Department Of Physics and Astronomy
University of Sheffield

Dynamic Single Molecule Imaging of DNA-Protein Interactions using Atomic Force Microscopy

Vinny Verma

Submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

at the Department Of Physics and Astronomy

University of Sheffield

2021

Declaration

This thesis is submitted to The University of Sheffield in support of my application for the degree of Doctor of Philosophy. I, the author, confirm that the thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means (www.sheffield.ac.uk/ssid/unfair-means). This work has not previously been presented for an award at this, or any other University.

The work presented herein has been conducted by me, except the following:

1. *Thermus aquaticus* DNA Polymerase I and *Thermus aquaticus* FEN domain producing *E. coli* strains had been produced by Prof. Jon. R. Sayers' group at the University of Sheffield and have been used to produce and purify FEN-inactive proteins (chapters 4 and 7)
2. T_7 FEN (active) had been produced by Prof. Jon. R. Sayers group (chapters 4 and 7) and has been used for dynamic imaging of flap DNA-FEN interactions
3. *parS* plasmid had been provided by Dr. Ling Chin Hwang at the University of Sheffield and has been used for imaging DNA (chapter 8)
4. 0-9 *parS* site containing DNA plasmids had been synthesized by Dr. Satpal Chodha in Dr. Ling Chin Hwang's lab at the University of Sheffield and have been used as a template for PCR to produce DNA for imaging (chapter 8)
5. ParB protein had been synthesized and purified by Dr. Ling Chin Hwang's lab at the University of Sheffield and has been used for imaging *parS*-ParB interactions using AFM (chapter 8).

Abstract

This thesis work applies Atomic Force Microscopy (AFM) for single-molecule imaging to understand the interactions between DNA and DNA-binding proteins. Flap Endonuclease (FEN) and ParB are the two proteins explored here, that bind DNA as an intermediate step in DNA replication and bacterial plasmid segregation respectively. With the objective to study FEN-flap DNA interactions, different immobilization methods, as well as protein and DNA modification methods, have been applied. The images of 100 bp flap DNA have revealed a Y-shaped flap DNA showing a thicker and higher double strand and a thinner less-elevated single strand attached to it. Helical resolution was obtained for DNA immobilized on Ca^{2+} treated mica and the molecules showed the major grooves of DNA. This was noteworthy as the helical resolution was difficult to resolve for such short DNA fragments and these observations have not been reported previously. However, there were no or very scarce ssDNA seen during flap DNA sample imaging, signifying the immobilization technique was not suitable for flap DNA. At a high salt concentration, salt precipitates were seen as noise on the mica surface. Ca^{2+} and Ni^{2+} was discovered to be more effective than Mg^{2+} for immobilization of dsDNA, with Ni^{2+} allowing the least amount of mobility of DNA in the consecutive frames. PLO treated mica proved to be the most suitable surface for DNA immobilization during imaging in buffer conditions, though it was not effective to visualize the short flaps. However, it allowed imaging with helical resolution where both the major and minor grooves of long DNA, like plasmids, was visualized. The imaging was simpler and more efficient if the length of the DNA strands was longer than 100 bp. Hence, the method of assembly of long DNA oligonucleotides to form 300 bp flap DNA with ~ 100 nt flap was applied such that the sample could be imaged with a good resolution and did not mobilize during consecutive scans.

Dynamic imaging of inactive FEN interaction with flap DNA showed the mobilisation of DNA on the mica surface, indicating that protein recognition and interaction affected the DNA conformation and disrupted the forces surrounding DNA that anchored it to the PLO treated mica surface. Active FEN could be seen intermittently mobilizing the DNA, threading it and cleaving the ssDNA branch. A method of motion tracking and quantification of the movement of sections of DNA was developed to allow the segmentation of DNA and tracking of each of the segments to recognise the DNA sections most affected by the protein. It was found that FEN caused the DNA to bend in the vicinity and time duration of its binding, and changes in DNA shape were witnessed in the short time span of the protein's appearance close to DNA, after which the DNA anchored to the mica surface and immobilized. The results corroborated the previously reported theories that FEN binds the dsDNA, bends the ssDNA strand and makes it thread through the arch of the active site. However, we have proposed that FEN initially binds to DNA (both flap and overhang) not at the branch point, as previously proposed by crystallography studies, but anywhere on the DNA and is capable of 'sliding' on it to find the ssDNA junction to which it can anchor and initiate the nuclease activity step.

ParB interactions with *parS* were performed by imaging the reaction mixture immobilized on PLO treated mica and comparing the trace lengths of the DNA molecules. When the ParB protein was bound to DNA, the molecules displayed a 'beadiness' and the length of DNA was reduced in comparison with the length of the DNA negative control. The presence of CTP in the buffer caused the protein to bind to DNA in a sequence non-specific manner, hence caused the sliding of the protein off the DNA. These observations were consistent with the previously reported research, yet how ParB recognised the DNA in the first place in the absence of any *parS* site, remains unknown. Potential future experiments, building from this work, are also discussed.

Acknowledgements

First of all, thank God for everything in my life!

To begin with, I would like to express sincere gratitude to my supervisors, Prof. Jamie K. Hobbs and Prof. Jon R. Sayers for giving me the opportunity to undertake this Ph.D. project. Their continuous guidance, invaluable training, encouragement, patience, and cordiality have been imperative for the completion of this project. I would like to express my heartfelt appreciation for Jamie for being extremely kind and supportive (especially during the COVID 19 pandemic hijacking the second half of the Ph.D.)and always making me feel welcome to ask questions and pose research ideas and for being acceptable to make mistakes. Acknowledgments to Dr. Ling Chin Hwang as well, for her co-supervision in the ParB-*parS* interactions project. I would also like to thank The University of Sheffield, Imagine: Imaging Life Institute and The Florey Institute for funding my Ph.D.

I would like to express thanks to my colleagues in the Hobbs group and Sayers group for their continuous help and advice. Thanks to Emma Brundell and Dr. Sarbendra Pradhananga in Sayers group for their help and initial training in the lab, Dr. Sandip Kumar formerly in the Hobbs group for the initial guidance on AFM and Dr. Nicholas Mullen for constant aid in my every catastrophe in the AFM lab and invaluable discussions on the subject. Special thanks to Dr. Laia Pasquina Lemonche for her unending guidance, ideas and consultations at every unexpected step, and for being available for me at every instance to help, making my Ph.D. life easier! I would also like to acknowledge Dr. Abimbola Feyisara Olulana for her assistance during the writing process. I am extremely grateful also to Dr. Satpal Chodha in Ling's group for his guidance and advice on the ParB-*parS* project, without whose support the work would not have expanded to be the chapter that it became. I would also like to thank my colleagues in Hobbs group, Raveen, Spyros, David, Xinyue, Anaam and Freya for their wonderful care and friendship throughout this journey.

Many thanks to the staff and students of the Imagine and Florey cohort, for their comradeship and assistance in functioning of these cohorts and sharing their experiences of research and organising numerous Symposia and events over the past 4 years. Heartfelt gratitude to the administrative staff at the Department of Physics for their services.

Most importantly, I would like to thank my parents, family and friends for their immense encouragement, understanding and faith in me throughout my life, making me capable to work on a Ph.D. project. It would not have been possible to dream of a Ph.D. or higher studies if not for the hard work, sacrifices and belief of my parents.

Finally, I would like to thank my best friend and husband, Nitin Suri, for being the pillar of strength through all my successful and failed lab experiments, late nights, work burnouts, laziness, and

craziness of research life. I am extremely grateful to him, without whose love, friendship, advice, patience, understanding, support, food and jokes this project would not have been completed!

This work is dedicated to
every child who dares to dream.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

2D	Two Dimensional
3D	Three Dimensional
Å	Angstrom
APS	1-(3-Aminopropyl) Silatrane Ammonium Persulfate
APTES	3-Aminopropyltriethoxysilane
bp	Base Pairs
$CaCl_2$	Calcium Chloride
CTP	Cytidine Triphosphate
DBS	DNA Binding Site
DNA	Deoxy-Ribonucleic Acid
DNA PolI	DNA Polymerase
DNM	Dynamic Nanomechanical Mapping
dsDNA	Double-Stranded Deoxy-Ribonucleic Acid
DTT	Diothiothreitol
E.coli	*Escherichia Coli*
EDTA	Ethylenediaminetetraacetic Acid
FEN	Flap Endonuclease
FISH	Fluorescent In-situ Hybridisation
FlAsH	Fluorescein Arsenical Helix binder
FM	Frequency Modulation
FRET	Förster Resonance Energy Transfer
HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)
HOPG	Highly Oriented Pyrolytic Graphite
HPLC	High-performance liquid chromatography
IEX	Ion-Exchange Chromatography
KCl	Potassium Chloride
mg	Milligram
$MgCl_2$	Magnesium Chloride
mL	Millilitre
N_2	Nitrogen gas
NaCl	Sodium Chloride
$NiCl_2$	Nickel Chloride
nm	Nanometre
nM	Nanomolar
nt	Nucleotide
NTP	Nucleoside Triphosphate
OBD	Optical Beam Deflection system

PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
PEG	Polyethylene Glycol
PEI	Polyethylenimine
PID	Proportional Integral Derivative
PLO	Poly-L-Ornithine
PNK	Polynucleotide Kinase
RBC	Red Blood Cell
ReAsH	Resorufin Arsenical Helix binder
RF	Replicative Form
RNA	Ribonucleic Acid
RPM	Rotations Per Minute
RT	Room Temperature
SDS	Sodium Dodecyl Sulphate
SEM	Scanning Electron Microscopy
SLB	Supported Lipid Bilayers
SP	Sulphopropyl
ssDNA	Single-Stranded Deoxy-Ribonucleic Acid
STM	Scanning Tunnelling Microscope
TaqPol	*Thermus Aquaticus* DNA Polymerase
TEM	Transmission Electron Microscopy
TEMED	Tetramethylethylenediamine
TREC	Topography and Recognition
Tris	Trisaminomethane
µg	Microgram
µL	Microlitre
µm	Micrometre
UV	Ultraviolet
WBC	White Blood Cell

Somewhere, something incredible is waiting to be known.

-*Carl Sagan*

Chapter 1

Introduction

Atomic Force Microscopy (AFM) is a technique that employs a sharp-tipped cantilever to scan over a sample and obtain a topographical image of the surface. The topography depends on a feedback loop that maintains a constant force between the tip and the sample. An image contrast is obtained due to this force, which depends on the tip sample interaction and their respective material properties. These interactions between the tip and the sample are used to study the properties of the sample, tip and the medium. AFM has been used to image various kinds of samples, both hard and soft, in a non-invasive manner without the application of complex sample processing methods. It has proved to be especially useful to image biological samples as it does not damage the soft cells, tissues and biomolecules to be imaged. It additionally provides the advantage of allowing imaging in buffer conditions that mimic the natural environment of the biological samples. Recently, advances in high-speed AFM have also allowed imaging of dynamic processes to understand biomolecular interactions.

In this research work, AFM has been used to image DNA-protein interactions. DNA, the genetic material in most living organisms, undergoes many processes during cell life, primarily DNA replication during cell division and gene expression (transcription and translation) to synthesize proteins as encoded by DNA. During these two processes, numerous proteins, along with ligands and cofactors, interact with DNA (and RNA counterparts) to complete the cellular tasks. An understanding of these interactions is crucial to comprehend, predict, manipulate and control cell functions in case of innumerable health issues such as cancer, neurodegenerative disorders, pathogen infections, organ regeneration and transplant, genetic disorders *etc*. Hence, a lot of research has been undertaken in this direction. With the advancement of imaging techniques, various cellular processes have come to be visualized and have thus aided in the understanding of cell functioning. This project also aims to contribute to this progress to understand the process of cell division.

The division of cells is a very elegant yet complex process that can be bifurcated into two main steps: the replication of the cell's genetic content to be segregated into the daughter cells, and the actual process of cell division. These two steps are highly regulated and coordinated by a

very wide array of proteins, nucleoproteins, ligands and cofactors in the cells as well as environmental triggers. The current work focusses on two such proteins: Flap endonuclease (FEN) and ParB.

FEN is a domain of the DNA Polymerase I, an enzyme that plays an important role in DNA replication. The FEN domain is responsible for DNA repair and nuclease activity to remove ssDNA branches formed during replication. ParB, on the other hand, participates in the partitioning or separation of the duplicated plasmids into daughter cells during cell division in bacteria. An understanding of both of these proteins' functions is essential for the development of treatments for disorders related to abnormal cell division and bacterial infections.

In this research work, AFM has been used to obtain dynamic images of FEN binding and nuclease activity to attempt to visualize the conformation changes effected on the DNA and the protein itself due to the protein's interactions. The technique has also been used to obtain images of ParB binding to DNA containing *parS* sites to assess the mechanism of this interaction. Attempts have been made to standardise the methods of sample immobilization for dynamic imaging suitable to allow AFM imaging of an unfamiliar DNA sample like flap DNA. Additionally, experiments have been performed to synthesize DNA and protein samples especially tailored to be immobilized for AFM imaging. Method of AFM image analysis has been developed to track the motion of DNA molecules due to protein interactions in the consecutive frames during dynamic imaging.

An Outline of Chapters

Chapter 2 describes the details of the structure and function of DNA and proteins FEN and ParB, and reviews the literature discussing the various models for binding and activity of these molecules. It also covers the research that is still required to understand the mechanism of action of these proteins fully.

Chapter 3 summarizes the working principle of the AFM, the various modes of the technique, the basis of the feedback operation controlling the microscope and the numerous cantilever tips, substrate surfaces and imaging environments employed for AFM. Additionally, the advances in the field of high-speed AFM have also been discussed with emphasis on the relevance of the technique to visualize catalysis and biomolecular mechanisms and the future prospects of the technique in the field of high-speed imaging of biological processes.

Chapter 4 details the materials and methods applied in the current research work.

Chapter 5 delves into the development of surface immobilization techniques for imaging short DNA samples of 100 bp length. The methods experimented with include treatment of mica with divalent cations and poly-L-ornithine.

Chapter 6 discusses the experiments performed to synthesize and purify DNA samples of varying lengths optimized for immobilization on the surface for AFM, such that the ssDNA branch could be visualized with an adequate resolution during interaction with FEN.

Chapter 7 examines the experiments and results obtained by dynamic imaging of FEN interaction with flap or overhang DNA. A discussion of the movement and cleaving of DNA due to mobilization caused by protein interactions has been made. Analysis of the movement has been performed by tracking the DNA molecules and quantifying the motion with time under the effect of FEN.

Chapter 8 describes the results obtained for AFM imaging ParB protein interaction with DNA containing a varied number of *parS* sites. An analysis of the variation in the length of DNA as a result of the interaction has been done, to understand the effect of the presence of CTP in the buffer and varying the number of *parS* sites on the binding potential of ParB.

Finally, chapter 9 of Conclusion and Future Prospects summarizes the results obtained in the current work and highlights the steps that could be taken to improve the experimentation and the developments in the technique of high-resolution AFM that could benefit the understanding of biochemical reactions.

Chapter 2

DNA and Proteins

"If you can't explain it to a six-year-old, you don't understand it yourself"- Albert Einstein

2.1. The Structure of DNA

Deoxyribonucleic acid or DNA is the genetic material in most organisms, with RNA (ribonucleic acid) being the genetic material in some viruses. The DNA molecule consists of two sugar-phosphate backbones that are linked via nitrogenous bases bonded by hydrogen bond. Each of the backbone strands is made up of nucleotides, that consist of a phosphate attached to a 2'-deoxyribose sugar molecule via a phosphodiester bond, and the sugar in turn attached to a nitrogenous base via an amino-glycosidic bond (figure 2.1). The nitrogenous bases are of two types: (double-ringed) purines (which are adenine, guanine) and (single ringed) pyrimidines (which are cytosine and thymine, latter replaced by uracil in the case of RNA).

The structure of the DNA double helix was first described by Watson and Crick[1,2], with the aid of electron diffraction studies by Franklin. The molecule is formed by the antiparallel alignment of the two polynucleotide chains, which are connected by hydrogen bonds between the purine of one chain and the pyrimidine of the other chain (figure 2.2). The DNA strand is labelled 5' to 3' as it has the phosphate group attached to the 5' C of deoxyribose sugar at one end, and a hydroxyl group attached to the 3' C of deoxyribose sugar at the other end. The two DNA strands being antiparallel, the DNA double helix has one strand 5' to 3' and other 3' to 5'. The phosphate groups provide a negative charge to the DNA ladder, which can be used for the attachment of DNA onto surfaces, as described later in chapters 3 and 4. The base adenine pairs with thymine and guanine pairs with cytosine, and therefore the two polynucleotide chains must have complementary sequences to be able to form a double strand. The double helix is formed due to the puckering of pentose sugar that causes the entire DNA chain to be tilted, hence wound around one another. The helix is stabilized by base pairing and stacking and has right-handed or left-handed orientation of the turns. The helix has two unequal grooves called the major and the minor grooves (figures 2.2, 2.3; table 2.1). Depending on the direction of rotation and the angle of puckering of the sugar molecules, the DNA can be of A, B, or Z type, B-DNA being the most commonly found in humans.[3,4]

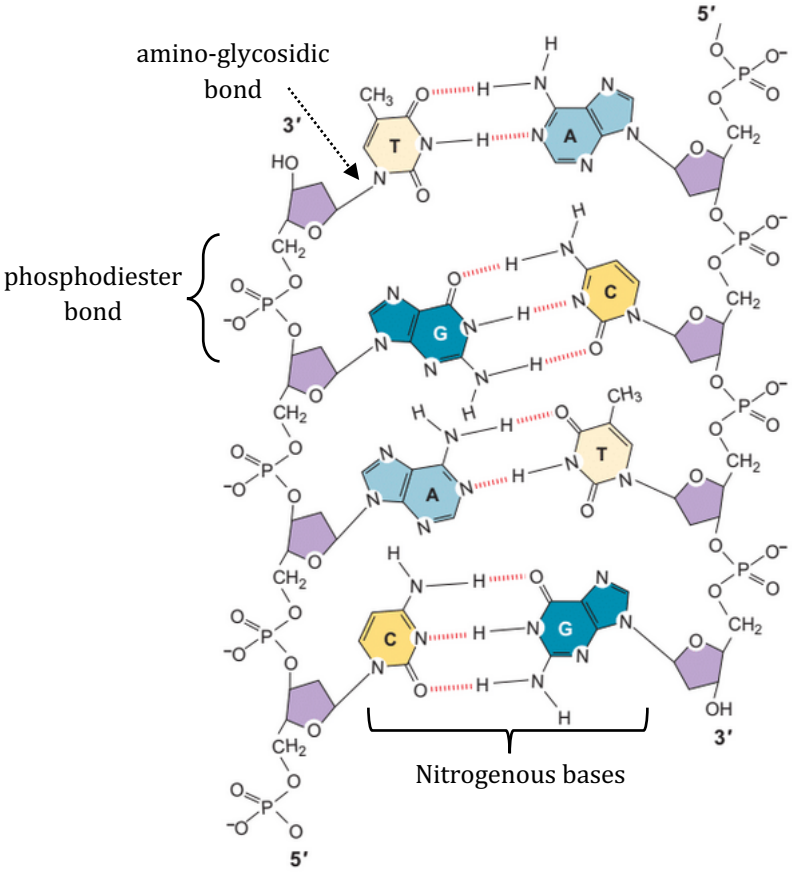


Figure 2.1: Diagrammatic representation of the detailed structure of the DNA double helix, reproduced from the chapter: *The Structure of DNA* [4].

