA Polymerase II Transcription Controlled by Elongation Factor TFIIS. *Proceedings of the National Academy of Sciences of the United States of America*, 1996. **93**(24): p. 13677-13682.
100. Kapanidis, A.N., et al., Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science*, 2006. **314**(5802): p. 1144-1147.
101. Kasas, S., Atomic Force Microscopy: Biomedical methods and applications, Chapter 25: Sample Preparation Method for Observing RNA Polymerase Activity by Atomic Force Microscopy. 2003: Humana Press.
102. Kasas, S., et al., Escherichia coli RNA polymerase activity observed using atomic force microscopy. *Biochemistry*, 1997. **36**(3): p. 461-468.
103. 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.
104. Kelly Jr, T.J. and H.O. Smith, A restriction enzyme from Hemophilus influenzae: II. Base sequence of the recognition site. *Journal of Molecular Biology*, 1970. **51**(2): p. 393-409.

105. Kim, J.M., et al., AFM phase lag mapping for protein-DNA oligonucleotide complexes. *Analytica Chimica Acta*, 2004. **525**(2): p. 151-157.
106. Kobayashi, M., K. Sumitomo, and K. Torimitsu, Real-time imaging of DNA-streptavidin complex formation in solution using a high-speed atomic force microscope. *Ultramicroscopy*, 2007. **107**(2-3): p. 184-190.
107. Kohonen, M.M. and H.K. Christenson, Capillary condensation of water between rinsed mica surfaces. *Langmuir*, 2000. **16**(18): p. 7285-7288.
108. Koltover, I., K. Wagner, and C.R. Safinya, DNA condensation in two dimensions. *Proceedings of the National Academy of Sciences of the United States of America*, 2000. **97**(26): p. 14046-14051.
109. Komissarova, N. and M. Kashlev, RNA polymerase switches between inactivated and activated states By translocating back and forth along the DNA and the RNA. *The Journal of biological chemistry*, 1997. **272**(24): p. 15329-38.
110. 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 of the United States of America*, 1997. **94**(5): p. 1755-1760.
111. Kumar, A., An Overview of Nested Genes in Eukaryotic Genomes. *Eukaryotic Cell*, 2009 **8**(9): p. 9.
112. Lang, D., Regular superstructures of purified DNA in ethanolic solutions. *Journal of Molecular Biology*, 1973. **78**(2): p. 247-250.
113. 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.
114. 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.
115. Li, J.W., et al., A convenient method of aligning large DNA molecules on bare mica surfaces for atomic force microscopy. *Nucleic Acids Research*, 1998. **26**(20): p. 4785-4786.
116. Li, J.W., et al., Possible multistranded DNA induced by acid denaturation-renaturation. *Journal of Vacuum Science & Technology B*, 1997. **15**(5): p. 1637-1640.
117. Limanskaya, O.Y. and A.P. Limanskii, Imaging of T7 RNA polymerase elongation complexes by atomic force microscopy. *Molecular Biology*, 2008. **42**(3): p. 469-477.
118. Lindsay, S.M., et al., STM and AFM images of nucleosome DNA under water. *Journal of Biomolecular Structure & Dynamics*, 1989. **7**(2): p. 279-287.
119. Linn, S. and W. Arber, Host specificity of DNA produced by *Escherichia coli* .x. In vitro restriction of phage FD replicative form. *Proceedings of the National Academy of Sciences of the United States of America*, 1968. **59**(4): p. 1300-&.
120. Lis, J., Promoter-associated pausing in promoter architecture and postinitiation transcriptional regulation. *Cold Spring Harbor Symposia on Quantitative Biology*, 1998. **63**: p. 9.

121. Liu, L.F. and J.C. Wang, Supercoiling of the DNA-template during transcription. *Proceedings of the National Academy of Sciences of the United States of America*, 1987. **84**(20): p. 7024-7027.