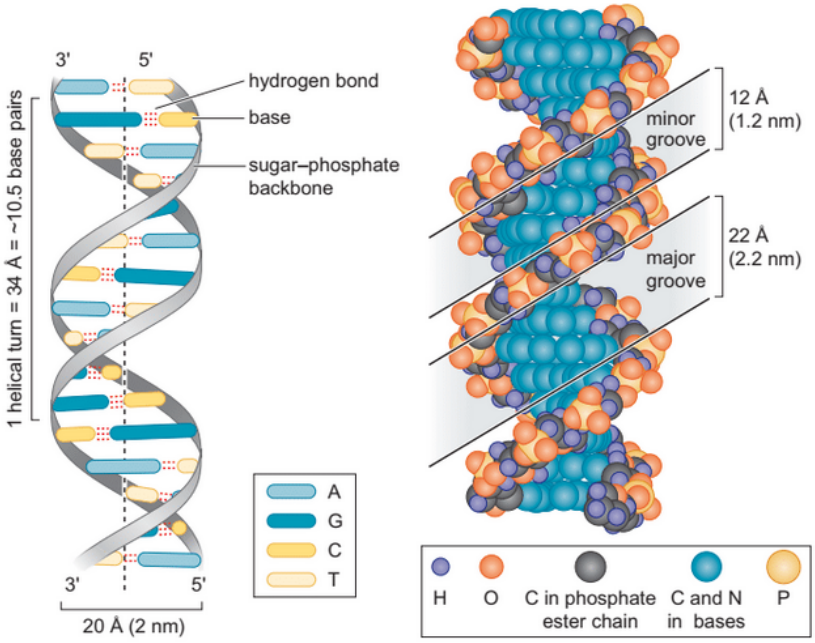


Figure 2.2: a. Schematic of the DNA double helix. b. model of DNA showing the major and minor grooves. reproduced from the chapter: *The Structure of DNA* [4].

	Helix type		
	A	B	Z
Overall proportions	Short and broad	Longer and thinner	Elongated and slim
Rise per base pair	2.3 Å	3.32 Å	3.8 Å
Helix pack diameter	25.5 Å	23.7 Å	18.4 Å
Base pairs per helix repeat	1	1	2
Base pairs per helix turn	~11	~10	12
Helical pitch	24.6 Å	33.2 Å	45.6 Å

Table 2.1: Comparison of the structures of A, B and Z DNA, as obtained by single crystal X-ray analysis. reproduced from the chapter: *The Structure of DNA* [4].

To understand the conformation changes of DNA in various buffer conditions and under the influence of various proteins, many measurements can be made during the study of DNA behaviour using imaging techniques. These techniques will be discussed in chapter 3. Numerous proteins interact with DNA and alter its flexibility by bending and looping, which can be quantified by the term 'persistence length'. Accurate determination of variations in the observed DNA persistence length under the action of proteins provides insight into the functioning of these cellular biomolecules[5]. The proteins causing the bending of DNA thereby change the persistence length and this can be used as a measure of the proteins' activity.

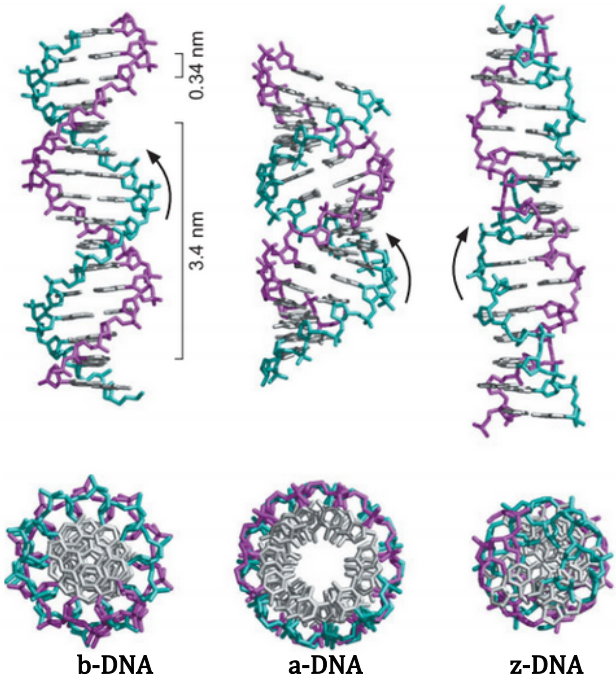


Figure 2.3: Comparison of the structure of B, A and Z DNA: top images show the lateral view, below is the horizontal cross section[4]. B-DNA is the narrower DNA with more prominent major groove, A DNA is the broader and shorter one with broad and not very deep major groove, Z-DNA is slim and compact with almost indistinguishable major and minor grooves. reproduced from the chapter: *The Structure of DNA.*

Persistence length denotes the stiffness of a molecules, measured as the average contour length between two points where the molecule direction changes by one radian. For DNA, the persistence length inversely relates to the DNA flexibility [6]. In case of AFM imaging, this parameter can be used to distinguish between

different types of DNA molecules, surfaces of immobilization [7] and solution compositions in which the sample was immobilized.

The cause of the stiffness of DNA is widely believed to be repulsions between the phosphate residues of the DNA backbone [8]. Although it has been reported that decrease in ionic concentration increases the persistence length of DNA[9] or reduces the flexibility of DNA[8], this has been refuted in studies by Garai *et al* [10]. It has been proposed that at high concentration of ions, the repulsion between the DNA phosphate residues reduces but the electrostatic repulsion within the strand increases. At a higher concentration of cations and their subsequent binding to the grooves by hydrophobic and polar interactions, the DNA becomes stiffer and resist bending, thus increasing the persistence length[10]. Much work has been done and numerous models proposed on the effect of ions on the persistence length of DNA, which though not in the scope of the current work, allow an understanding of the DNA strand conformation changes[7,9,11–25].

Proteins interact with DNA either in a sequence specific manner, where the protein is able to recognise the bases of DNA to interact with them, or in a sequence non-specific manner. In case of the sequence specific interactions, the protein recognising the DNA sequence usually contains helix-turn-helix (HTH) motif that interacts with the major groove of DNA[26]. On the contrary, the sequence non-specific interactions occur by the recognition of distinct shape or structure of the substrate DNA.

In this project, AFM has been used to study the interaction of two kinds of proteins with DNA: firstly flap endonuclease (FEN), that interacts with DNA in a sequence non-specific manner by recognising the substrate structure, and secondly, ParB that recognise a specific sequence of DNA. Both the proteins play an important role in the process of genome replication and segregation during cell division, the details of which are discussed in sections 2.2 and 2.3. The rest of this chapter describes the structures of these proteins and the proposed models for protein binding to the substrate, and summarizes some of the literature in the advancements of the understanding of protein functions.

2.2. Flap Endonuclease (FEN)

DNA polymerases (of various types in prokaryotic and eukaryotic cells) are the chief enzymes employed for DNA replication during cell division. The enzymes' roles can be summarized to be primarily three-fold: DNA polymerisation (primarily DNA Polymerase I and III [27]), 3'-5' proofreading exonuclease activity and 5'-3' exonuclease activity (DNA Polymerase I) [28,29], the domain of which is also called the Flap endonuclease (FEN). The exonuclease activities of the enzyme allow proofreading during DNA replication and repair of damaged DNA. These enzymes are conserved across all species[30] and paralogous FEN-encoding genes have been found in many eubacteria[31].

2.2.1. Structure and Function of FEN

Limited proteolysis of *E.coli* DNA polymerase I produces a large fragment (Klenow fragment) with polymerase and 3' - 5' exonuclease activity, and a small fragment with 5' -3' exonuclease activity, also called the Flap endonuclease (FEN) (figure 2.4[32]). Studies have shown that FEN is limited to the amino acids 1-298 of PolI (see appendix 1 for details of protein sequence), and is essential for cell viability[33]. In eukaryotic cells, the 42 kDa FEN-1 has both the structure specific 5' endonucleolytic DNA cleavage activity (flap DNA cleaving)[34–37] and dsDNA 5'-exonuclease activity, *i.e.* exonucleolytic hydrolysis of blunt-ended duplex DNA substrates [38]. It also recognises and cleaves a branched DNA structure called DNA flap, and its derivative, called a pseudo Y-structure, acting as a phosphodiesterase, with divalent metal ion dependency[39–42]. These activities are important for excision repair of damaged DNA and processing of Okazaki fragments [31,33,43] (see figure 2.5).

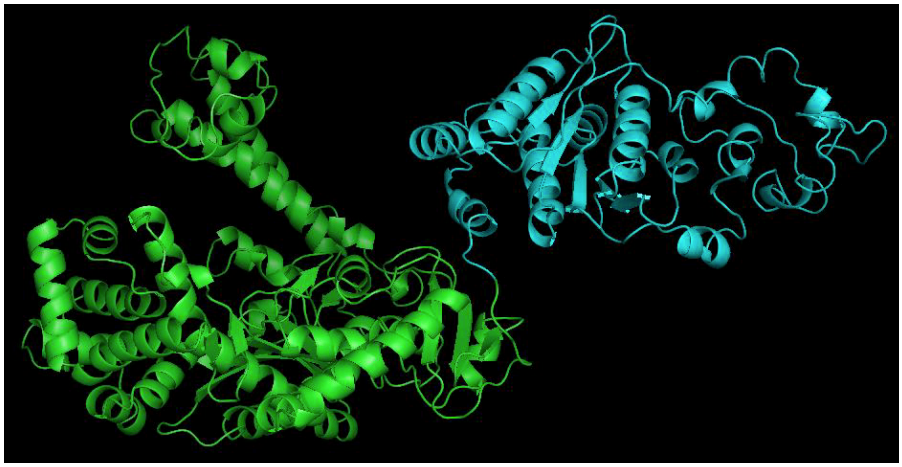


Figure 2.4: Structure of *Thermus aquaticus* DNA polymerase I/Taq Pol reproduced from the Protein Data Base[32] showing the Klenow fragment (green) with the distinct shape of 'hand-fingers and thumb', and the Flap endonuclease or the 5' endonuclease (blue).

In humans, during the lagging strand DNA replication, 5' flaps are generated 5 million times per cell cycle, and failure to eliminate them affects cell viability, causes genome instability, mutations, and predisposition to cancer in humans and mice. In humans, FEN1 is also required for non-homologous end joining of dsDNA breaks. These proteins act as enzymes to cleave the 5' single stranded RNA or DNA segment of the DNA double helix[30,44]. During DNA replication, 5' flaps are made from the RNA primer extensions formed from replication of the

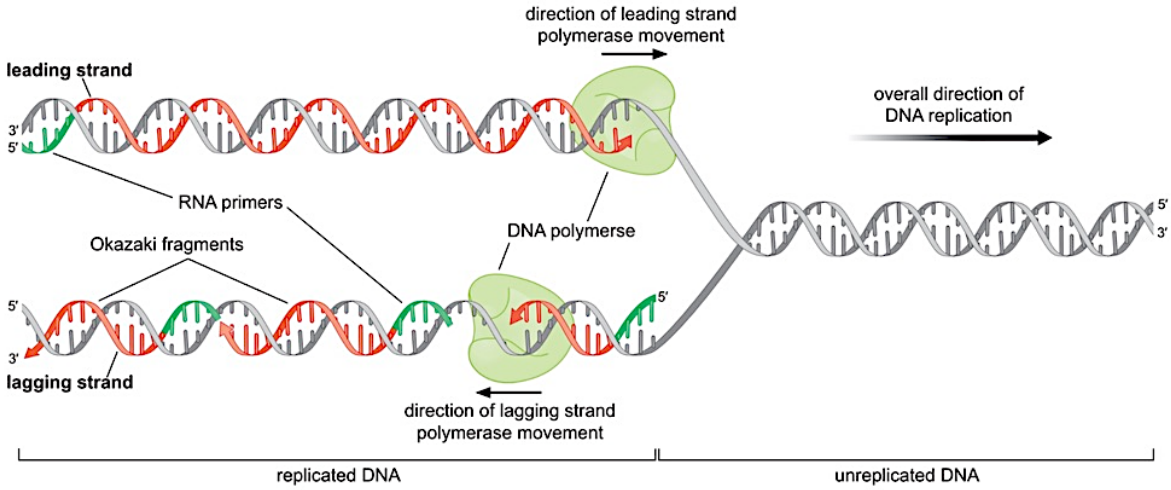


Figure 2.5: Diagrammatic representation of the DNA replication fork (reproduced from the chapter: *The Replication of DNA* [27]) **:** old DNA or template DNA is shown in grey, new DNA in red, DNA primers in green. The leading strand is replicated continuously while the lagging strand is replicated by formation of Okazaki fragments. Okazaki fragments are short DNA sections (~ 100 bp) that are replicated from 5' to 3' direction and ligated later. During extension of the Okazaki fragments, the RNA primers of the previous Okazaki fragment are displaced and lead to the formation of flap DNA.

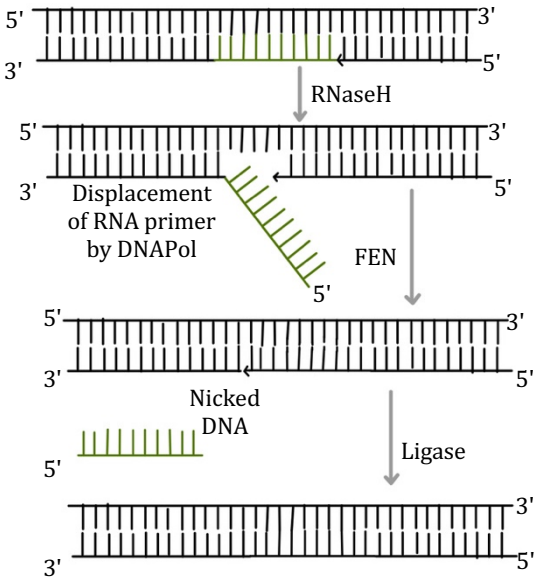


Figure 2.6: Displacement of RNA primer to produce flap DNA, reproduced from the chapter: *The Replication of DNA* [27] and *Lindahl T.(1995)* [45]. The RNA primer (green) is cleaved from the DNA by RNaseH allowing the DNA Polymerise to displace the primer and consequently form the flap. FEN activity cleaves the flap and leaves the single phosphodiester bond break (nick) that is sealed by ligases.

Okazaki fragment and are displaced during replication (figure 2.5[27] and figure 2.6[27,45]). These must be removed by FEN [46–53] to regain the smooth double-helix DNA without any branches or they would interfere with the interaction of RNA polymerase during gene expression.

2.2.2. Structure of the Active Site

To understand the mechanism of enzymatic cleavage of flap DNA, it is imperative to examine the structure of the enzyme's active site.

The core region of a lot of FEN has ß-sheets and α-helices which form the base of the pocket for divalent cations required for catalysis. Conserved amino acid sequences and motifs have been found in the active sites of FEN from different sources, with a hole in the middle of the protein (observed in FENs from T_5 exonuclease, where the hole is formed by 2 helices arched around the active site, one of them is positively charged and the other consisting of hydrophobic residues. The base of the arch contains the active site and consists of two metal binding sites[54]) or disordered around the active site (like in T_4 RNase H [55] and TaqPol [32]). The presence of a hole in T_5 FEN, and in other FENs similar in structure to human FENs indicate that the flap DNA might be threading through the enzyme during its activity [44]. Such a mechanism is believed to allow a close proximity of the ssDNA with the cation cofactors during catalysis [56].

In the active site of FEN, conserved amino acid residues are found that primarily chelate the divalent cations. In T_5 bacteriophage FEN, the base of the archway contains a tyrosine residue that is believed to play a role in stabilizing the ternary complex of protein-DNA-metal ions during catalysis [57] (see figure 2.7 and 2.12). Experiments have suggested that lysine and arginine are highly conserved residues in the active site, and interact with DNA close to the phosphodiester bond immediately 5' to the target bond, while another arginine residue interacts with the phosphodiester bond immediately 3' to the target bond. These residues have been found in *E. coli* and T_5 to form the two ends of the base of the helical arch that binds the ssDNA[54,58]. Additionally, research has shown that nine conserved carboxylates present in nucleases and FEN family, play some role in the substrate binding, and signify the involvement of the divalent cations in catalysis [55,59–66].

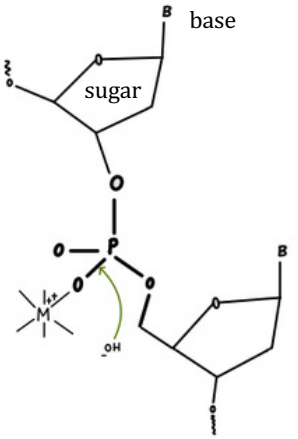


Figure 2.7 Divalent cations stabilize the phosphate ions during nucleophilic attack on the phosphodiester bond nuclease reaction, acting as Lewis Acid[42].

2.2.3. Substrate Recognition and Binding by FEN

Since the RNA primer or flap and the newly synthesized DNA strand complementary to the lagging strand of DNA is also complementary to the newly synthesized strand complementary to the leading strand (see figure 2.5), complex structures might form that would make it difficult for FEN to bind. To overcome this problem, FEN have evolved to recognise the double flap structure, which is a 5' flap of any length or sequence branching next to a 3' flap that is precisely 1 nucleotide long (figure 2.8). The FEN activity hence produces a nicked DNA[30]. Work has been done to study the different kinds of substrates [67] that can be cleaved by FEN. It has been found that FEN cleaves 5' ssDNA on flap DNA as well as on hairpin loops showing a variety of activity [38].

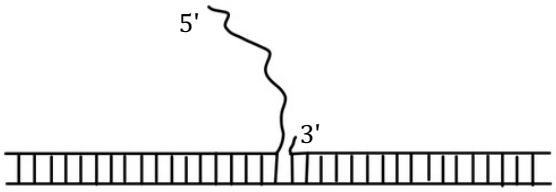


Figure 2.8: Schematic of structure of double flap: 5' flap of any length or sequence branching next to a 3' flap of length 1 nt long.

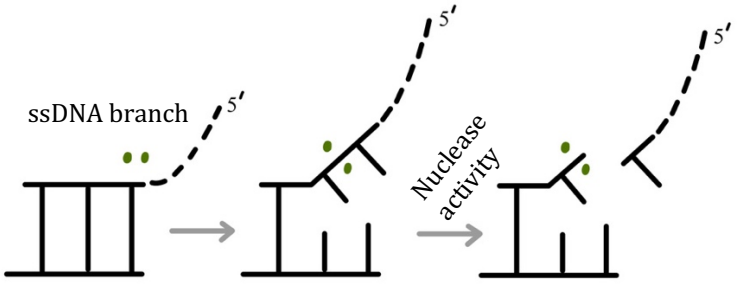


Figure 2.9: Denaturation of the single base pair at the branch point to accommodate the flap DNA threading [68]: The green points are the two divalent cations placed around the scissile phosphodiester bond in the active site.

The family of 5'-nucleases consists of structurally related proteins which have a similar active site but recognise different kinds of substrates. This has raised the doubt as to whether FEN bind the 5' end of flap or bind to the branch point by complementary strand interactions [38,46,47,69–73]. It has been proposed that substrate recognition occurs by electrostatic interactions between the phosphate backbone and FEN in a non-specific manner. The enzyme recognises the 3'-OH at the DNA end near the branch point of flap DNA (or any other branching DNA) and this amounts to the efficiency of binding [58,74]. It selects the phosphate diester bond by denaturation of the dsDNA (figure 2.9)[68] ends for a few nucleotides and binds the DNA non-specifically by detection of structures different from the double helix [30,41,46,72,75]. Hence, it has become difficult to determine how the ssDNA reaches the active site of FEN if the enzyme does not open during binding.