122. Luna, M., J. Colchero, and A.M. Baro, Intermittent contact scanning force microscopy: The role of the liquid necks. *Applied Physics Letters*, 1998. **72**(26): p. 3461-3463.
123. Lyubchenko, Y., et al., Atomic force microscopy of long DNA - imaging in air and under water. *Proceedings of the National Academy of Sciences of the United States of America*, 1993. **90**(6): p. 2137-2140.
124. 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.
125. Maaloum, M., A close encounter with DNA. *European Biophysics Journal*, 2003. **32**(6): p. 585-587.
126. Marmur, A., Tip-surface capillary interactions. *Abstracts of Papers of the American Chemical Society*, 1992. **203**: p. 366-COLL.
127. Martin, Y., C.C. Williams, and H.K. Wickramasinghe, Atomic force microscope force mapping and profiling on a sub 100-Å scale. *Journal of Applied Physics*, 1987. **61**(10): p. 4723-4729.
128. Mate, C.M., et al., Atomic-scale friction of a tungsten tip on a graphite surface. *Physical Review Letters*, 1987. **59**(17): p. 1942-1945.
129. 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.
130. McClure, W.R., Mechanism and Control of Transcription Initiation in Prokaryotes. *Annual Review of Biochemistry*, 1985. **54**(1): p. 171-204.
131. Meselson, M. and R. Yuan, DNA Restriction Enzyme from *E. coli*. *Nature*, 1968. **217**(5134): p. 1110-1114.
132. Metzger, W., P. Schickor, and H. Heumann, A cinematographic view of *Escherichia coli* RNA-polymerase translocation. *EMBO Journal*, 1989. **8**(9): p. 2745-2754.
133. Meyer, G. and N.M. Amer, Novel optical approach to atomic force microscopy. *Applied Physics Letters*, 1988. **53**(12): p. 1045-1047.
134. 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*, 1998. **38**(1): p. 218-225.
135. Minchiotti, G. and P.P. Dinocera, Convergent transcription initiates from oppositely oriented promoters within the 5' end regions of *Drosophila melanogaster* f elements. *Molecular and Cellular Biology*, 1991. **11**(10): p. 5171-5180.
136. Miranda, P.B., et al., Icelike water monolayer adsorbed on mica at room temperature. *Physical Review Letters*, 1998. **81**(26): p. 5876-5879.
137. Moreno-Herrero, F., J. Colchero, and A.M. Baró, DNA height in scanning force microscopy. *Ultramicroscopy*, 2003. **96**(2): p. 167-174.
138. Morgan, T.H., Sex limited inheritance in *Drosophila*. *Science (New York, N.Y.)*, 1910. **32**(812): p. 120-122.
139. Muller, K. and C.C. Chang, Electric dipoles on clean mica surfaces. *Surface Science*, 1969. **14**(1): p. 39-51.

140. Nagami, F., et al., Time-lapse imaging of conformational changes in supercoiled DNA by scanning force microscopy. *Analytical Biochemistry*, 2002. **300**(2): p. 170-176.
141. Neidle, S., *Nucleic Acid Structure*. 2002: Oxford University Press.
142. Neish, C.S., et al., Direct visualization of ligand-protein interactions using atomic force microscopy. *British Journal of Pharmacology*, 2002. **135**(8): p. 1943-50.
143. Nelson, P., Transport of torsional stress in DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 1999. **96**(25): p. 14342-14347.
144. Neubauer, G., et al., Force microscopy with a bidirectional capacitance sensor. *Review of Scientific Instruments*, 1990. **61**(9): p. 2296-2308.
145. Neuman, K.C., et al., Ubiquitous Transcriptional Pausing Is Independent of RNA Polymerase Backtracking. *Cell*, 2003. **115**(4): p. 437-447.
146. 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.
147. Nishimura, S., et al., Molecular-scale structure of the cation modified muscovite mica basal-plane. *Langmuir*, 1994. **10**(12): p. 4554-4559.
148. 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.