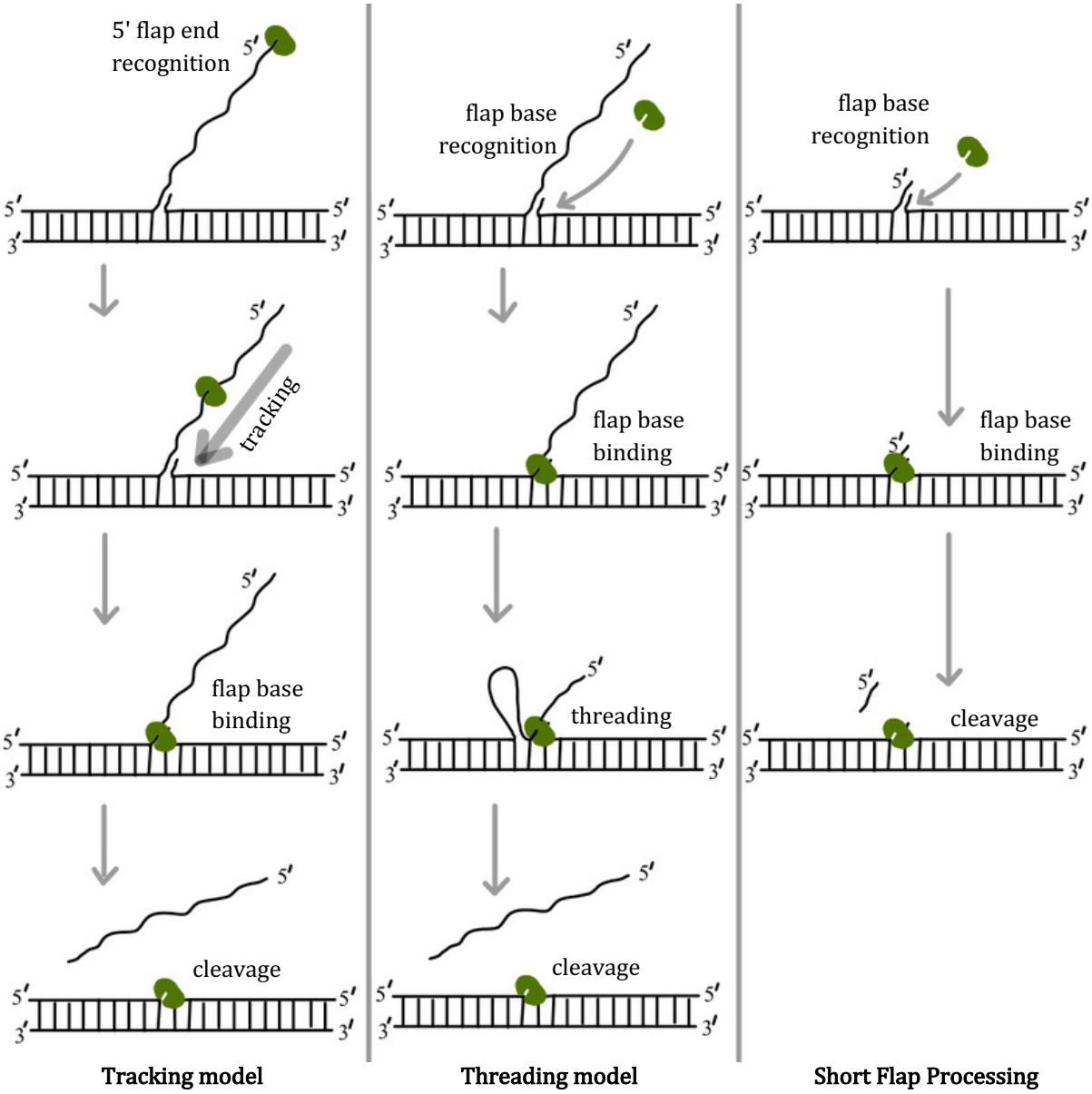


Figure 2.10: Various models for flap DNA recognition and biding by FEN[71]:
In the tracking model, FEN binds the 5' ends and tracks the ssDNA to reach and bind the branch point to cleave the branch. In the Threading model, FEN binds the branch point, followed by bending and threading of the 5' end and cleaving of the ssDNA. The third model explains how FEN binds directly to the branch point and cleaves short ssDNA flap.

Research has led to the proposal of three possible mechanisms of FEN binding to flap (figure 2.10). According to the 'Tracking model', FEN binds to the 5' ssDNA end, tracks the ssDNA to reach the branch point and binds followed by nuclease reaction. In the 'Threading model', the enzyme first binds the branch point, then bends the ssDNA to allow the 5' end to thread through the hole in the enzyme. The third model proposes the cleaving of flaps shorter than 2-

4 nucleotides by binding with no tracking or threading of ssDNA involved [71]. Out of these propositions, the 'tracking model' has been disregarded due to the low rate of recognition by the enzyme of the 5' end of long flap (assuming RNA primers to be usually 5-10 nt long), hanging far from the dsDNA in the cell.

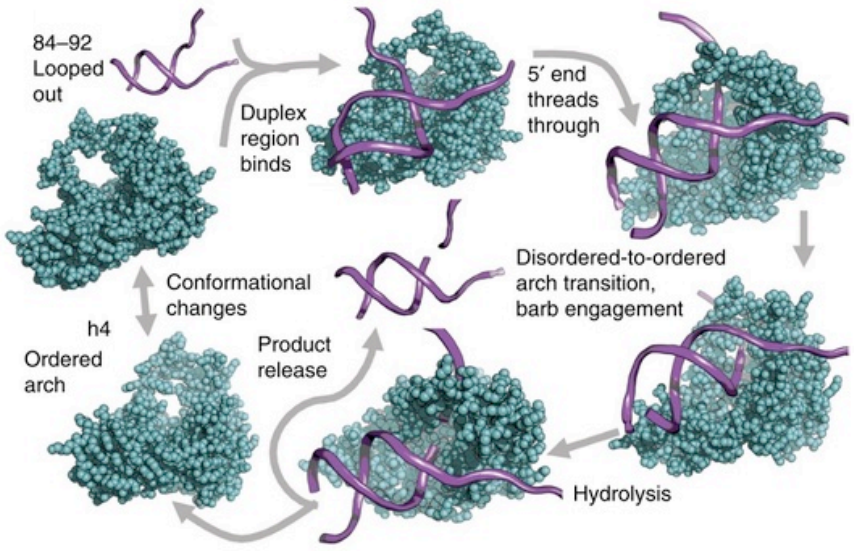


Figure 2.11: Overview of 'fly-casting, thread, bend and barb' mechanism of substrate recognition and binding by T_5 FEN[76].

Studies have shown that human FEN specifically recognises the ssDNA branch and not the sequence of DNA. The protein causes bending of the DNA junction, followed by 5' flap threading and protein recognition of the 3' OH at the branch point [46,70,77–80]. In general, FEN's structure specific recognition of double flap substrates (figure 2.8) is believed to occur using two separate DNA binding sites that sharply bend the flexible junction of flap DNA. The 3' unpaired nucleotide specifically interacts with a cleft on the protein adjacent to the upstream dsDNA binding site. FEN1 encloses a single 3' nucleotide, allowing the cleavage product to be ready for ligation and uses the threading mechanism to guide the 5' ssDNA flap into a conserved helical arch [40,46,47,54,70,81–88]. Thus, the mechanism termed as 'fly-casting, thread, bend and barb' has been proposed (figure 2.11). The amino acid residues at the hole of the active site of FEN (section 2.2.2) cause the DNA substrate to conform into a 'barb' shape thus allowing the divalent cations in the active site to interact with the phosphate group of the diester bond to be broken[76]. As already discussed, FEN possibly denatures the dsDNA segment (attached to the ssDNA flap) (figure 2.9)[68] by a few nucleotides such that the flap's target phosphodiester bond can be placed in the active site of the enzyme by DNA threading.

Hydrophobic residues of the protein form a 'hydrophobic wedge' that inhibit the base stacking of the next base pair before the flap/overhang and stabilize the bases[74], aiding in the threading. Basic residues of the enzyme have been found to steer the inverted flap into a gateway above the active site and move the dsDNA for catalysis[78]. Interestingly, the DNA binding produces changes in the active site of the enzyme which further aid in catalysis. Thus, the flexibility of the enzyme plays a role in the enzyme activity[89,90]. It has been speculated that varied mechanisms for nuclease activity may be employed by FEN during different conditions in the cell for catalysis[91].

When experiments were performed to block the 5' site with streptavidin and then allowed the FEN to act on the flap DNA, it was found that the enzyme could not bind, indicating that FEN recognised and bound to the 5' flap. This helped to discard the theory that FEN bound to the branch point of flap or overhang DNA by clamping. From the crystal structures of the enzyme, a 'disorder-thread-order' model has been proposed, indicating that the flap does not thread directly but initially passes through the arch, modifying the arch instead of denaturing the dsDNA section. After the 5' flap passes through the disordered loop, in order for the metal ions to interact with the scissile phosphodiester bond, the helix positions itself correctly with the lysine and arginine residues in the active site. This model explains how FEN can interact with even those 5' flap that have modifications. The binding of the protein to the complementary DNA strand junction allows the ssDNA to orient itself into the protein active site and removes the need of a force causing the ssDNA to thread [92].

It has also been found that human FEN has an additional 3' flap binding pocket that increases the turnover rate of catalysis when the DNA substrate has an additional 3' flap besides the 5' flap, though the catalysis rate is limited by the product release step irrespective of the presence or absence of the 3' flap. The results reported showed that the 3' pocket is required for substrate binding and recognition[70].

2.2.4. Role of Divalent Cations

Divalent cations are essential for the activity of FEN (primarily Mg^{2+}, Mn^{2+} and Co^{2+} , but also Zn^{2+}, Fe^{2+} and Ni^{2+}, with a varying degree of effect on the enzyme activity [56]), like all phosphodiesterase enzymes. Two active site divalent cations are present that interact with the phosphate diester and act as Lewis acid catalysts [42,93] (see figure 2.12). Their crystal structure shows that they have two conserved metal binding sites.

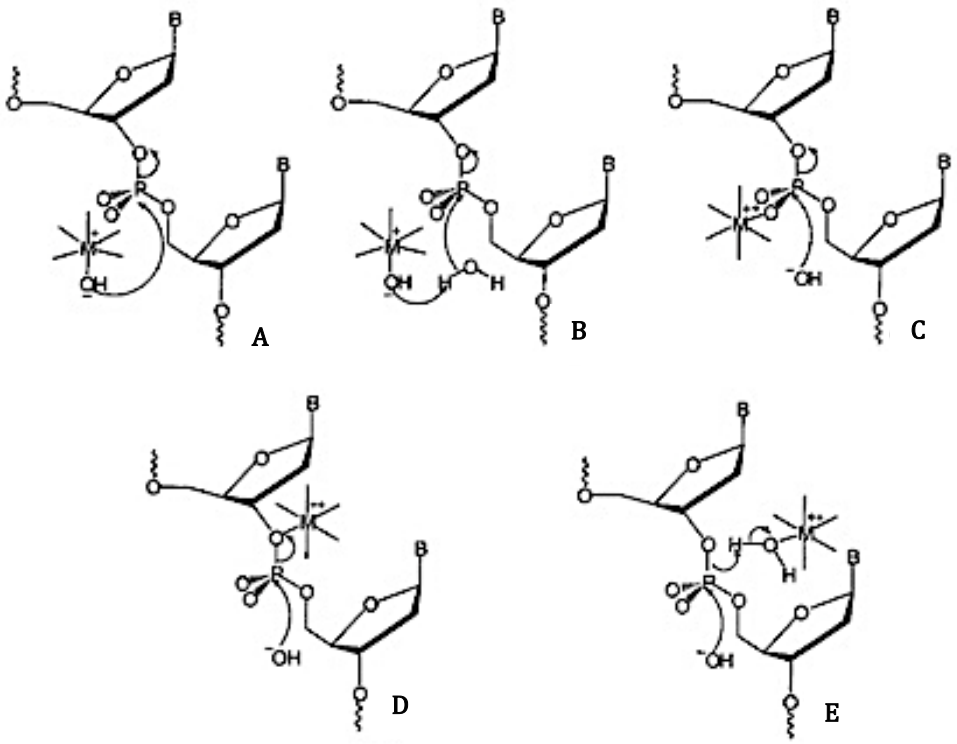


Figure 2.12: Role of divalent cations in phosphodiester bond hydrolysis: A. metal bound hydroxide acts as the nucleophile. B. nucleophilic attack on the phosphate group. C. metal ion acts as Lewis acid catalyst and binds to the non-bridging oxygen. D. leaving group is activated by the metal ion. E. metal bound water acts as a general acid. Reproduced from Tock *et al.* (2003)[42].

In T_5 FEN, however, the spacing between the divalent cations in the enzyme's structure indicates that more than 2 divalent cations might be required for enzyme activity [41]. This is because the gap between the cations site has been found to be 8-10 Å[91], which is more than the required 4 Å to bind to the same oxygen atom of the target phosphodiester bond. Crystallography studies have shown that actually 3 cations are involved in the enzyme activity, the third one believed to bind FEN when it interacts with the DNA substrate, though its function is still unknown [75].

Additionally, the rate limiting step in the reaction has been found to be the Mg^{2+} dependent step of conformational changes of the DNA in the enzyme's active site, while FEN binding and DNA bending are fast steps of catalysis (which is limited by the ease by which the enzyme encounters the substrate by diffusion) [94].

Feng *et al*.[34] have reported that Mg^{2+}, Mn^{2+}, Ni^{2+} and Zn^{2+} favour the nuclease activity of FEN more successfully than other divalent cations. Ca^{2+} has been found to not support the nuclease activity but allow the FEN to bind tightly to the substrate [41]. This binding to the enzyme has shown to inhibit catalysis, probably by inhibiting the binding of or replacing the other essential cofactor ions [75]. Hence, Ca^{2+} has emerged as a successful candidate to modify mica for DNA immobilisation to study DNA-protein binding using AFM [95].

2.2.5. Applicability of AFM in Imaging Flap DNA-FEN Interactions:

To understand flap DNA-FEN interactions, single molecule recognition reactions can be carried out on modified mica surface and observed via time lapse AFM. The technique can be applied to image in buffer conditions without damaging the sample molecules or interfering with catalysis. A low concentration of the enzyme and divalent cations in the imaging buffer can be used to control the rate of the processes to be observed under AFM [96]. In the current research work, different immobilization methods as well as protein and DNA modification methods have been applied to observe the DNA molecules interacting with FEN on modified mica surface, by formulating AFM protocol optimized for dynamic imaging. The experiments are especially complicated in case of immobilizing flap DNA sample which consists of ssDNA and dsDNA present together in one molecule. The AFM images have been analyzed and quantified by DNA molecule tracking and length measurements to inspect the modifications in the conformation of FEN and flap DNA due to their interactions. The technique can prove to be useful to understand the specificity of DNA structure for FEN binding, how the protein locates the DNA binding site in the long length of DNA and what conformation changes occur during catalysis.

2.3. ParB-parS Interactions

The second DNA-protein interaction that has been studied using AFM in this project is the binding of ParB protein with *parS* site containing DNA. ParB binds DNA as an imperative step in the process of plasmid partitioning (separation of the duplicated bacterial genome into the daughter cells). Various molecular interactions, conformation changes and enzymatic activity occurs during ParB-*parS* DNA recognition and binding and despite a lot of research and years of experimentation, there are many gaps in the understanding of the details of the process.

This section aims to summarize the structures of the *parS* site and ParB protein, and the theories and models of the processes involved in plasmid partitioning.

2.3.1. The Structure of parS Site

During cell division in bacteria, the plasmids in the cell are replicated and segregated into the daughter cells. Since the prokaryotic cells do not have a chromosome segregation machinery as found in the eukaryotic cells, there are many voids in our understanding of the separation of the duplicated genome into the progeny. To overcome this, the P1 plasmid in bacteria has been used as a standard for studying plasmid partitioning in bacteria during cell division.

AATTTCAAGGTGAAATCGCCACGATTTCACCTTG

Figure 2.13: The conserved *parS* site sequence: the 34 bp sequence with 13 bp inverted repeat (green). The sequence of 3 nucleotides separates the palindrome [97].

The P1 plasmid was defined in 1987 by Martin *et al*. The plasmid has the advantage of not being lost over consecutive divisions like the other plasmids (<1 plasmid loss in 10^5 cell divisions [98,99]) even though its copy number is 2. This indicates that there is a mechanism of efficient transfer of the plasmid through the generations. P1 has been found to contain 2.5 kb sequence (called the partition region) that participates in this separation and is located adjacent to the plasmid's origin of replication. This sequence is considered equivalent to the centromere of the eukaryotic chromosome. A 13 bp inverted repeat with a 3 bp spacer sequence has been found as the conserved sequence of 29 bp in the 34 bp sequence, the latter called the par sequence *parS* (figure 2.13). This 34 bp sequence is considered to be the cis factor for plasmid segregation[97].

The *parS* sequence can stabilize the P1 plasmid (or the F plasmid, if present in it) if the plasmid encoded proteins necessary for partition are present. It was hence proposed that the presence of this sequence in any plasmid would ensure its stable inheritance as long as the DNA replication produces two copies of the plasmid. When first discovered, the palindromic nature of the *parS* sequence also indicated that the protein binding to it was possibly binding as dimers. This has proved to be true in the subsequent research. Mutations in either of the inverted sequences have inhibited the efficient partition signifying the importance of the conservation of this sequence [97]. The inverted repeat sequence being conserved, the number, sequence and length varies in species, providing a diversity of sites for the ParB protein (explained in the following section) to bind.

2.3.2. Structure of ParB Protein

Two proteins required for partition have been recognised: ParA and ParB. ParB and ParB-like proteins are dimeric. They have a bipartite structure with the N-terminal being disordered and the C-terminal consisting of ribbon-helix-helix (RHH) fold. The N-terminal domain has been found to be the DNA-binding domain (DBD) while the C-terminal is the dimerization domain. The homologues of ParB are flexible multidomain proteins[100]. From the study of ParG protein[101], a ParB homologue in *E.coli*, it has been deduced that the protein binds the *parS* site by forming a dimer of dimer to assemble into a double-stranded ß sheet that is inserted into the major groove of DNA. C-terminal domain (of KorB protein, another homologue of ParB), on the other hand, is made up of eight helices. Two of these helices compose the helix-turn-helix motif that binds DNA, with additional residues outside this motif aiding in sequence-specific recognition of DNA during binding[100]. Recently, it has been reported from studies in *Bacillus subtilis* that the C-terminal domain consists of a dimer with lysine on the surface that allows non-specific DNA binding hence enabling DNA bridging and condensation[102].

Though ParB varies in the bacterial species, they have a conserved general structure. The N-terminal is a flexible ParB/Srx like domain with probable role in ParA binding but functions largely unknown, the middle DNA-binding domain with HTH motif that specifically interacts with *parS* site and the C-terminal responsible for dimerization of ParB and spreading by sequence non-specific binding to *parS* DNA[100,103,104] (figure 2.14). The N-terminal is called Srx-

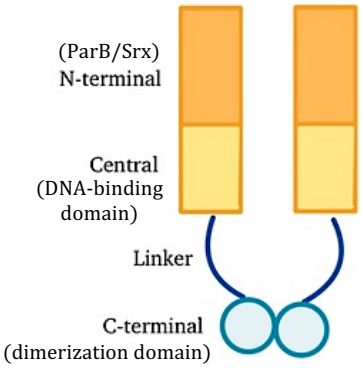


Figure 2.14: Diagrammatic representation of ParB dimer.

like owing to the resemblance of this domain with the nucleotide binding sites of ATP/ADP dependent enzymes sulfiredoxin[105] and SerK[106], hence the N-terminal may be involved in the enzymatic role of ATPase activity. Jalal *et al.* have proposed two functions of ParB, DNA bridging and parB spreading on the DNA, the latter requiring CTP [107], and recent work [104] has shown that the N-terminal of ParB might be the domain responsible for this CTP binding and switch-like property (the details of which are discussed in section 2.3.3).

2.3.3. Role of CTP

Recently, the involvement of CTP in the binding of ParB to DNA has been discovered. To understand this, experiments have been performed to analyse the change in the behaviour of ParB in accordance with varying concentration of CTP in the buffer. Additionally, the change in the ParB binding specifically to *parS* vs to DNA in a sequence non-specific manner has also been found to be affected by the presence of CTP.

It has been reported that increasing the concentration of CTP decreased the specific binding of ParB to *parS* and that bound ParB dissociated about 7 times faster from *parS* in the presence of CTP. This result was observed for short DNA of 20 bp. In longer DNA of 169 bp with the ends blocked, CTP enhanced the binding of ParB to the entire DNA segments, not just the *parS* site. Additionally, CTP increased the *parS* dependent spreading of ParB. However, similar results were not obtained when the DNA ends were free, *i.e.* ParB aggregation on the DNA was not enhanced by the presence of CTP when the DNA was open-ended. Interestingly, if the DNA ends were blocked by the binding of other ParB, the results obtained were similar to those for closed DNA and the presence of CTP enhanced the ParB spreading and non-specific binding[107].

Hence, it can be understood that ParB binds to DNA specifically to the *parS* site but the presence of CTP causes the interaction to convert to non-specific binding. The protein dissociates from the DNA, probably by sliding off the DNA ends, in the absence of something blocking the DNA ends or DNA being closed. Capping the DNA with any molecule causes the DNA to behave in a manner similar to closed ends for ParB interaction [107]. Moreover, CTP binding by ParB has been found to decrease in the absence of *parS* sites on the DNA[108,109]. This has shown that though CTP encourages ParB to switch to non-specific binding, the initial recognition of the *parS* site is imperative for the protein and its associated CTP molecule to bind DNA.

Additionally, experiments with *Bacillus*[103] have shown that ParB-CTP forms a sliding clamp that slides on the DNA. This has explained why the DNA bound to ParB-CTP complex does not dissociate quickly and has indicated that CTP stabilises the binding of ParB to DNA[107]. CTP hydrolysis has been reported to be involved in controlling the time for which ParB binds the DNA and the length to which it slides. But CTP has been found to not be important for DNA loading or sliding, though it plays an important role in allowing the ParB to concentrate near the plasmid in a sequence non-specific manner[102].

Work has been done to understand the mechanism of binding and sliding of ParB on DNA (Figure 2.15). Biochemical experiments have shown that ParB exists in the CTP bound state until it encounters the *parS* site, which causes the adjacent ParB/Srx domain to dimerize, stimulating CTP hydrolysis. The CDP thus formed has weaker interactions with ParB and dissociates from the ParB/Srx domain. This CTP switch ensures that the *parS* sites are not crowded by the protein[104]. ParB also interacts with the DNA non-specifically[102] making the *parS* site available for the next ParB molecule and allowing the single *parS* site to attract multiple ParB molecules[110]. The *parS* site is believed to possibly acts as a CTPase activating factor that causes conformation changes in ParB[104]. Though a lot of research has undergone to understand the mechanism of ParB spreading, the process has not been reproduced *in vitro*, signifying that there is a lack of information on the requirements of the process.

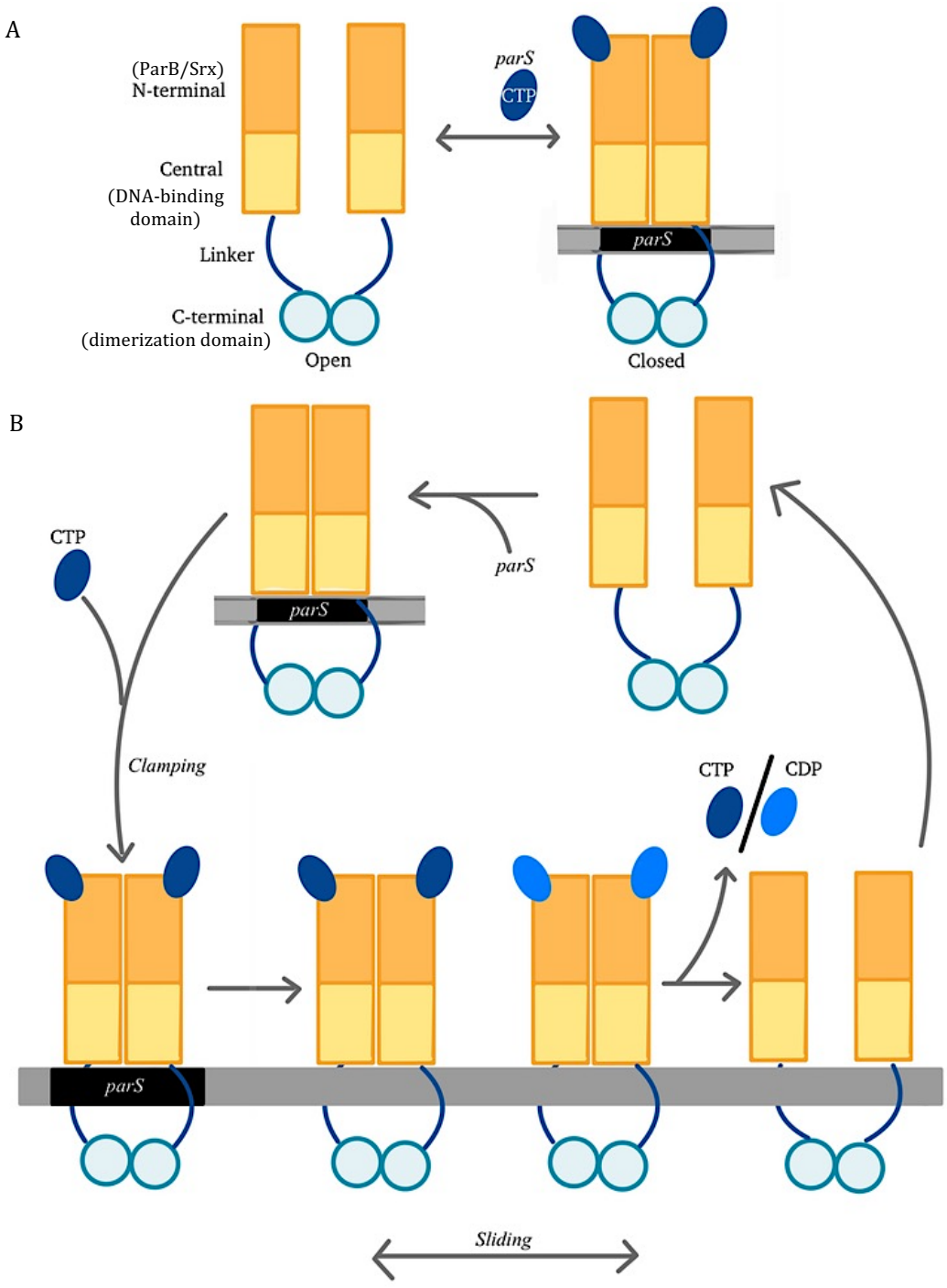


Figure 2.15: Schematic of ParB binding and sliding [104,107,111] **:**

A. The open and closed states of the ParB dimer. The open conformation converts to closed form on binding CTP and *parS* site and vice versa.

B. The open state of ParB can bind *parS* site and convert to closed state that can bind CTP. Consequently ParB undergoes a conformation change that causes the protein to clamp around *parS*, leading to unbinding from *parS* and be able to slide on the DNA . This frees the *parS* site to bind a new parB molecule. Due to the CTPase activity trigger by *parS* site, the CTP is converted to CDP and dissociates from ParB, converting the protein into open state and release from the DNA.

2.3.4. *Mechanism of Partition*

Duplicated plasmids in the prokaryotic cell separate during cell division into the daughter cells, the mechanism of which involves ParA and ParB protein homologues and corresponding nucleoside triphosphates (NTPs). The details of the mechanism of partition of plasmids are not in the scope of this thesis, nevertheless, it is important to briefly outline the major events of the process. Two theories have primarily been put forward to explain this process: the filament pulling model, as described below in figure 2.16 [112,113], and the diffusion ratchet model where the concentration gradient of ParA dimers drives the motion of plasmids [113–117].

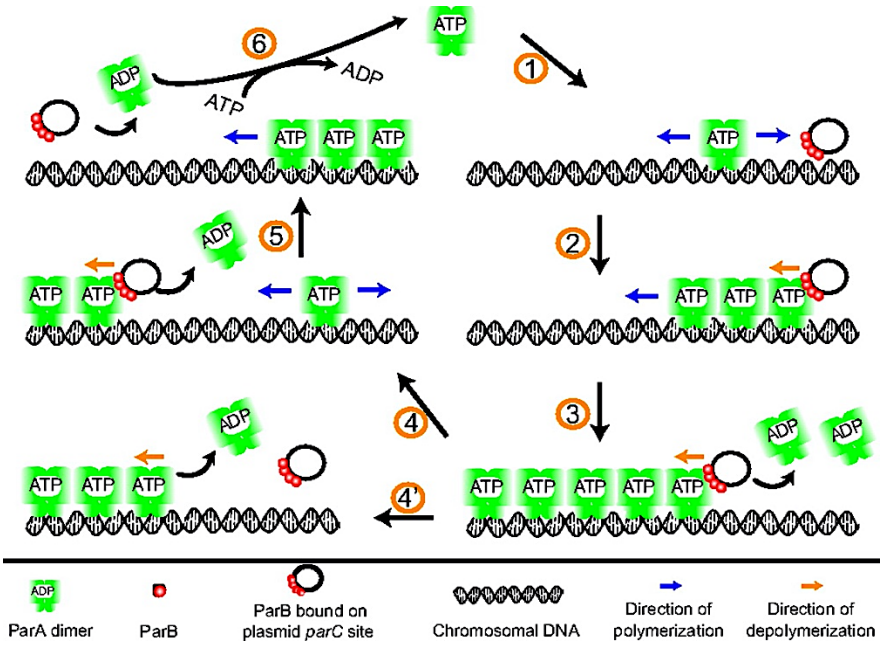


Figure 2.16: Filament Pulling model: Sequence of events in the partition of plasmids: 1. ATP bound ParA dimerizes and assembles into filaments. 2. Plasmid bound ParB recognises ParA and binds. 3. ParB triggers the ATPase reaction and the loss of ATP causes ParA to depolymerize and disengage. 4. ParB bound to plasmid is transferred to the next ParA-ATP dimer. 5. The movement of plasmid continues until there are no ParA-ATP dimers left to pass onto [112].

According to the filament pulling model, ParB binds the *parS* site on the newly replicated plasmids and subsequently leads to multiple ParB molecules to spread around the *parS* sequence to form the partition complex, also called as segrosome [100].This step has already been described in the section 2.3.3. Simultaneously, ParA binds ATP, dimerizes and assembles in the nucleoid (region of the prokaryotic cell where the DNA localizes) into large filaments[112]. The replicated plasmids are paired, mediated by the ParB bound to them, and the partition complex is recognised by ParA[100]. ParB triggers the ATPase activity of ParA which in turn

stimulates the depolymerization of ParA[100]. As ParA depolymerises, it moves away from the filament assembly and the partition complex (consisting of the plasmid bound ParB) is attached to the next ParA-ATP dimer [112]. This, in a way, causes the ParB bound plasmid to be passed from ParA to ParA, thus moving the plasmids in the cell.

According to the more widely accepted diffusion ratchet model[113–117], the partition is based on the concentration gradient of ParA-ATP dimers. ParA-ATP dimers bind DNA to the nucleoid surface by making use of the ParB-*parS* DNA interaction. Formation of this complex triggers the ATPase activity of ParA causing it to disengage from the complex but unable to bind again due to the reactivation of ParA being a slow conformation change. This lag-time causes a concentration gradient that in-turn causes the plasmid to detach from the nucleoid and diffuse to the edge of the gradient to reattach at the point of high ParA-ATP concentration. Thus the movement of plasmid occurs through ParA-ATP diffusion zones [114].

2.3.5. Future Prospects in Understanding of parS-ParB Interactions

There are many aspects of the *parS*-ParB interaction, and the partition process, that are yet to be clear. Firstly, the structural properties of ParB that allow it to bind the DNA sequence are not fully understood. Additionally, the steps of assembly of the segrosome (*parS* DNA bound to ParB protein) are not very clear[100]. The mechanism by which ParB, bound to the *parS* site, recruits more protein molecules to assemble near the *parS* site and form a large nucleoprotein complex is not known[107]. Also, the conformation changes on ParB that cause the sliding of the protein on DNA need to be visualized to understand how the molecule opens and closes. Attempts likewise have to be made to understand how does CTP hydrolysis occur and is related to ParB sliding[113]. Lastly, research has to be undertaken to comprehend the variation in the mechanisms of ParB-*parS* interaction and the involvement of CTP, ParA and ATP in different species of bacteria[113].

To answer the above questions, high resolution AFM could be used to image the protein to observe its conformation. Each of the steps of segrosome assembly, parB binding and sliding on the DNA and plasmid partition could be performed on the modified mica surface to allow dynamic imaging of DNA-protein and protein-protein interactions. With further improvements in the time and spatial resolution of AFM, the details of this complex nucleoprotein assembly could be possibly studied.

Bibliography

1. Watson D, Acid N. Molecular Structure of Nucleic Acids. Nature. 1953;171.
2. Crick F, Watson J. Molecular Structure of Deoxypentose Nucleic Acids. Nature. 1953;171:738-740.
3. Berg JM, Tymoczko JL, Stryer L. DNA, RNA, and the Flow of Genetic Information. In: Biochemistry. 7th ed. W. H. Freeman and Company, New York; 2012:109-138.
4. James D. Watson *et al*. The Structure of DNA. In: Molecular Biology of the Gene. 7th ed. Pearson; 2014:77-105.
5. Murugesapillai D, Bouaziz S, Maher LJ, Israeloff NE, Cameron CE, Williams MC. Accurate nanoscale flexibility measurement of DNA and DNA–protein complexes by atomic force microscopy in liquid. Nanoscale. 2017;9(31):11327-11337.
6. Hansma HG *et al*. Properties of Biomolecules Measured from Atomic Force Microscope Images: A Review. J Struct Biol. 1997;119(2):99-108.
7. Hansma HG, Revenko I, Kim K, Laney DE. Atomic force microscopy of long and short double-stranded, single-stranded and triple-stranded nucleic acids. Nucleic Acids Res. 1996;24(4):713-720.
8. Podestà A *et al*. Positively charged surfaces increase the flexibility of DNA. Biophys J. 2005;89:2558–2563.
9. Lo K, Simonov A. Strong deformations of DNA: Effect on the persistence length. Eur Phys J E. 2018;41:114.
10. Garai A, Ghoshdastidar D, Senapati S, Maiti PK. Ionic liquids make DNA rigid. J Chem Phys. 2018;149:45104.
11. Lysetska M, Knoll A, Boehringer D, Hey T, Krauss G, Krausch G. UV light-damaged DNA and its interaction with human replication protein A: an atomic force microscopy study. Nucleic Acids Res. 2002;30(12):2686-2691.
12. Cristofalo M *et al*. Nanomechanics of Diaminopurine-Substituted DNA. Biophys J. 2019;116:760–771.
13. Chakraborty D, Hori N, Thirumalai D. Sequence-Dependent Three Interaction Site Model for Single-and Double-Stranded DNA. J Chem Theory Comput. 2018;14:3763−3779.
14. Matsumoto A, Olson WK. Sequence-Dependent Motions of DNA: A Normal Mode Analysis at the Base-Pair Level. Biophys J. 2002;83(1):22-41.
15. Stellwagen E, Stellwagen NC. Determining the electrophoretic mobility and translational diffusion coefficients of DNA molecules in free solution. Electrophoresis. 2002;23:2794-2803.
16. Herrero-Gala E *et al*. Mechanical Identities of RNA and DNA Double Helices Unveiled at the Single-Molecule Level. J Am Chem Soc. 2013;135:122-131.
17. Lamour G, Kirkegaard JB, Li H, Knowles TP, Gsponer J. Easyworm: an open-source software tool to determine the mechanical properties of worm-like chains. Source Code Biol Med. 2014;9.
18. Guilbaud S, Salomé L, Destainville N, Manghi M, Tardin C. Dependence of DNA Persistence Length on Ionic Strength and Ion Type. Phys Rev Lett. 2019;122(2).
19. Tinland B, Pluen A, Sturm J, Weill G. Persistence Length of Single-Stranded DNA. Macromolecules. 1997;30:5763-5765.
20. Roth E, Glick Azaria A, Girshevitz O, Bitler A, Garini Y. Measuring the Conformation and Persistence Length of Single-Stranded DNA Using a DNA Origami Structure. Nano Lett. 2018;18:6703-6709.
21. Chi Q, Wang G, Jiang J. The persistence length and length per base of single-stranded DNA obtained from fluorescence correlation spectroscopy measurements using mean field theory. Physica A. 2013;392(5):1072-1079.
22. Vemulapalli S, Hashemi M, Kolomeisky AB, Lyubchenko YL. DNA Looping Mediated by Site-Specific SfiI−DNA Interactions. J Phys Chem B. 2021;125:4645-4653.
23. Maaloum M, Muller P, Harlepp S. DNA-intercalator interactions: structural and physical analysis using atomic force microscopy in solution. Soft Matter. 2013;9:11233–11240.
24. Rivetti C, Guthold M, Bustamante C. Scanning force microscopy of DNA deposited onto mica: Equilibration versus Kinetic trapping studied by statistical polymer chain analysis. J Mol Biol. 1996;264(5):919-932.
25. Lam PM, Zhen Y. Twisting, supercoiling and stretching in protein bound DNA. Phys A Stat Mech its

Appl. 2018;496:200-208.

26. Alberts B *et al*. Chapter 3: Proteins. In: Molecular Biology of the Cell. sixth.
27. Watson JD *et al*. Molecular Biology of the Gene. 7th ed. Cold Spring Harbour Laboratory Press; 2013.
28. Liu Y, Kao H-I, Bambara RA. Flap Endonuclease 1: A Central Component of DNA Metabolism. Annu Rev Biochem. 2004;73(1):589-615. doi:10.1146/annurev.biochem.73.012803.092453
29. Balakrishnan L, Bambara RA. Flap Endonuclease 1. Annu Rev Biochem. 2013;82(1):119-138.
30. Beddows A, Patel N, Finger D, Atack JM, Williams DM, Grasby JA. Interstrand disulfide crosslinking of DNA bases supports a double nucleotide unpairing mechanism for flap endonucleases. Chem Commun. 2012;48:8895–8897.
31. Allen LM, Hodskinson MRG, Sayers JR. Active site substitutions delineate distinct classes of eubacterial flap endonuclease. Biochem J. 2009;418:285–292.
32. Kim Y, Hyun Eom S, Wang J, Lee D-S, Won Suht Be Thomas Steitz SA. Crystal structure of Thermus aquaticus DNA polymerase. Nature. 1995;376:612-616.
33. Sayers JR. Computer aided identification of a potential 5'-3' exonuclease gene encoded by escherichia coli. J Theor Biol. 1994;170:415-421.
34. Feng M *et al*. Roles of divalent metal ions in flap endonuclease-substrate interactions. Nat Struct Mol Biol. 2004;11:450-456.
35. Lyamichev V, Brown MA, Dahlberg JE. Structure-specific endonucleolytic cleavage of nucleic acids by eubacterial DNA polymerases. Science (80-). 1993;260:778-783.
36. Lieber MR. The FEN-1 family of structure-specific nucleases in eukaryotic DNA replication, recombination and repair. 1969.
37. Zheng L, Jia J, David Finger L, Guo Z, Zer C, Shen B. Functional regulation of FEN1 nuclease and its link to cancer. Nucleic Acids Res. 2011;39:781-794.
38. Williams R *et al*. Comparison of the Catalytic Parameters and Reaction Specificities of a Phage and an Archaeal Flap Endonuclease. J Mol Biol. 2007;371(1):34-48.
39. Harrington JJ, Lieber MR. DNA structural elements required for FEN-1 binding. J Biol Chem. 1995;270(9):4503-4508.
40. Craggs TD, Hutton RD, Brenlla A, White MF, Penedo JC. Single-molecule characterization of Fen1 and Fen1/PCNA complexes acting on flap substrates. Nucleic Acids Res. 2014;42:1857-1872.
41. Tomlinson G *et al*. Neutralizing Mutations of Carboxylates That Bind Metal 2 in T5 Flap Endonuclease Result in an Enzyme That Still Requires Two Metal Ions. J Biol Chem. 2011;286:30878-30887.
42. Tock MR, Frary E, Sayers JR, Grasby JA. Dynamic evidence for metal ion catalysis in the reaction mediated by a flap endonuclease. EMBO J. 2003:995-1004.
43. Kao H-II, Bambara RA. The Protein Components and Mechanism of Eukaryotic Okazaki Fragment Maturation. Crit Rev Biochem Mol Biol. 2003;38(5):433-452.
44. Sayers JR, Artymiuk PJ. Flexible loops and helical arches. Nat Struct Mol Biol. 1998;5(8):668-670.
45. Lindahl T. Recognition and processing of damaged DNA. J Cell Sci Suppl. 1995;19:73-77.
46. Tsutakawa SEE *et al*. Human Flap Endonuclease Structures, DNA Double-Base Flipping, and a Unified Understanding of the FEN1 Superfamily. Cell. 2011;145(2):198-211.
47. Murante RS, Rust L, Bambara RA. Calf 5' to 3' Exo/Endonuclease Must Slide from a 5' End of the Substrate to Perform Structure-specific Cleavage. J Biol Chem. 1995;270:30377-30383.
48. Turchi JJ, Bambara RA. Completion of mammalian lagging strand DNA replication using purified proteins. J Biol Chem. 1993;268(20):15136-15141.
49. Siegal G *et al*. Structural relationships between two forms of DNA polymerase epsilon from calf thymus. J Biol Chem. 1992;267(6):3991-3999.
50. Siegal G, Turchi JJ, Myers TW, Bambara RA. A 5' to 3' exonuclease functionally interacts with calf DNA polymerase epsilon. Proc Natl Acad Sci. 1992;89(20):9377-9381.
51. Huang L, Rumbaugh JA, Murante RS, Lin RJR, Rust L, Bambara RA. Role of Calf RTH-1 Nuclease in Removal of 5′-Ribonucleotides during Okazaki Fragment Processing. Biochemistry. 1996;35:9266-9277.
52. Bambara RA, Murante RS, Henricksen LA. Enzymes and Reactions at the Eukaryotic DNA Replication Fork. J Biol Chem. 1997;272:4647–4650.
53. Murante RS, Huang L, Turchi JJ, Bambara RA. The calf 5'- to 3'-exonuclease is also an endonuclease with both activities dependent on primers annealed upstream of the point of cleavage. J Biol Chem.