149. Pakarinen, O.H., et al., High-resolution scanning force microscopy of gold nanoclusters on the KBr (001) surface. *Physical Review B (Condensed Matter and Materials Physics)*, 2006. **73**(23): p. 235428.
150. Pashley, R.M., DLVO and hydration forces between mica surfaces in Li⁺, Na⁺, K⁺, and Cs⁺ electrolyte solutions: A correlation of double-layer and hydration forces with surface cation exchange properties. *Journal of Colloid and Interface Science*, 1981. **83**(2): p. 531-546.
151. Pashley, R.M. and J.N. Israelachvili, A comparison of surface forces and interfacial properties of mica in purified surfactant solutions. *Colloids and Surfaces*, 1981. **2**(2): p. 169-187.
152. Pastre, 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.
153. Pastre, 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.
154. Pastre, 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.
155. 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.
156. Polyakov, A., E. Severinova, and S.A. Darst, Three-dimensional structure of *E. coli* core RNA polymerase: Promoter binding and elongation conformations of the enzyme. *Cell*, 1995. **83**(3): p. 365-373.

157. Putman, C.A.J., et al., Tapping mode atomic force microscopy in liquid. *Applied Physics Letters*, 1994. **64**: p. 2454-2456.
158. Quate, C.F., The AFM as a tool for surface imaging. *Surface Science*, 1994. **300**(1-3): p. 980-995.
159. Quintana, C. and N. Bonnet, Observation and simulation of unstained DNA images in bright field TEM. *Ultramicroscopy*, 1982. **10**(3): p. 253-261.
160. Rees, W.A., et al., Evidence of dna bending in transcription complexes imaged by scanning force microscopy. *Science*, 1993. **260**(5114): p. 1646-1649.
161. Revyakin, A., et al., Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science*, 2006. **314**(5802): p. 1139-1143.
162. Richardson, J.P., Structural Organization of Transcription Termination Factor Rho. *Journal of Biological Chemistry*, 1996. **271**(3): p. 1251-1254.
163. Ring, B.Z., W.S. Yarnell, and J.W. Roberts, Function of E. coli RNA Polymerase [sigma] Factor- [sigma]70 in Promoter-Proximal Pausing. *Cell*, 1996. **86**(3): p. 485-493.
164. Rippe, K., et al., Transcriptional activation via DNA-looping: Visualization of intermediates in the activation pathway of E-coli RNA polymerase center dot sigma(54) holoenzyme by scanning force microscopy. *Journal of Molecular Biology*, 1997. **270**(2): p. 125-138.
165. 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*, 2000. **87**(1-2): p. 55-66.
166. 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.
167. 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.
168. Rivetti, C., M. Guthold, and C. Bustamante, Wrapping of DNA around the E.coli RNA polymerase open promoter complex. *Embo Journal*, 1999. **18**(16): p. 4464-4475.
169. Roberts, R.J., Restriction and modification enzymes and their recognition sequences. *Gene Amplification and Analysis*, 1987. **5**: p. 1-49.
170. 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.
171. Saiki, R.K., et al., Primer-directed enzymatic amplification of DNA with a thermostable DNA-polymerase. *Science*, 1988. **239**(4839): p. 487-491.
172. 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**: p. 1599-1605.
173. 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**(3-4): p. 151-158.

174. Sanchez, H., R. Kanaar, and C. Wyman, Molecular recognition of DNA-protein complexes: A straightforward method combining scanning force and fluorescence microscopy. *Ultramicroscopy*, 2010. **110**(7): p. 844-851.
175. 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.
176. Saunders, G.C., H.C. Parkes, and S.B. Primrose, Analytical Molecular Biology: Quality and Validation. 1999: RSC Publishing
177. Schickor, P., et al., Topography of intermediates in transcription initiation of Escherichia-coli. *EMBO Journal*, 1990. **9**(7): p. 2215-2220.
178. 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.
179. 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.