1994;269(2):1191-1196.

54. Ceska TA, Sayers JR, Stier G, Suck D. A helical arch allowing single-stranded DNA to thread through T5 5'-exonuclease. Nature. 1996;382(6586):90-93.
55. Mueser TC, Nossal NG, Hyde CC. Structure of bacteriophage T4 RNase H, a 5′ to 3′ RNA–DNA and DNA–DNA exonuclease with sequence similarity to the RAD2 family of eukaryotic proteins. Cell. 1996;85(7):1101-1112.
56. Garforth SJ, Patel D, Feng M, Sayers JR. Unusually wide co-factor tolerance in a metalloenzyme; divalent metal ions modulate endo-exonuclease activity in T5 exonuclease. Nucleic Acids Res. 2001;29(13):2772-2779.
57. Patel D *et al*. A conserved tyrosine residue aids ternary complex formation, but not catalysis, in phage T5 flap endonuclease. J Mol Biol. 2002;320:1025-1035.
58. Xu Y, Potapova O, Leschziner AE, Grindley NDF, Joyce CM. Contacts between the 5' Nuclease of DNA Polymerase I and Its DNA Substrate. J Biol Chem. 2001;276:30167-30177.
59. Bhagwat M, Meara D, Nossal NG. Identification of Residues of T4 RNase H Required for Catalysis and DNA Binding. J Biol Chem. 1997;272:28531–28538.
60. Shen B, Nolan JP, Sklar LA, Park MS. Essential Amino Acids for Substrate Binding and Catalysis of Human Flap Endonuclease 1. J Biol Chem. 1996;271:9173–9176.
61. Xu Y, Derbyshire V, Ng K, Sun C, Grindley NDF, Joyce CM. Biochemical and Mutational Studies of the 5 H-3 H Exonuclease of DNA Polymerase I of Escherichia coli. J Mol Biol. 1997;268:284-302.
62. Gutman PD, Minton KW. Conserved sites in the 5'-3' exonuclease domain of Escherichia coli DNA polymerase. Nucleic Acids Res. 1993;21(18):4406-4407.
63. Shen B, Nolan JP, Sklar LA, Park MS. Functional analysis of point mutations in human flap endonuclease-1 active site. Nucleic Acids Res. 1997;25(16):3332–3338.
64. Mizrahi V, Huberts P. Deoxy-and dideoxynucleotide discrimination and identification of critical 5′ nuclease domain residues of the DNA polymerase I from Mycobacterium tuberculosis. Nucleic Acids Res. 1996;24(24):4845-4852.
65. Joyce CM, Steitz TA. Function and Structure Relationships in DNA Polymerases. Annu Rev Biochem. 1994;63:777-822.
66. Shen B, Qiu J, Hosfield DJ, Tainer JA. Flap endonuclease homologs in archaebacteria exist as independent proteins. 1998.
67. Nazarkina ZK, Lavrik OI, Khodyreva SN. Flap endonuclease 1 and its role in eukaryotic DNA metabolism. Mol Biol. 2008;42(3):357-370. doi:10.1134/S0026893308030035
68. David Finger L *et al*. Observation of unpaired substrate DNA in the flap endonuclease-1 active site. Nucleic Acids Res. 2013;41:9839–9847.
69. Wu X, Li J, Li X, Hsieh C-L, Burgers PMJ, Lieber MR. Processing of branched DNA intermediates by a complex of human FEN-1 and PCNA. Nucleic Acids Res. 1996;24(11):2036-2043.
70. Blanchard MS *et al*. The 3'-Flap Pocket of Human Flap Endonuclease 1 Is Critical for Substrate Binding and Catalysis. J Biol Chem. 2009;284:22184-22194.
71. Gloor JW, Balakrishnan L, Bambara RA. Flap Endonuclease 1 Mechanism Analysis Indicates Flap Base Binding Prior to Threading. J Biol Chem. 2010;285(45):34922-34931.
72. Orans J *et al*. Structures of Human Exonuclease 1 DNA Complexes Suggest a Unified Mechanism for Nuclease Family. Cell. 2011;145(2):212-223.
73. Hitomi K, Iwai S, Tainer JA. The intricate structural chemistry of base excision repair machinery: Implications for DNA damage recognition, removal, and repair. DNA Repair. 2007;6(4):410-428.
74. Nishino T, Ishino Y, Morikawa K. Structure-specific DNA nucleases: structural basis for 3D-scissors. Curr Opin Struct Biol. 2006;16(1):60-67.
75. Syson K *et al*. Three metal ions participate in the reaction catalyzed by T5 flap endonuclease. J Biol Chem. 2008;283:28741–28746.
76. AlMalki FA *et al*. Direct observation of DNA threading in flap endonuclease complexes. Nat Struct Mol Biol. 2016.
77. Chapados BR *et al*. Structural Basis for FEN-1 Substrate Specificity and PCNA-Mediated Activation in DNA Replication and Repair. Cell. 2004;116(1):39-50.
78. Tsutakawa SE *et al*. Phosphate steering by Flap Endonuclease 1 promotes 5'-flap specificity and incision to prevent genome instability. Nat Commun. 2017;8:15855.

79. Rashid F *et al*. Single-molecule FRET unveils induced-fit mechanism for substrate selectivity in flap endonuclease 1. Elife. 2017;6:1-23.
80. Algasaier SI *et al*. DNA and Protein Requirements for Substrate Conformational Changes Necessary for Human Flap Endonuclease-1-catalyzed Reaction. J Biol Chem. 2016;291(15):8258-8268.
81. Henricksen LA, Tom S, Liu Y, Bambara RA. Inhibition of Flap Endonuclease 1 by Flap Secondary Structure and Relevance to Repeat Sequence Expansion. J Biol Chem. 2000;272:16420–16427.
82. Devos JM, Tomanicek SJ, Jones CE, Nossal NG, Mueser TC. Crystal Structure of Bacteriophage T4 5 Nuclease in Complex with a Branched DNA Reveals How Flap Endonuclease-1 Family Nucleases Bind Their Substrates. J Biol Chem. 2007;282(43):31713-31724.
83. Bornarth CJ, Ranalli TA, Henricksen LA, Wahl AF, Bambara RA. Effect of Flap Modifications on Human FEN1 Cleavage. Biochemistry. 1999;38:13347-13354.
84. Ceska TA, Sayers JR. Structure-specific DNA cleavage by 5′ nucleases. Trends Biochem Sci. 1998;23(9):331-336.
85. Tsutakawa SE, Tainer JA. Double strand binding–single strand incision mechanism for human flap endonuclease: Implications for the superfamily. Mech Ageing Dev. 2012;133(4):195-202.
86. Shaw SJ, Finger LD, Grasby JA. Human Exonuclease 1 Threads 5′-Flap Substrates through Its Helical Arch. Biochemistry. 2017.
87. Liu R, Qiu J, David Finger L, Zheng L, Shen B. The DNA-protein interaction modes of FEN-1 with gap substrates and their implication in preventing duplication mutations.
88. Qiu J, Liu R, Chapados BR, Sherman M, Tainer JA, Shen B. Interaction Interface of Human Flap Endonuclease-1 with Its DNA Substrates. J Biol Chem. 2004;279(23):24394-24402.
89. Bennet IA *et al*. Regional conformational flexibility couples substrate specificity and scissile phosphate diester selectivity in human flap endonuclease 1. Nucleic Acids Res. 2018;46(11):5618-5633.
90. Storici F, Henneke G, Ferrari E, Gordenin DA, Hubscher U, Resnick MA. The flexible loop of human FEN1 is required for flap cleavage during DNA replication and repair. EMBO J. 2002;21:5930-5942.
91. Garforth SJ, Ceska TA, Suck D, Sayers JR. Mutagenesis of conserved lysine residues in bacteriophage T5 5'-3' exonuclease suggests separate mechanisms of endo-and exonucleolytic cleavage. Proc Natl Acad Sci U S A. 1999;96(1):38-43.
92. Patel N *et al*. Flap endonucleases pass 5'-flaps through a flexible arch using a disorder-thread-order mechanism to confer specificity for free 5'-ends. Nucleic Acids Res. 2012;40:4507–4519.
93. Yang W, Lee JY, Nowotny M. Making and Breaking Nucleic Acids: Two-Mg^{2+} Ion Catalysis and Substrate Specificity. Mol Cell. 2006;22(1):5-13.
94. Song B, Hamdan SM, Hingorani MM. Positioning the 5'-flap junction in the active site controls the rate of flap endonuclease-1–catalyzed DNA cleavage. J Biol Chem. 2018;293(13):4792-4804.
95. Sorel I *et al*. The EcoRI-DNA complex as a model for investigating protein-DNA interactions by atomic force microscopy. Biochemistry. 2006;45(49):14675-14682.
96. Chao J *et al*. Single-molecule imaging of DNA polymerase I (Klenow fragment) activity by atomic force microscopy. Nanoscale. 2016;8:5842.
97. Martin KA, Friedman SA, Austin SJ. Partition site of the P1 plasmid. Proc Natl Acad Sci U S A. 1987;84(23):8544-8547.
98. Austin S, Ziese M, Sternberg N. A novel role for site-specific recombination in maintenance of bacterial replicons. Cell. 1981;25(3):729-736.
99. Rosner JL. Formation, induction, and curing of bacteriophage P1 lysogens. Virology. 1972;48(3):679-689.
100. Schumacher MA. Structural biology of plasmid segregation proteins. Curr Opin Struct Biol. 2007;17:103–109.
101. Golovanov AP, Barillà D, Golovanova M, Hayes F, Lian LY. ParG, a protein required for active partition of bacterial plasmids, has a dimeric ribbon-helix-helix structure. Mol Microbiol. 2003;50(4):1141-1153.
102. Fisher GLM *et al*. The structural basis for dynamic DNA binding and bridging interactions which condense the bacterial centromere. Elife. 2017;6:1-25.
103. Soh Y-M *et al*. Self-organization of *parS* centromeres by the ParB CTP hydrolase. Science (80-). 2019;366:1129-1133.
104. Osorio-Valeriano M *et al*. ParB-type DNA Segregation Proteins Are CTP-Dependent Molecular Switches. Cell. 2019;179(7):1512-1524.

105. Jönsson TJ, Murray MS, Johnson LC, Lowther WT. Reduction of Cysteine Sulfinic Acid in Peroxiredoxin by Sulfiredoxin Proceeds Directly through a Sulfinic Phosphoryl Ester Intermediate. J Biol Chem. 2008;283(35):23846-23851.

106. Nagata R *et al*. Structural Study on the Reaction Mechanism of a Free Serine Kinase Involved in Cysteine Biosynthesis. ACS Chem Biol. 2017;12:1514−1523.

107. Jalal ASB, Tran NT, Le TBK. ParB spreading on DNA requires cytidine triphosphate in vitro. Elife. 2020;9:1-24.

108. Wu LJ, Errington J. Coordination of Cell Division and Chromosome Segregation by a Nucleoid Occlusion Protein in Bacillus subtilis. Cell. 2004;117(7):915-925.

109. Jalal ASB *et al*. Diversification of DNA-binding specificity via permissive and specificity-switching mutations in the ParB/Noc protein family. bioRxiv. 2019:724823.

110. Wang X, Brandão HB, Le TBK, Laub MT, Rudner DZ. Bacillus subtilis SMC complexes juxtapose chromosome arms as they travel from origin to terminus. Science (80-). 2017;527(February):524-527.

111. Walter J-C *et al*. Physical modeling of a sliding clamp mechanism for the spreading of ParB at short genomic distance from bacterial centromere sites. iScience. 2020;23(12):101861.

112. Ringgaard S, Van Zon J, Howard M, Gerdes K. Movement and equipositioning of plasmids by ParA filament disassembly. Proc Natl Acad Sci. 2009;106:19369–19374.

113. Le TBK, Jalal ASB. Bacterial chromosome segregation by the ParABS system. Open Biol. 2020;10:200097.

114. Vecchiarelli AG *et al*. ATP control of dynamic P1 ParA-DNA interactions: A key role for the nucleoid in plasmid partition. Mol Microbiol. 2010;78(1):78-91.

115. Vecchiarelli AG, Neuman C, Mizuuchi K. A propagating ATPase gradient drives transport of surface-confined cellular cargo. Proc Natl Acad Sci. 2014;111(13):4880–4885.

116. Vecchiarelli AG, Hwang LC, Mizuuchi K. Cell-free study of F plasmid partition provides evidence for cargo transport by a diffusion-ratchet mechanism. Proc Natl Acad Sci. 2013:E1390–E1397.

117. Hwang LC *et al*. ParA-mediated plasmid partition driven by protein pattern self-organization. EMBO J. 2013;32:1238-1249.

Chapter 3

Atomic Force Microscopy

3.1. Introduction

Atomic force microscope (AFM) was invented by Binnig *et al*[1] as a type of scanning probe microscope, the first one of which was the scanning tunnelling microscope (STM) that was used to image conducting/semi conducting material surfaces, primarily single atoms. AFM is an imaging technique that allows topographical mapping of conducting and insulating surfaces with atomic resolution[2,3]. It primarily consists of a very sharp tipped cantilever that scans a sample surface to produce a topographical map of the area of interest. The topography thus measured is dependent on a feedback loop that maintains a constant force between the tip and the sample. The interactions between the tip and the sample can be used to study the properties of the sample, tip and the medium. For imaging with AFM, it is essential that the sample is immobilized on a flat surface to prevent any movement of the sample by the cantilever tip[3]. This has been perfected over the years by different sample preparation methods in different laboratories[4–13].

AFM can image the structure of the surface of objects, but it lacks the ability to identify target molecules that don't have a distinct structural feature, especially in a crowded system containing multiple species. Hence, AFM can be used as a complementary technique to other biophysical imaging platforms (*eg.* florescence microscopy) to obtain the structure as well as the identity of molecules [14,15].

3.2. Types of AFM

Two main categories of AFM can be described, depending on the way the cantilever tip scans the sample surface.

3.2.1. Contact Mode

In contact mode AFM, the tip always touches the sample surface during scanning to obtain sample topography. This method is primarily used for hard samples as the cantilever causes a vertical and lateral force. The former is caused due to the force of the cantilever on the surface and the lateral force occurs as the cantilever tip pushes the sample during scans[16]. These forces cause damage and deformation of biological samples[17,18], or displacement of weakly attached samples[16]. To counter this situation, cantilevers with low spring constants are used to reduce the force applied on the sample surfaces, yet for lateral friction forces, the spring constant remains high[17,18].

3.2.2. Intermittent Contact Mode (Tapping Mode/Amplitude Modulation Mode)

Tapping mode AFM is a type of non-contact mode of AFM where the amplitude of oscillation of the cantilever is larger[19]. In this mode, the cantilever oscillates near the first resonance frequency. When the tip interacts with the sample, the oscillation amplitude decreases due to energy dissipation and resonance frequency shifts (leading to phase shift *w.r.t.* excitation signal). The feedback loop adjusts the tip-sample separation to keep the amplitude at a constant value. This mode gives the advantage that the time duration of contact of the tip with the sample is very small, which decreases the shearing forces. Therefore, the tip induced sample deformation is minimized (figure 3.1), allowing tapping mode to become suitable for soft biological samples [17,20]. There is also the freedom of not requiring the sample to be immobilized hard onto the surface, allowing imaging in conditions closer to the physiological environment[21].

There are different parameters that can be imaged in tapping mode. The height signal reveals the topography details of the sample. The amplitude signal denotes the error in the feedback loop (*i.e.* the difference between the target amplitude and the actual amplitude) which varies with the topography of the sample. The phase signal denotes the angle lag between the piezo

drive and the cantilever motion and is related to the elastic and dissipative properties of the sample. This parameter can be used to image a poorly attached sample and for higher speed of imaging [16].

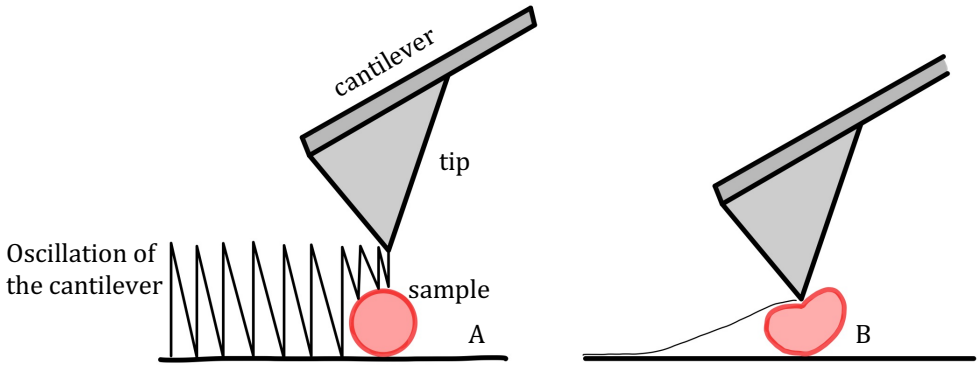


Figure 3.1: Difference between A. tapping mode and B. contact mode AFM. In tapping mode, the cantilever oscillates and does not shift the sample. In contact mode, the lateral force generated is high and the sample is either moved or deformed. Reproduced from Bezanilla *et al* [20].

3.3. *Feedback Operation*

AFM works on a feedback loop that maintains the force between the sample and cantilever tip. In contact mode AFM, a laser beam is incident on a cantilever and reflected into a split photodiode, with the optical lever amplifying the deflection[19]. The variation in the deflection signal relative to the setpoint is fed into the feedback loop. The normalized difference between the top and bottom photodiode signals provides a signal proportional to the cantilever deflection. The signal controls the z-motion of the *xyz* translator to keep the cantilever deflection and therefore the force constant for imaging[22].