180. Seong, G.H., et al., Atomic force microscopy identification of transcription factor NFkappaB bound to streptavidin-pin-holding DNA probe. *Analytical Biochemistry*, 2002. **309**(2): p. 241-247.
181. Shaevitz, J.W., et al., Backtracking by single RNA polymerase molecules observed at near-base-pair resolution. *Nature*, 2003. **426**(6967): p. 684-687.
182. 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.
183. Shearwin, K.E., B.P. Callen, and J.B. Egan, Transcriptional interference a crash course. *Trends in genetics : TIG*, 2005. **21**(6): p. 339-345.
184. Shlyakhtenko, L.S., et al., Atomic Force Microscopy Imaging of DNA Covalently Immobilized on a Functionalized Mica Substrate. *Biophysical journal*, 1999. **77**(1): p. 568-576.
185. Shlyakhtenko, L.S., et al., Specificity of Binding of Single-Stranded DNA-Binding Protein to Its Target. *Biochemistry*, 2012. **51**(7): p. 1500-1509.
186. Shlyakhtenko, L.S., et al., Structure and dynamics of supercoil-stabilized DNA cruciforms. *Journal of Molecular Biology*, 1998. **280**(1): p. 61-72.
187. 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.
188. Sirghi, L., et al., Atomic Force Microscopy Study of the Hydrophilicity of TiO₂ Thin Films Obtained by Radio Frequency Magnetron Sputtering and Plasma Enhanced Chemical Vapor Depositions. *Langmuir*, 2001. **17**(26): p. 8199-8203.
189. Sitko, J.C., E.M. Mateescu, and H.G. Hansma, Sequence-dependent DNA condensation and the electrostatic zipper. *Biophysical Journal*, 2003. **84**(1): p. 419-431.
190. 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.

191. Stadler, L.J., Mutations in barley induced by X-rays and radium. *Science*, 1928. **68**(1756): p. 186-187.
192. 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.
193. Straney, D.C. and D.M. Crothers, A stressed intermediate in the formation of stably initiated RNA chains at the Escherichia coli lac UV5 promoter. *Journal of Molecular Biology*, 1987. **193**(2): p. 267-278.
194. Sushko, M.L., A.L. Shluger, and C. Rivetti, Simple model for DNA adsorption onto a mica surface in 1 : 1 and 2 : 1 electrolyte solutions. *Langmuir*, 2006. **22**(18): p. 7678-7688.
195. Thomas, M.J., A.A. Platas, and D.K. Hawley, Transcriptional Fidelity and Proofreading by RNA Polymerase II. *Cell*, 1998. **93**(4): p. 627-637.
196. Thomson, N.H., *Chapter 15: Atomic Force Microscopy of DNA structure and interactions*, in *Applied Scanning Probe Methods Vol VI: Characterization*, B. Bhushan, H. Fuchs, and S. Hosaka, Editors. 2006, Springer-Verlag: Berlin.
197. Thomson, N.H., et al., Reversible binding of DNA to mica for AFM imaging. *Langmuir*, 1996. **12**(24): p. 5905-5908.
198. Thundat, T., D.P. Allison, and R.J. Warmack, Stretched DNA structures observed with atomic-force microscopy. *Nucleic Acids Research*, 1994. **22**(20): p. 4224-4228.
199. Thundat, T., et al., Atomic force microscopy of deoxyribonucleic acid strands adsorbed on mica: The effect of humidity on apparent width and image contrast. *Journal of Vacuum Science & Technology* 1992. **10**: p. 630-635.
200. Tomanee, P. and J.T. Hsu, Transition between supercoiled and open circular plasmid DNA during alcohol precipitation. *Journal of Liquid Chromatography & Related Technologies*, 2004. **27**(10): p. 1483-1490.
201. Tortonese, M., R.C. Barrett, and C.F. Quate, Atomic resolution with an atomic force microscope using piezoresistive detection. *Applied Physics Letters*, 1993. **62**(8): p. 834-836.