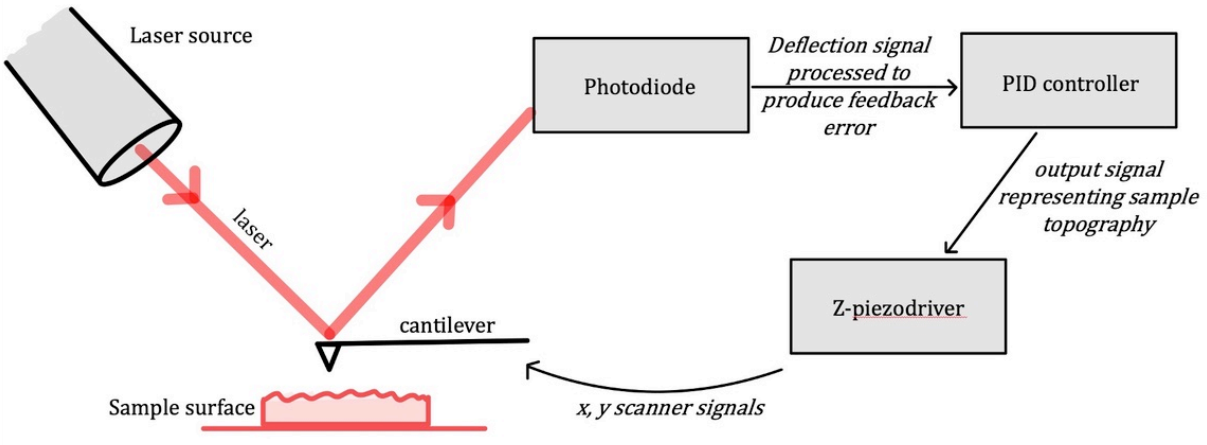


Figure 3.2: Schematic of the feedback loop of AFM.

In tapping mode, the cantilever has a free amplitude of oscillation (A_0) that depends on the nature of the cantilever and the medium in which the sample is being imaged. A_0 is the amplitude at which the cantilever oscillates when there is no surface damping its motion. When the cantilever is brought close to the sample surface, it is set to oscillate at an amplitude, called the amplitude setpoint (A_s), that allows the cantilever to oscillate to detect the topography of the sample surface without damaging the details of the sample. When the sample surface is rigid enough *w.r.t.* the cantilever,

$$A_s = rA_0;\ 0 < r < 1,$$

i.e. the amplitude for oscillation is reduced from the free amplitude by a factor of r [21,23] and the actual oscillation amplitude A is fixed at A_s by feedback control[18,23]. As already mentioned, the reflection of the laser beam from the back of the cantilever is detected by the position-sensitive photodetector. The small displacement of the cantilever is converted into a large motion of the beam on the detector by the optical lever arm, which thus functions similar to a mechanical lever [24,25]. The cantilever deflection is measured by optical beam deflection system (OBD). The laser beam reflected from the cantilever (figure 3.2) is detected by the photodiode sensor and processed into a signal that is sent to an amplitude convertor to obtain the feedback error. This value is sent to the proportional integral derivative (PID) feedback controller, producing an output that represents the sample height and gives the topography.

The feedback machinery consists of a Z-peizo that controls the vertical height to which the cantilever oscillates. The output signal produced by the PID feedback controller (described above) is fed to the Z-peizodriver. It amplifies the signal to displace the Z-scanner such that the feedback error approaches zero. Additionally, using a two-channel digital/analogue converter (D/A), the computer generates the X- and Y- scanner signals that define the scan area and are fed into their respective peizos [18,23]. Thus, the next oscillation of the cantilever is directed into the subsequent loop of surface detection, laser deflection, feedback error generation and *xyz* piezo modulation.

3.4. Imaging Environment

AFM can be used to image a sample in vacuum, air and liquid. The imaging in vacuum, though difficult to maintain, provides the advantage of being the medium that provides the best resolution out of the all the environments. Imaging in air, on the other hand, is the simplest of all environments, as in the sample preparation and the maintenance of the imaging conditions

is very straightforward. However, imaging in air provides the disadvantage of having the least amount of spatial resolution. Also, the capillary forces of the tip produced due to the presence of a water layer between the tip and sample cause the deformation of sample details by the larger force of interaction between sample and tip (figure 3.3). This is the reason for the high resolution of imaging in vacuum as the absence of a water layer avoids the above-mentioned limitation[16]. Also, when temperature need to be drastically low, cryo-AFM can be used in a high vacuum[26]. The resolution can be further enhanced by imaging at low temperature which reduces mobility of the sample during imaging.

The sample can also be imaged in liquid, either water or buffer, and sometimes also alcohol and toluene. Imaging in liquid has the advantage of providing better resolution than imaging in air (explained below) and that the sample can be imaged in buffer conditions mimicking the physiological conditions, especially useful in studying biochemical reactions and molecular interactions. Interestingly, AFM imaging of DNA in liquid has been found to preserve the B-form and its related mechanical properties, contrary to imaging in air that disturbs the helical pitch at various salt concentrations[27] (see section 2.1).

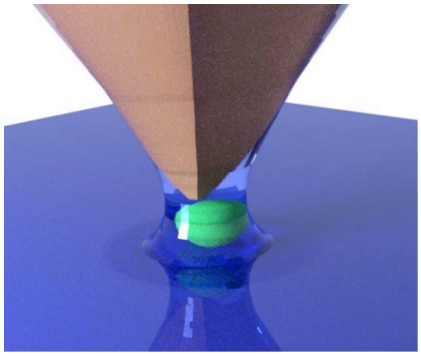


Figure 3.3: Water covering the tip during imaging in air increases the forces of attraction between the sample and the tip. Image reproduced from Kasas and Dietler, 2017 [16].

The resolution of the images depends on the kind of sample and its preparation, and the imaging parameters. There are many reasons for the better resolution of images obtained in liquid compared to imaging in air, though many images in air have been obtained that resolve features with comparable resolution. Since most of the tip-surface forces in AFM are electrostatic, water and electrolyte solutions screen out long range electrostatic forces, so the force the tip feels is more localised to the features on the sample. This enhances the amount of details that tip can detect on the sample surface. Also, in air, water condenses well below the saturation vapour pressure close to highly curved surfaces, like the tip-sample junction (figure 3.3) . To stop the

tip getting stuck in the resulting capillary, a larger amplitude is utilized in air than in water, which means that the AFM is more sensitive to long range forces and cannot detect small details of the sample. Moreover, on hydrophilic surfaces in ambient air (like mica, nucleic acids and proteins) surface water tends to fill gaps and smoothens the surface features to minimise the energy cost due to surface tension. This obscures the small features. Additionally, biological samples are usually immobilized from solution containing salts and buffer, which most likely cover up the minute details of the surface despite meticulous rinsing with water. Furthermore, the structures that biological samples form in the ambient liquid environment might not be formed in air, hence are not observed.

Imaging in propanol has been reported to provide the advantage that the DNA is not damaged after even an hour of imaging, possibly owing to the insolubility of DNA in propanol. On the other hand, imaging in buffer has been found to cause the DNA to appear larger and wider because the tip also images the hydration layer around the DNA[28,29]. It is also found that slower scan rates and large scan sizes cause less damage to the DNA sample, with smaller scans as small as 50 nm possible in propanol compared to about 500 nm required in buffer imaging. With smaller scans, as with more scan lines, the tip encounters the sample more frequently and damages the sample in the consecutive scans, limiting the number of images that can be obtained. Slower scans allow more time for the feedback system to respond to height variations during imaging, while large scans increase the gap between scan lines thus reducing the sample damage. Interestingly, it has also been reported that imaging in buffer displays a repulsion between the tip and DNA compromising the details of the molecule that can be resolved, and the addition of $MgCl_2$ to the buffer eliminates the repulsion[28].

The understanding of the apparent height of features of the sample depends on the force of oscillation, deformation or contamination of the cantilever tip, presence of hydration layer on the surface and tip-sample forces of interaction. General expressions have been developed to interpret the surface topography. Using models and approximations it has become possible to recover true height, error in height and contributions from geometrical and chemical factors[30].

3.5. Cantilever Tips:

AFM measurements depends on the precise geometry and chemical properties of the cantilever tip apex. Imaging in liquid is more complicated by the effect of solvation forces. The tip can greatly affect measurements because of its chemical composition, geometry and surface charge. But since tip characterization is difficult, time consuming, affected by possible contamination and possibly damaging to the tip itself, the structural and chemical properties of the tip are not known well. For very sharp tips, however, the measurements are dominated by the solvation structure of the sample and the influence of the tip is minimal [31].

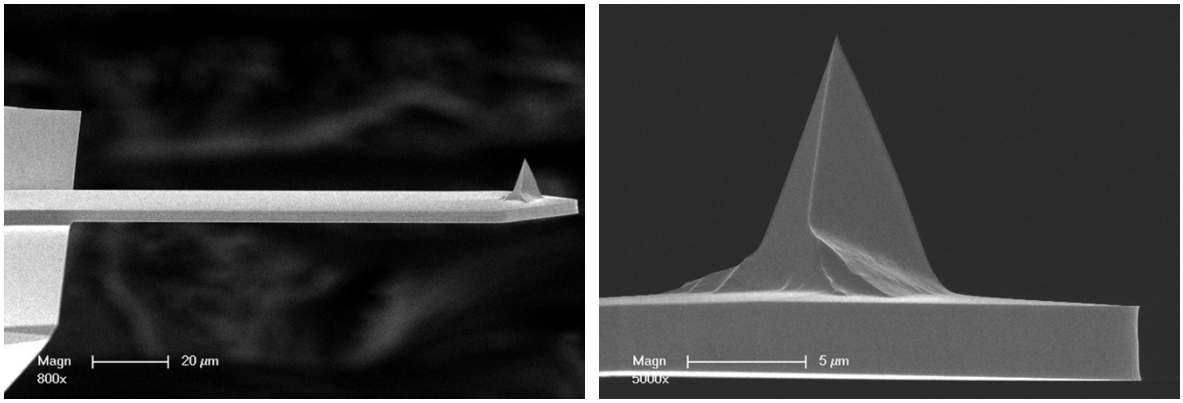


Figure 3.4: Bruker Tespa-V2 for imaging in air.

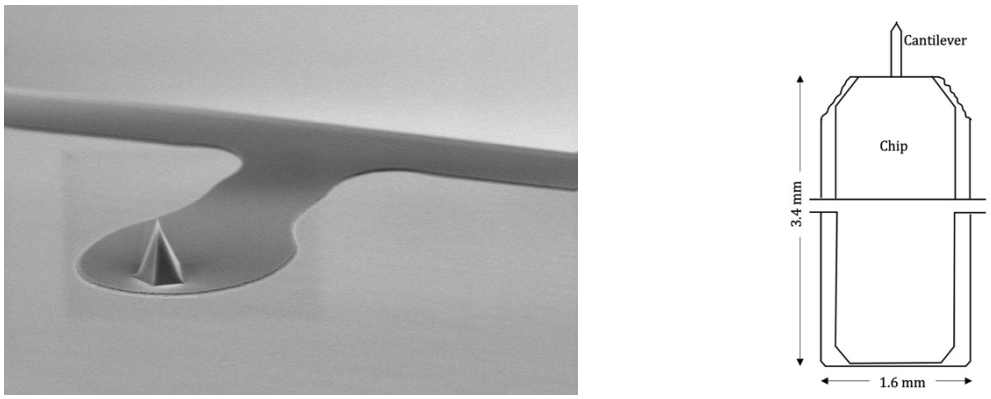


Figure 3.5: Bruker Fastscan D and AC-40 (also called the biolever mini) for imaging in liquid, see appendix for details.

Commercially available microfabricated cantilever tips are usually made of silicon (*e.g.* Bruker's Tespa V2) or silicon nitride (*e.g.* Bruker's MLCT and AC40-for details see appendix 2). Both the materials are oxidized and tuned to modify their properties[32]. It has been demonstrated by Akrami *et. al.* that the efficiency of the tip depends on the presence of stable hydration sites at the apex [33]. These sites can be created by Si-sputtering on commercially available tips, thus controlling the influence of tip properties on high-resolution images. Other

techniques have also been employed to produce stable hydration sites, like UV/O_3 cleaning, Ar plasma and Ar sputtering, though they are not as efficient as Si-sputtering. The method has been found to provide the additional advantage of removing organic contaminants when immersed in water rapidly after treatment [31].

From the earlier experiments, it has been observed that cantilever tips can damage or move the DNA and that there are repulsive forces between the DNA and the tip, which can be removed if the silicon nitride tips are SEM modified with carbon. Also, it is found that sharp tips and blunt tips are equally gentle while imaging DNA, because the DNA damage is caused by the electrostatic repulsion, and not by puncturing of the sample owing to the sharpness of the tip[28].

Organic contamination of the tip has been found to be detrimental to the AFM measurements, especially imaging in liquid, as the tip is rendered hydrophobic and the tip hydration made unstable. Moreover, tip treatments have sometimes proved to be harmful for the tip sharpness [31] as additional molecules of reagents used for tip treatment stick to the tip and add to its radius. But interestingly, a protrusion or slight sharp artefact on the tip have produced some rare instances of very high resolution images [34,35]. Recently, the tip modification has facilitated DNA interactions with crosslinked porous protein crystals to be measured by high-resolution AFM, and has allowed resolution of protein details without damaging the protein sample [36]. Also, measurements of dsRNA at consecutive frames with different oscillation forces[37] has assisted the already reported observation that the force of tip oscillation affects the AFM height of the sample [35,38,39]. Simulation studies of dsRNA images have revealed that to obtain high-resolution images of DNA, sharp tips of radius below 2.5 nm is required. It has come to light that to observe the major and minor grooves of double-stranded nucleic acids, this is the suitable tip radius. However, the sharpness of tip has been reported to improve the resolution of major or minor grooves to different extents and a suitable explanation has not been proposed[37].

Thus, it has become clear that the characteristics of the cantilever significantly affect the image quality and progress in this direction is imperative in the expansion of the role of AFM in biological and dynamic imaging.

3.6. Substrate Surfaces (for Nucleic Acids and Proteins)

As already mentioned, for AFM imaging, the sample must be immobilized on an atomically flat surface. For dynamic imaging, the substrate should be selectively bound such that it is immobilized, and at the same time partially free so as to prevent any interference with the process/interaction under study[40–42].

Numerous types of surfaces have been used for sample immobilization [6,13,18,40,43–63]. The types of surfaces used for HS-AFM so far are: bare or chemically modified mica [63], supported lipid bilayers (SLB) on mica, 2D crystals of streptavidin grown on planar lipid bilayers containing biotin lipid, highly oriented pyrolytic graphite (HOPG), DNA origami tiles placed on mica surface[18,23], glass coverslips (usually roughness of about 1-2 nm[19]), plain or silanized, graphite (HOPG) and gold. The substrate can also be derivatized if required. For immobilization, glutaraldehyde has been used for liquid imaging of proteins [26] and most recently, Wang *et. al*. have demonstrated that porous protein crystals can be used to bind DNA for AFM imaging [36].

For immobilization, the buffer composition of the DNA solution plays a role in the stability of the sample on the surface and DNA- protein interaction. Ideally, the DNA concentration used is ~ 10 nM with buffer containing $MgCl_2$ (10 mM), HEPES (4 mM) and KCl (50 mM)[64]. This buffer composition closely resembles the buffer employed for biochemical assays and at the same time, does not contain components that may interfere with the tip-sample interactions during imaging.

3.6.1. Muscovite/synthetic Flourophlogopite Mica

Mica is one of the most commonly used surfaces for sample immobilization for AFM imaging. It is made of layers of aluminium phyllosilicate lattice ionically bound by interstitial K^+ ions [65] and is known as sheet silicates owing to the property of forming distinct layers. Cleaved mica has been used for the synthesis of carbon support films, growing cells and thin films in electron microscopy.

Ruby muscovite (mica with colour ranging between white to red to brown [66]) is a commonly used type of mica for AFM imaging and is a hard material that cleaves easily without cracking and splitting along natural cleavage planes to from extremely flat surfaces. This mica has the advantage of being quite transparent and not get scratched or contaminated, and is a suitable substrate for dynamic imaging DNA and DNA-protein complexes[67].

3.6.2. Divalent Cation Treatment of Mica

Mica provides an atomically flat surface over a large area for AFM imaging of samples a few nanometres high[19]. When cleaved and washed with water, it becomes hydrophilic and negatively charged as the K^+ ions located in between the layers become desorbed. Thus, different samples can be immobilized on the mica surface by charge interactions [65] (figure 3.6). In this aspect, monovalent cations (K^+, Na^+, and Li^+) reduce the binding affinity while negatively charged molecules like DNA can be anchored on mica with the aid of divalent cations (Mg^{2+}, Ca^{2+},Ni^{2+} or Zn^{2+}) in the buffer [23,26,68]. This is because on cleavage of the mica layers, vacancies are created due to the splitting and clustering of interstitial K^+ ions. These sites can be filled by divalent cations, or the residual K^+ can be displaced by transition metal cations like Ni^{2+}, Co^{2+}, and Zn^{2+} to produce a carpet of anchor points on the mica surface of cations that continuously interchange between K^+, H^+, and other cations such as Mg^{2+}. Interaction with the OH^- of mica surface produce a positively charged phase separated domain. The transition metal cations bind negatively charged sample to mica by directional bonding of their d-orbitals and restrict the motion of sample molecules [65], thus immobilizing them for AFM imaging.

Mg^{2+}, on the other hand, cannot directionally bond as it interchanges continuously with residual K^+ and H^+ ions on the mica surface and the imaging buffer. Therefore, it allows a relatively weak and diffused adhesion between DNA molecules and mica, as compared to Ni^{2+}, causing DNA to adopt an equilibrated form. Hence, they can be used for immobilization of the sample, such as nucleic acids, loosely on the surface to allow biomolecular interactions that can be imaged using dynamic AFM [65,69]. Additionally, Mg^{2+} ions provide the advantage of acting as enzyme co-factors for a large number of biological processes studied by AFM [65], hence they further facilitate the dynamic imaging process.

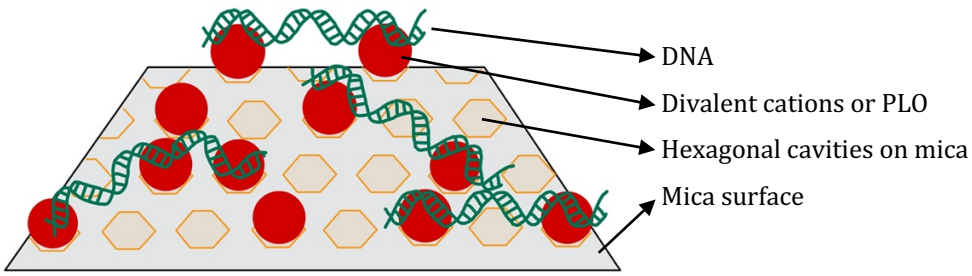


Figure 3.6: Schematic of the surface of mica treated with divalent cations or PLO with DNA immobilized on it.

It has been reported that DNA binds tightly to the mica surface in the presence of those divalent

transition metals whose cationic radius ranges from 0.69 Å to 0.74 Å, namely Ni^{2+}, Co^{2+} and Zn^{2+}. This has been explained based on how well the cations fit into the hexagonal cavities above the deeper seated and exposed hydroxyl ions of freshly cleaved mica. Although Mg^{2+} has a cationic radius close to the above-mentioned cations, Mg^{2+} treatment to mica does not bind DNA as well as Ni^{2+} does [69,70], which might be related to higher enthalpies of hydration of the latter, thus affecting its ability to bind DNA (or any other ligand) to from stable complexes. Also, the DNA binding has been reported to peak at cation (primarily Ni^{2+}) concentration of 1 mM, below which the concentration is speculated to be too low to bind DNA, and above which the cations possibly saturate the gaps in mica lattice (0.5 nm hexagonal cavities [60] as compared to the size of 0.3 nm for phosphate groups of DNA). Therefore, at the suitable 1 mM concentration of cations, every third base of DNA can bind to every second mica site[45]. This difference in the binding capacity can be applied to alter the DNA surface adhesion by regulating the concentrations and ratio of monovalent and divalent cations, and their exposure time.[65]

It has been proposed that for larger cations, like Ca^{2+} to counter the negative charge of mica, the washing of cationic solution treated mica with water might play a role in immobilization of nucleic acids, as the H^{+} ions would replace the larger divalent cations in the mica cavities and hence facilitate the binding of DNA to the thus treated mica surface[71]. It is worth mentioning here that when immobilizing using Ni^{2+}, if the ions form complexes with the base pairs to absorb the double-stranded nucleic acids to the surface (like it has been done for dsRNA[72]), the helical structure of the nucleic acid may be affected and prohibit the resolution of the grooves, especially the minor groove[37].