202. Tripathi, T. and D. Chowdhury, Interacting RNA polymerase motors on a DNA track: Effects of traffic congestion and intrinsic noise on RNA synthesis. *Physical Review E*, 2008. **77**(1): p. 011921.
203. van de Water, W. and J. Molenaar, Dynamics of vibrating atomic force microscopy. *Nanotechnology*, 2000. **11**(3): p. 192-199.
204. Velichko, Y.S., K. Yoshikawa, and A.R. Khokhlov, Surface-induced DNA superhelicity. *Biomacromolecules*, 2000. **1**(3): p. 459-465.
205. Vesenska, J., et al., Substrate preparation for reliable imaging of DNA molecules with the scanning force microscope. *Ultramicroscopy*, 1992. **42-44**(Part 2): p. 1243-1249.
206. von Hippel, P.H. and O.G. Berg, Facilitated target location in biological systems. *Journal of Biological Chemistry*, 1989. **264**(2): p. 675-678.
207. von Hippel, P.H. and Z. Pisman, Reaction pathways in transcript elongation. *Biophysical Chemistry*, 2002. **101-102**: p. 401-423.
208. Wang, A.H.J., et al., Molecular-structure of a left-handed double helical DNA fragment at atomic resolution. *Nature*, 1979. **282**(5740): p. 680-686.

209. Watson, J.D. and F.H.C. Crick, Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature*, 1953. **171**(4356): p. 737-738.
210. Wilson, K.S. and P.H. von Hippel, Transcription Termination at Intrinsic Terminators: The Role of the RNA Hairpin. *Proceedings of the National Academy of Sciences of the United States of America*, 1995. **92**(19): p. 8793-8797.
211. Wing, R., et al., Crystal-structure analysis of a complete turn of B-DNA. *Nature*, 1980. **287**(5784): p. 755-758.
212. Xiao, X.D. and L.M. Qian, Investigation of humidity-dependent capillary force. *Langmuir*, 2000. **16**(21): p. 8153-8158.
213. Xu, L., et al., Wetting and capillary phenomena of water on mica. *Journal of Physical Chemistry B*, 1998. **102**(3): p. 540-548.
214. Yarnell, W.S. and J.W. Roberts, Mechanism of Intrinsic Transcription Termination and Antitermination. *Science*, 1999. **284**(5414): p. 611-615.
215. 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.
216. Yokokawa, M., K. Takeyasu, and S.H. Yoshimura. *Fast-scanning AFM is now applicable to the analyses of single-molecule reactions in nanobiophysics*. in *Micro-NanoMechatronics and Human Science, 2005 IEEE International Symposium on*. 2005.
217. Yoshimura, S.H., et al., Molecular mechanisms of DNA end-loop formation by TRF2. *Genes to Cells*, 2004. **9**(3): p. 205-218.
218. Yu, P., D.L. Ma, and M.X. Xu, Nested genes in the human genome. *Genomics*, 2005. **86**(4): p. 414-422.
219. Zhang, G.Y., et al., Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 angstrom resolution. *Cell*, 1999. **98**(6): p. 811-824.
220. Zhong, Q., et al., Fractured polymer/silica fiber surface studied by tapping mode atomic force microscopy. *Surface Science*, 1993. **290**(1-2): p. L688-L692.
221. Zhou, X.F., et al., Direct measurement of compression spring constant of single DNA molecule with AFM. *Chinese Science Bulletin*, 2005. **50**(10): p. 954-957.
222. Zitzler, L., S. Herminghaus, and F. Mugele, Capillary forces in tapping mode atomic force microscopy. *Physical Review B*, 2002. **66**(15).
223. Zuccheri, G., et al., Conformational fluctuations of supercoiled DNA molecules observed in real time with a scanning force microscope. *Applied Physics a-Materials Science & Processing*, 1998. **66**: p. S585-S589.
224. Zuker, M., Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, 2003. **31**(13): p. 3406-3415.