Characteristics of DNA Immobilized in Different Solutions

In the early studies of DNA using AFM, DNA was deposited on Mg^{2+} treated mica but it mobilized during imaging. It was found that DNA could be dehydrated with propanol or imaged with an SEM modified tip to avoid the scraping. The immobilization could be done on magnesium acetate [6,28,73,74] or calcium acetate treated mica and the DNA would immobilize in the B-conformation. Dehydration of DNA allowed it to bind better to the mica by removing the hydration layer. However, it was interesting to see that DNA displayed A-conformation when imaged in propanol followed by imaging in buffer. The length of DNA (as a function of the configuration of DNA, whether A or B) was dependent on the environment that the DNA was immobilized in and did not change when the environment of imaging was changed. But

when DNA was imaged in propanol after imaging in buffer, it probably became loosened and precipitated again on the surface as A-form [28,73,74]. It was also fascinating that long DNA fragments appeared to retain the conformation of B-DNA while short DNA (100-200 bp) had the dimensions of A-form [75,76]. This was probably because longer DNA have more anchor points to the treated mica and hence bound firmly. Short DNA, on the other hand, have comparatively fewer anchor points thus giving the DNA freedom to move and take the conformation of A-DNA during dehydration [76]. Longer DNA strands have a smaller fraction of DNA base pair conformation changing upon drying. But, the total length of DNA that changes conformation remains the same as the length of the B-form of DNA[77].

Short DNA, about 50 bp long, immobilized in Mg^{2+} containing buffer, have been found to appear not as rod shapes but as globules. It has been proposed that DNA was so short that assuming that they have the base pair spacing as the longer 100 bp or 200 bp DNA, the dimensions were almost spherical. Alternatively, there was also the possibility that some denaturation of dsDNA ends occurred that gave the DNA a globular shape[76].

It is worth mentioning that Mg^{2+} in the imaging buffer facilitated base pairing between DNA molecules and caused them to be immobilized with lumps [76]. It has been reported that Mg^{2+} can cause DNA to aggregate by stacking together when the sample is immobilized on Mg^{2+} treated glass and more so in Mg^{2+} solution during AFM [78]. It has also been found that Mn^{2+} deposited DNA displays conformation similar to silane treated molecules, contrary to Mg^{2+} bound DNA. This can be due to stronger binding of DNA to the surface with Mn^{2+} than Mg^{2+}[77].

To prevent the aggregation of long DNA molecules when immobilized on the surface, Heenan and Perkins have proposed a method to immobilize the DNA gradually, with the buffer composition on the mica surface controlled to be same as the buffer of DNA solution. It was ensured that the transfer of DNA was gradual from the 3-D solution onto the 2-D surface of mica to allow the DNA to spread, and anchor on the surface using Mg^{2+}. The anchor was then sealed with Ni^{2+} to strongly bind the DNA to the surface for imaging [79].

3.6.3. Poly-L-Ornithine (PLO) Treatment of Mica

Recently, poly-L-lysine$_x$ -b-PEG$_y$ block of polymers have been used to treat mica surfaces to selectively bind DNA for imaging DNA-protein interactions such that the protein does not non-specifically bind to the surface to mask information[80].

PLO is a positively charged synthetic amino acid chain and consists of one hydrobromide per unit of ornithine. It is primarily used for cell adhesion to plastic and glass surfaces and in combination with laminin for neuronal and neural stem cell cultures. The molecular weight of PLO varies and as a result produces varying number of binding sites with a range of viscosity[81]. It has been used to immobilize DNA as well as DNA-protein complexes on mica for AFM imaging [61] by countering the negative charge of the mica surface.

3.6.4. Silane Treatment of Mica

The mica surface contains silanol Si-OH, which can react with organosilanes to form siloxane bonds Si-O-Si, thus allowing attachment to various functional groups. Aminosilanes and silatranes have therefore been used to functionalize the mica surface to image nucleic acids and nucleoproteins [13,82]. These allow imaging in a variety of ionic, pH and temperature conditions. The samples are stable once prepared, and don't absorb any contaminants for weeks, with minimal storage precautions [82]. This method allows strong binding of DNA, allowing 2D projection of the 3D DNA structure [77].

3-Aminopropyltriethoxysilane (APTES) has been used frequently to attach amino groups to mica and provide a positively charged surface [3,58,70]. APTES, can polymerize to form rough surfaces, and hence need to be handled carefully. 1-(3-Aminopropyl) silatrane (APS) can also be used the same way and has the advantage of being less reactive, and resistant to polymerization and hydrolysis at neutral pH, ensuring a smooth surface. Amino groups can be further used for protein immobilization to the surface, *e.g.* using amino-reactive glutaraldehyde[18]. APS treated mica is found to be a stable method of binding DNA, but since the attachment is very strong, DNA molecules entangle and don't spread on the mica surface[43]. A 'molecular combing' mechanism [83] has allowed DNA spreading on the modified glass surface and the method was extended on APS treated mica to obtain DNA immobilized on a mica surface in an extended form to be imaged stably in a buffer solution [40].

This method is a frequently used method of DNA immobilization but has not been used in the current work.

3.7. Speed limits of AFM:

AFM is a slow imaging technique, though a lot of progress has been made to improve the speed of imaging since its birth in 1986. There is an inevitable delay in the feedback loop through several devices [23] which affects the speed of imaging. The main time delays are the response time of an oscillating cantilever, measurement of cantilever oscillation time, integration of the error signal with the feedback controller, z-scanner's response time and parachuting [68].

Parachuting is the loss of contact between the cantilever tip and the surface because while scanning the surface, especially the downhill regions, it takes a long time to make tip-surface contact if it is lost. This causes a major hindrance in obtaining high-speed and low invasiveness simultaneously. A 'dynamic PID control' has been used to add a false error signal proportional to the real error signal to the latter when $A>A_s$ (the amplitude of oscillation A is more than the amplitude setpoint A_s) *i.e.* during downhill region scan. Also, feedback controllers have been developed by Ando *et. al*. that change the gain parameters depending on the cantilever oscillation amplitude. Thus, the feedback gain is increased while going downhill on the sample[23]. This reduces parachuting [68].

3.8. Sample Disruption:

The tip sample forces may disrupt the sample, so improvement in signal-to-noise ratio by use of small cantilevers, allows high resolution imaging at low forces. In tapping mode AFM, the time averaged tip-sample force F_{TS} is given by:

$$\langle F_{TS} \rangle = \frac{k}{2Q}\sqrt{A_0^2 - A_{SP}^2}$$

Assuming negligible tip-sample power dissipation with Q as the quality factor, A_0 as the free amplitude and A_{SP} as the set-point amplitude, Q is nearly unity in liquid. Therefore, for low tip-sample forces, cantilevers with low k (stiffness) are used with a small A_0 , and A_{SP} as close to A_0 as possible. When k and A_0 are both reduced, the thermal noise increases with decrease in k. This noise, as well the noise in the optical detection system can be reduced by the use of small cantilevers. Shorter cantilevers have been found to produce a greater angular displacement for a given spatial displacement of the tip, giving a large signal at the detector and reducing the noise contribution of the noise sources like the laser fluctuations. Additionally, since the tip spends a greater fraction of time close to the surface, application of

a small amplitude allows high resolution imaging and lower tip-sample forces with all other parameters being equal[84].

The mechanical quantity that influences the samples by causing a momentum change is not force but the impulse (product of force and time) of the cantilever [57]. Therefore, even if the cantilever resonant frequency is high (*e.g.* in water and for small cantilevers), the acting time remains low. Hence the F_{max} will not be high enough to damage fragile samples. Additionally, during high-speed imaging, it is difficult to reduce the invasiveness if the amplitude setpoint is very close to the free oscillation amplitude. This causes parachuting (as already described) that can been avoided by the development of an improved PID controller [68].

Drift in lateral dimensions is an issue associated with AFM that compromises the quality of the images. It inhibits the to return to a particular sample position, which poses difficulties especially during long durations of imaging. Lateral drift also limits the signal to noise ratio in an image, especially in cases when an image must be acquired too quickly to average away the thermal noise, like in HS-AFM. Vertical drift, on the other hand, occurs when unwanted vertical motion happens due to the movement of sample or bending of the cantilever. There is another kind of drift, the force drift where images of biomolecules vary with force. This can be avoided in contact mode by manually updating the force set point. Thus, force drift diminishes the results, interpretability, and ease of the imaging experiments [25].

3.9. Dynamic Imaging of Biomolecular Processes -High Speed AFM

HS-AFM has developed over the years with the capability of imaging molecular dynamics with a rate approaching video optical microscopy and nanometre scale resolution[59]. HS-AFM can be used for dynamic imaging of biological processes as well as observing non-biological dynamic processes at solid-liquid interface, *e.g.* electrochemical reactions, corrosions, catalysis, action of detergents *etc*. to facilitate the creation of 'wet nanotechnology'[68]. In this direction, remarkable developments have been made in the field of HS-AFM by Uchihashi *et al.*[14,15,18,23,56,57,85–91] and Ando *et al.*[18,23,56,68,85,86,89,90,92–94]

To achieve the fast rate of imaging, some steps have been taken in the direction of improvement of the technique as a whole, namely the reduction of cantilever size and the corresponding changes to the instruments to accommodate the smaller cantilevers,

modification of the laser focus system and data acquisition electronics to ensure the faster and higher data transfer, reduction of the scan size and limiting the sample to the ones that are not very high[16,95] (not more than 3 μm [96]). Phase imaging has been used instead of the amplitude parameter in some cases for imaging at a high speed[16].

HS-AFM requires reduction in time delay in feedback loop[97], damping of mechanical vibrations due to rapid displacement of scanner and elimination of parachuting. Considering a given displacement of the cantilever free end in the z direction during tapping, a shorter cantilever has a larger angle change, thus a better displacement detection sensitivity compared to a larger cantilever. This, and small thermal noise effect, allows a very small displacement to be detected, even with high bandwidth detection[23].

The HS-AFM instrumentation has been developed (as already mentioned) consisting of a fast scanner, feedback controller, an optical beam deflection (OBD) detector and fast electronics to detect deflection of a very small cantilever, the latter having a very high resonant frequency. In tapping mode, the cantilever is oscillated to allow the tip to intermittently contact the sample, thereby imaging at ~5-20 frames/second (fps), without affecting the structure, interactions and functioning of the delicate proteins/nucleic acids[57]. However, to aid the imaging of reaction processes using HS-AFM, it is also useful to slow the reaction process, for instance by reducing the temperature of imaging to reduce the rate of perforin protein assembly experimented by Leung *et. al* [98]. Also, the concentration of nucleotide triphosphates (NTPs) has been used to slow the rate of transcription for AFM imaging. For studying biological processes, balance has to be maintained between the scan rate vs the scan size, depending on whether the focus of the study is percentage activity of an enzyme or the best sequence of intermediate steps [41].

During HS-AFM, the tip tends to be degraded during consecutive scans, hence, steps must be taken to improve the AFM setup so as to avoid tip damage. Attempts have been made to develop methods to increase the scan speed, and allow dynamic imaging without damaging the tip, enabling sample modification during reactions and simultaneous AFM scanning [99]. As a result of these measures, the speed of imaging has been reported to increase to as much as 70 frames /s with some images collected in less than 15 ms, using mechanical feedback loops and micro-resonant scanner [100].

Many processes related to DNA have been studied using HS-AFM like action of methyltransferase restriction enzymes, unwrapping of DNA-histone complex, DNA base

excision repair [68] and poly(ADP-ribose) polymerase during DNA repair[101]. Recently, the conformation changes of DNA on interaction with polycomb repressive complex 2, which is a histone methyltransferase, were imaged using AFM [102]. The technique has been used for real-space and real-time dynamic imaging of CRISPR-Cas9 in action, with nanoscale resolution, accompanied by complex assembly, searching for targets, target cleavage and product release. Also, the structural transitions of the nuclease domains during Cas9 mediated DNA cleavage have been visualized[103].

Streptavidin 2D crystals on biotin lipid bilayers have been used as the surface of HS-AFM, *e.g.* for the observation of actin filaments [68]. Work has been done on the activity of intrinsically disordered proteins to identify the folded and disordered regions and the transitions in between, along with the estimation of amino acid residues in the disordered regions allowing an understanding of the dynamic structure of proteins[104].

It is difficult to determine the polarity (5' or 3' end) of DNA strands in AFM. Several methods have been devised to solve this problem. Biotin-streptavidin labelling or streptavidin conjugated gold particles attached to one end of the DNA strand to form a small blob can be easily detected under the AFM, or DNA origami tiles have been attempted to determine strand polarity[17]. Attempts have also been made to automate the progressive detection of DNA strands in images starting from small segments of the molecule to the neighbouring pixels to allow automatic tracing of DNA strands in their entirety [105,106].

Despite the promising applications, there are currently many limitations of this technique. HS-AFM is currently limited to a small scan range of few μm to obtain high resonant frequencies for fast scan and a substantial deformation of very soft molecules (like cell membranes) occurs during repeated scanning to obtain multiple frames. Work is underway to overcome these limitations so as to image over a wider sample area, allow in-situ imaging of dynamic processes, obtain faster wide area observation of biological samples and high speed non-contact imaging [23].

3.10. Applications of AFM:

AFM has been used to image the surface features and conformation of biological samples and the interactions between biomolecules.

RBCs and bacteria immobilized on coverslips being one of the first cells to be imaged using AFM[107], the technique became applicable to image other types of cells like WBCs neurons, muscle cells[108], and cell organelles like chromosomes [109]. Work has been done to investigate if cells can be described with elastic modulus in a self-consistent way [110].

Many protein molecules like immunoglobulins, enzymes, actin filaments, fibrin, membrane proteins, *etc.* have been imaged [111–114], though the resolution of images has not been very good to observe the individual atoms. This can be credited to imaging forces, motion of protein molecules and their hydrations layers inhibiting detection of the actual surface of proteins [115]. Using AFM, the volume of proteins has been related with the molecular weights and the polymeric state of the proteins resolved[116]. Many attempts have been made to image lipids and fatty acids as well[117].

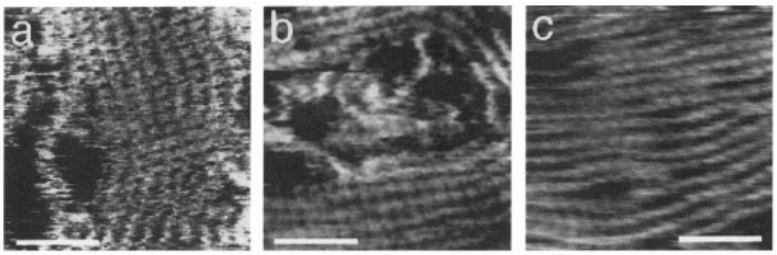


Figure 3.7: The right-handedness of the dsDNA can be seen resolved. a, b and c are different sources of DNA, namely pBR322(4.36 kB), pBR325(6 kB) and *HaeIII* restriction fragments of ØX174. Scale bar 20 nm. Reproduced from J. Mou *et. al.* [10].

AFM has made numerous developments, enabling its use for high resolution, routine and dynamic imaging of DNA[77,118]. AFM has been used to detect UV damage in DNA molecules[119]. From the height measurements by AFM, the chirality or handedness of supercoiled plasmid DNA has been determined [120]. DNA double helix and right-handedness has been visualized on AFM by imaging DNA immobilized on cationic lipid bilayer surface (figure 3.7)[10]. The helical resolution of DNA has been obtained by FM-AFM [35]. Though, it has been difficult to visualize the minor grooves of dsDNA, despite being deeper than the major grooves. This has possibly been because of the former being narrower, as in the distance between the phosphate sugar backbone in the consecutive turns is shorter for minor groove

than the major. This dilemma has been found to be further enhanced in case of dsRNA where the minor groove is slimmer[37].

In the early 1990s, scanning force microscopy (SFM) was used for real-time and in-physiological imaging of biomolecules in aqueous medium. Guthold *et al*. imaged the action of RNA polymerase on DNA[121]. DNA fragments in HEPES/$MgCl_2$ immobilized mica surface were imaged before and after the addition of *E.coli* RNA polymerase. The results represented the first step in the technique of dynamic imaging of protein-DNA interactions. Before this, experiments were performed to image living mammalian cells in changing physiological conditions and to study the process of crystal growth in cell lysozyme [121]. Time lapse AFM of dynamics of cruciform and H-DNA structure have been performed, that showed the interplay between local DNA conformations, global conformations, and dynamics of the entire molecule. Such imaging has been possible because the sample could be imaged without drying up. This has allowed the imaging of genetic processes.

dsDNA complexed with cytochrome c and klenow fragment of DNA Polymerase I bound to ssDNA of M13 was immobilized on carbon coated mica surface and imaged on home-designed AFMs as early as 1992 [5]. Dynamic imaging of lysozyme was performed in 1994 to observe the height changes in the enzyme immobilized on mica, corresponding to the conformation changes in the enzyme during hydrolysis [122]. Enzymatic breakdown of DNA immobilized on Ni^{2+} treated mica was imaged via dynamic AFM in 1994 to observe the motion and degradation of DNA by DNase I [20]. Enzymes phosphorylase b and phosphorylase kinase involved in the muscle cell activity [123] have also been imaged (1990) to understand catalysis mechanism.

It has been shown that AFM can be used to cut DNA at specific locations [6,124,125]. DNA-protein binding happening at biotin-streptavidin labels immobilized on ruby-mica was imaged very early-on using AFM demonstrating that protein complexes with tag molecules could be seen bound to DNA[126]. The length of DNA in a nucleosome has been calculated from plots of DNA vs nucleosomes assembled on it [8]. DNA has been imaged complexed with RNA polymerase[28,127,128], DNA polymerase [5] and EcoRI restriction enzyme [129]. The DNA bending by RNA Polymerase [130] and most recently MCM helicase [131] has been studied using AFM. The technique has been used to measure the unbinding kinetics of DNA-protein interactions by a method called pulling assay [132]. Chromatin structure constituting DNA wound around histone proteins immobilized on glass substrate has been imaged and has revealed the conformation of chromatin. The details have assisted the information obtained by electron microscopy [8]. AFM,

in assistance with Raman spectroscopy and light microscopy has been used to study bands and interbands of chromosomes [133]. Overall, from the static images of DNA-protein interactions, measurements of DNA length[134], location of protein binding sites and conformation changes on DNA due to protein binding have been performed[135–138].

Research has been done to suggest that many proteins that interact with DNA during replication and gene expression, first bind non-specifically to the DNA, and then reach their target by one dimensional diffusion [17,42,139]. The possible site searching models that occur either individually or simultaneously are random collision, hopping or sliding [42]. In this direction, scanning force microscopy has been used to study the same during transcription[139]. Most recently, the technique has provided evidence that a previously unremarkable tyrosine residue in AP-endonuclease I is important for anchoring the DNA intercalating loop and stabilizing the protein's active site [140].

Force measurements with functionalized/ modified microfabricated tips or colloidal probes are called "chemical force microscopy" [2]. Force curves are plotted using the cantilever deflection as a function of vertical displacement of the piezoelectric scanner. The cantilever deflection can be used to calculate the force using Hooke's law: F= -kd where k is the cantilever spring constant[32], which in turn can be applied to estimate the strength and elasticity of bonds based on the force measurements. AFM based dynamic nanomechanical mapping (DNM) has allowed an understanding of the functions of biological systems by interpretation of 'force', which is difficult with other techniques [141]. The method has been applied to image folding and unfolding of proteins which has facilitated the estimation of bond strengths and the study of equilibrium during molecular folding and unfolding [25].

Molecular recognition imaging, a combination of AFM imaging and force measurements, has allowed simultaneous topography and recognition images (TREC) of systems like antibody-protein, aptamer-protein, peptide-protein, chromatin, antigen-antibody, cells, *etc.* to be obtained. It requires AFM tip functionalization (silicon, silicon nitride, gold, *etc*.) by interactions like silanization and gold thiol chemistry. Since this technique is based on the specific interactions between the ligand and receptor, it has enabled the recognition of a protein of interest among a lot of proteins, or monitor a biological process in its physiological state [142].

3.11. Selection of Imaging Technique for DNA-Protein Interactions: Why AFM?

The study of function, dynamics and structure of proteins requires a technique that allows imaging of the sample in liquid, obtain high spatial and temporal resolution, low invasiveness to the sample and direct imaging of the sample without markers[18]. Over the years, there have been numerable techniques that have come up in the field of imaging. The classic light microscope is an instrument that has proved to be extremely useful, though it is limited by the difference between the wavelength of light and the size of the sample (<35 nm) making it ineffective for sub-nanometre imaging [143].

Laser Dark Field Imaging is one technique that has proved useful for imaging biological samples, considering its application for measuring movements of molecular motors. The gold nanoparticles used as probes in this technique have low viscous friction and can follow the fast movement of molecules. Though gold nanoparticles are too big (40–200 nm) to allow visualisation of DNA conformation changes under study[143], the method provides the benefit of not suffering from photo-bleaching or blinking. However, the large probe can impair catalysis [144], and relatively large and complicated equipment are needed for the required spatial and temporal resolutions[143].

Fluorescent microscopy is another technique that tracks the molecules by detecting the fluorescence probes attached to the molecules, ensuring that the objects observed are indeed the protein/DNA one is interested in, and not some artefact or contamination[145]. FRET (Förster Resonance Energy Transfer) can be used as a molecular ruler for small conformation changes (≤ 1 nm) or to check if the molecules are interacting or not. Despite the limitations of less than 100% binding efficiency of the organic dyes used in this technique, dyes having chances of being photo-inactive and bleed through into the sample [144], FRET has proved to be beneficial in determining the locations of the prominent domains and substrate binding sites on the proteins. The method has been applied to assist the current project by obtaining structural details of the Flap endonucleases before their study using AFM. An alternate to using probes for imaging, even during AFM imaging, is imaging proteins with enzyme activity cleaving florescent molecules that would generate fluorescent signal when the reaction happens, in combination with various microscopy techniques [146,147].

NMR, FRET and super resolution optical microscopy techniques have a good degree of spatial resolution and have allowed the study of oligonucleotide structures in physiological conditions. However, these methods are limited by the distance between probes, *i.e*. if the probes are located far away on the molecule under study, the features between the probes are not seen [39]. Fluorescence microscopy, optical spectroscopy and magnetic and optical tweezers have a limited degree of directness of measurement, as the protein molecules are invisible themselves. It therefore becomes necessary to image proteins 'in action' at high spatiotemporal resolution[18,57]. Numerous fluorescence probes are available for imaging a large variety of biological samples and can be used as an aiding technique, though not the sole instrument of choice for the current research work, with some limitations. The Immunofluorescence labels, Quantum dots and bead probes used frequently are large (~10 to 100s nm radius) when compared to DNA and proteins structural details (few nm) to allow a resolution (~1 nm) suited to study DNA and protein conformation changes during protein aided catalysis. Biotin labels conjugated to cysteine residues have been used to locate and tracking the proteins molecule without a very high resolution. In these techniques, attention has to be paid to the precise introduction of cysteine residues without affecting the biological structure and functionality of the proteins. Organic dyes like Fluorescein Arsenical Helix binder (FlAsH) and Resorufin Arsenical Helix binder (ReAsH), though smaller (1-2 nm), have been found to bind non-specifically and thus pose difficulty in their introduction to the sample [145].

Fluorescent In-situ Hybridisation (FISH) labelling technique is limited by its long incubation time and irreversible photo-bleaching of reporter molecules like organic dyes and fluorescent amino acids, which restricts their observation time. Yet, the florescent protein tags, HaloTags and SNAP/CLIP tags are used to label proteins frequently and offer many advantages to study changes in molecular conformations. However, they don't allow visualisation of the conformation changes the project pertains to. These are less bright than the probes previously mentioned, less photo-stable and absorb fewer photons. Some of these use secondary tags that are large compared to the DNA or protein sample[144]. At higher resolutions, nonlinear imaging contributions, diffraction of electrons and inelastic scattering affect the imaging results, therefore giving chance to oversimplified interpretations, even with the most advanced microscopes[145].

Super-resolution microscopy is a powerful method of single molecule imaging. FIONA-type single molecule imaging, sub-stochastic labelling approach, PALM, STORM and other photo-blinking reconstructive methods are some excellent examples in this category and have allowed

the understanding of the molecule behaviour and interactions in biological system. Yet, these pose several problems that make them not very suitable for the current research work. Firstly, some of these methods use long UV that may damage the structure of the DNA under study (dimerization of thymine bases) [148]. Secondly, some of these techniques are restricted by a high Abbe limit for probing molecular processes. Lastly, the blinking in these techniques (also in Quantum dots) limits their application in dynamic imaging of fast reactions completing within a few seconds[144]. Other imaging techniques like Near-field optical microscopy and Structured Illumination method, despite their lower spatial resolution (~20 nm lateral and ~5 nm axial) and speed of imaging than required, and associated photo-damage to the sample, have a promising future in single molecule imaging to understand the structure of molecules.

Electron Microscopy has the advantage of offering very high-resolution images and the recent remarkable advancements in the technique of cryo-EM have allowed atomic resolution imaging of protein structures[149–152]. But, there are chances that the exposure of the biological sample to large dose of high energy electron beam damages some biological samples [1,18]. Transmission Electron Microscopy (TEM) does not allow dynamic imaging of a single molecule. However, it can be used to image biological molecules in multiple structural states and an average image of each structural state can be obtained, while not offering an exact conformational information of a single molecule and possibly imposing structures that don't not actually exist[144]. The temporal resolution of EM is not very good and a large amount of data processing is involved that may be unnecessary (when simpler technique can be used instead). The requirement of imaging in vacuum and intense sample preparation disrupting the sample structure [153] restricts its and Scanning Electron Microscopy's (SEM) application to the current project as the sole imaging technique. However, the application of EM along with AFM could allow ground-breaking advancements in the field of single molecule imaging of biological interactions.

Biomolecular structure determination has been performed using X-ray crystallography, electron microscopy and NMR spectroscopy, but these techniques make use of ensemble averaging from images of fixed conformations and don't consider deviations from the average molecular structures[18,39]. Synthesis of crystalline samples is required which needs the sample in a large concentration and the process is long, unpredictable and tedious. These techniques can be used in collaboration with AFM to overcome the limitations of the respective techniques, which becomes especially important to study structural variations that can't be

studied by methods dependent on ensemble averaging, *e.g.* variations in groove depth and deformations of DNA, which in turn determine the protein binding specificity and gene expression [38].

In comparison to the above mentioned techniques, AFM allows nanometre-scale resolution, imaging in ambient buffer conditions, dynamic imaging and force measurements (if required)[39,153], without staining, probe labelling, recognition and localization of specific molecules [17,154]. There are no high energy electrons or UV light required that may damage the sample, the technique is not an averaging technique, does not require heavy equipment and allows high speed imaging. Also, the sample preparation method is comparatively straightforward, non-invasive and economical with a short incubation time. Considering all the above factors, AFM can be adopted as the imaging technique for dynamic single molecule imaging of DNA-protein interactions under native conditions. In combination with the other techniques mentioned here, AFM can be used to improve the understanding of various proteins' substrate recognition, binding and activity.

3.12. Limitations of AFM and the Future Prospects

AFM has numerous limitations despite being a useful tool for non-invasive biological imaging. Firstly, the resolution of the sample depends on the tip, hence the tip must be very sharp and suitably flexible. At the same time, the sharpness of tips has been known to compromise the structure of soft samples[16].

Secondly, large section of the sample actually remain unimaged as the tip scans only the surface and not the sides and bottom details[16]. Since attempts have been made to image biomolecular reactions inside the cell, AFM falls short in that direction as it is only able to image the surface [94]. It has been demonstrated that a tip can be inserted into the cell without damaging it [155,156] but whether that would translate to AFM inside the living cell in the future, is uncertain. Also, AFM studies isolated reactions, that are not always true for physiological conditions [157]. Hence, it is required that AFM be able to image reactions in the presence of complex reaction assemblies (like signal pathways). Additionally, in cases where biomolecular interactions are studied using force measurements, the understanding of forces in multiple directions needs to be taken into consideration[157], and not just along the z-axis that force measurements focus on.

Currently, AFM has been used for imaging the force of bonds and the elasticity of biomolecules[158], but not those two together, and the development of which would be an important step in the direction of imaging the conformation details of proteins and DNA during interaction. Hence, a combination of optical tweezers with HS-AFM could aid in this dilemma[94], especially to observe protein folding and unfolding, as these processes can be speed-regulated and are already studied by elastic network modelling [159,160].

Since AFM usually overestimates the *xy* plane size and underestimates the height of the sample, an indirect quantification of the size of molecules (protein or DNA molecules) is performed by measurement of volumes from the AFM images[161]. The resolution of AFM is currently insufficient to resolve the atomic conformation of proteins and DNA during their interactions. Additionally, the resolution is not suitable to detect ligands that play an important role in biochemical reactions as well as individual proteins in holoenzymes[94].

To achieve a greater speed of imaging, a smaller cantilever could be synthesized with a resonant frequency matching the required response time of AFM for imaging fast processes [94]. In this direction, the current OBD detector is replaced by a faster detector that also accommodates the smaller cantilever behaviour. Better speed and resolution can aid in the study of specificity of DNA binding for different types of protein, how the protein locates the DNA binding site in the long length of DNA and what conformation changes occur during catalysis. With such developments in imaging soft biological samples, AFM could allow ground-breaking advances in the understanding of biochemical processes.

Bibliography

1. Binnig GK. Atomic force microscope and method for imaging surfaces with atomic resolution. August 1986. https://patents.google.com/patent/US4724318.
2. Butt HJ, Cappella B, Kappl M. Force measurements with the atomic force microscope: Technique, interpretation and applications. Surf Sci Rep. 2005.
3. Lyubchenko YL, Shlyakhtenko LS, Gall AA. Atomic force microscopy imaging and probing of DNA, proteins, and protein-DNA complexes: Silatrane surface chemistry. Methods Mol Biol. 2009;543:337-351.
4. Bustamante C, Vesenka J, Tang CL, Rees W, Guthold M, Keller R. Circular DNA Molecules Imaged in Air by Scanning Force Microscopy1. Biochemistry. 1992;31:22-26.
5. Yang J, Takeyasu K, Shao Z. Atomic force microscopy of DNA molecules. FEBS Lett. 1992;301(2):173-176.
6. Vesenka J, Guthold M, Tang CL, Keller D, Delaine E, Bustamante C. Substrate preparation for reliable imaging of DNA molecules with the scanning force microscope. Ultramicroscopy. 1992:1243-1249.
7. Lyubchenko YL *et al*. Atomic Force Microscopy Imaging of Double Stranded DNA and RNA.2012;10(3):589-606.
8. Allen MJ *et al*. Atomic Force Microscope Measurements of Nucleosome Cores Assembled along Defined DNA Sequences. Biochemistry. 1993;32(33):8390-8396.
9. Hegner M, Wagner P, Semenza G. Immobilizing DNA on gold via thiol modification for atomic force microscopy imaging in buffer solutions. Fed Eur Biochem Soc. 1993;336(3):452-456.
10. Mou J, Czajkowsky DM, Zhang Y, Shao Z. High-resolution atomic-force microscopy of DNA: the pitch of the double helix. Mol Physiol. 1995;371(3):279-282.
11. Bustamante C, Erie DA, Keller D. Biochemical and structural applications of scanning force microscopy. Curr Opin Struct Biol. 1994;4(5):750-760.
12. Bustamante C, Rivetti C, Keller DJ. Scanning force microscopy under aqueous solutions. Curr Opin Struct Biol. 1997;7(5):709-716.
13. Lyubchenko YL, Gall AA, Shlyakhtenko LS. Atomic force microscopy of DNA and protein-DNA complexes using functionalized mica substrates. Methods Mol Biol. 2001;148:569-578.
14. Uchihashi T, Scheuring S. Applications of high-speed atomic force microscopy to real-time visualization of dynamic biomolecular processes. Biochim Biophys Acta - Gen Subj. 2018;1862(2):229-240.
15. Fukuda S *et al*. High-speed atomic force microscope combined with single-molecule fluorescence microscope. Rev Sci Instrum. 2013;84(7).
16. Kasas S *et al*. DNA-protein interactions explored by atomic force microscopy. Semin Cell Dev Biol. January 2017.
17. Suzuki Y, Endo M, Sugiyama H. Studying RNAP-promoter interactions using atomic force microscopy. Methods. 2015;86:4-9.
18. Ando T, Uchihashi T, Scheuring S. Filming biomolecular processes by high-speed atomic force microscopy. Chem Rev. 2014;114:3120−3188.
19. Hansma HG, Hoh JH. Biomolecular Imaging with Atomic Force Microscopy. Annu Rev Biophys Biomol Struct. 1994;23:115-139.
20. Bezanilla M, Drake B, Nudler E, Kashlev M, Hansma PK, Hansma HG. Motion and enzymatic degradation of DNA in the atomic force microscope. Biophys J. 1994;67(6):2454-2459.
21. Bruker. Imaging of the DNA Double Helix with PeakForce Tapping Mode Atomic Force Microscopy.
22. Gould SAC *et al*. The atomic force microscope: A tool for science and industry. Ultramicroscopy. 1990;33:93-98.
23. Ando T, Uchihashi T, Kodera N. High-Speed AFM and Applications to Biomolecular Systems. Annu Rev Biophys. 2013.

24. Hoof S, Gosvami NN, Hoogenboom BW. Enhanced quality factors and force sensitivity by attaching magnetic beads to cantilevers for atomic force microscopy in liquid. J Appl Phys. 2012;114324(2012).

25. Churnside AB, Perkins TT. Ultrastable atomic force microscopy: Improved force and positional stability. FEBS Lett. 2014;588:3621–3630.

26. Gaczynska M, Osmulski PA. AFM of biological complexes: What can we learn? Curr Opin Colloid Interface Sci. 2008;13:351–367.

27. Murugesapillai D, Bouaziz S, Maher LJ, Israeloff NE, Cameron CE, Williams MC. Accurate nanoscale flexibility measurement of DNA and DNA–protein complexes by atomic force microscopy in liquid. Nanoscale. 2017;9(31):11327-11337.

28. Hansma HG, Bezanilla M, Zenhausern F, Adrian M, Sinsheimer RL. Atomic force microscopy of DNA in aqueous solutions. Nucleic Acids Res. 1993;21(3):505-512.

29. Vesenka J, Manne S, Giberson R, Marsh T, Henderson E. Colloidal gold particles as an incompressible atomic force microscope imaging standard for assessing the compressibility of biomolecules. Biophys J. 1993;65(3):992-997.

30. Lai C-Y, Santos S, Chiesa M. General interpretation and theory of apparent height in dynamic atomic force microscopy. RSC Adv. 2015;5(97):80069-80075.

31. Voïtchovsky K. High-resolution AFM in liquid: What about the tip? Nanotechnology. 2015;26(10):6-9.

32. Alsteens D, Müller DJ, Dufrêne YF. Multiparametric Atomic Force Microscopy Imaging of Biomolecular and Cellular Systems. Acc Chem Res. 2017;50(4):924-931.

33. Akrami SMR, Nakayachi H, Watanabe-Nakayama T, Asakawa H, Fukuma T. Significant improvements in stability and reproducibility of atomic-scale atomic force microscopy in liquid. Nanotechnology. 2014;25(45).

34. Leung C *et al*. Atomic force microscopy with nanoscale cantilevers resolves different structural conformations of the DNA double helix. Nano Lett. 2012;12:3846−3850.

35. Ido S *et al*. Beyond the helix pitch: Direct visualization of native DNA in aqueous solution. ACS Nano. 2013;7(2):1817-1822.

36. Wang D, Stuart JD, Jones AA, Snow CD, Kipper MJ. Measuring interactions of DNA with nanoporous protein crystals by atomic force microscopy. Nanoscale. 2021;13:10871.

37. Ares P *et al*. High resolution atomic force microscopy of double-stranded RNA. Nanoscale. 2016;8(23):11818-11826.

38. Moreno-Herrero F, Colchero J, Baró AMM. DNA height in scanning force microscopy. Ultramicroscopy. 2003;96(2):167-174.

39. Pyne A, Thompson R, Leung C, Roy D, Hoogenboom BW. Single-molecule reconstruction of oligonucleotide secondary structure by atomic force microscopy. Small. 2014;10:3257–3261.

40. Hu J *et al*. Imaging of Single Extended DNA Molecules on Flat (Aminopropyl)triethoxysilane-Mica by Atomic Force Microscopy. acs J surfaces colloids. 1996;12:1697-1700.

41. Kasas S *et al*. Escherichia coli RNA Polymerase Activity Observed Using Atomic Force Microscopy. Biochemistry. 1997;36.

42. Suzuki Y, Shin M, Yoshida A, Yoshimura SH, Takeyasu K. Fast microscopical dissection of action scenes played by Escherichia coli RNA polymerase. FEBS Lett. 2012;586(19):3187-3192.

43. Bezanilla M, Manned S, Laney DE, Lyubchenko YL, Hansma HG. Adsorption of DNA to Mica, Silylated Mica, and Minerals: Characterization by Atomic Force Microscopy. Langmuir. 1995;11:655-659.

44. Luda S, Alexander AG, Yuri L, Shlyakhtenko LS, Gall AA, Lyubchenko YL. Mica Functionalization for Imaging of DNA and Protein-DNA Complexes with Atomic Force Microscopy. Methods Mol Biol. 2012;931.

45. Hansma HG, Laney DE. DNA Binding to Mica Correlates with Cationic Radius: Assay by Atomic Force Microscopy. Biophys J. 1996;70(4):1933-1939.

46. Oliveira Brett AM, Chiorcea Paquim A-M. DNA imaged on a HOPG electrode surface by AFM with controlled potential. Bioelectrochemistry. 2005;66(1-2):117-124.

47. Zhang X, Kumar S, Chen J, Teplyakov A V. Covalent attachment of shape-restricted DNA molecules on amine-functionalized Si(1 1 1) surface. Surf Sci. 2009;603(16):2445-2457.
48. James Cleaves H, Crapster-Pregont E, Jonsson CM, Jonsson CL, Sverjensky DA, Hazen RA. The adsorption of short single-stranded DNA oligomers to mineral surfaces. Chemosphere. 2011;83(11):1560-1567.
49. Pastré D *et al*. Anionic Polyelectrolyte Adsorption on Mica Mediated by Multivalent Cations: A Solution to DNA Imaging by Atomic Force Microscopy under High Ionic Strengths. Langmuir. 2006;22:6651-6660.
50. Nangreave J, Han D, Liu Y, Yan H. DNA origami: A history and current perspective. Curr Opin Chem Biol. 2010;14(5):608-615.
51. Endo M, Yang Y, Sugiyama H. DNA origami technology for biomaterials applications. Biomater Sci. 2013;1(4):347-360.
52. Sun X *et al*. Surface-Mediated DNA Self-Assembly. J Am Chem Soc. 2009;131:13248-13249.
53. Barinov NA, Prokhorov V V., Dubrovin E V., Klinov D V. AFM visualization at a single-molecule level of denaturated states of proteins on graphite. Colloids Surfaces B Biointerfaces. 2016;146:777–784.
54. Yamamoto D, Nagura N, Omote S, Taniguchi M, Ando T. Streptavidin 2D Crystal Substrates for Visualizing Biomolecular Processes by Atomic Force Microscopy. Biophys J. 2018;97(8):2358-2367.
55. Shlyakhtenko LS, Gall AA, Filonov A, Cerovac Z, Lushnikov A, Lyubchenko YL. Silatrane-based surface chemistry for immobilization of DNA, protein-DNA complexes and other biological materials. Ultramicroscopy. 2003;97(1-4):279-287.
56. Yamamoto D *et al*. High-Speed Atomic Force Microscopy Techniques for Observing Dynamic Biomolecular Processes. In: Methods in Enzymology. Vol 475. 1st ed. Elsevier Inc.; 2010:541-564.
57. Uchihashi T, Kodera N, Ando T. Guide to video recording of structure dynamics and dynamic processes of proteins by high-speed atomic force microscopy. Nat Protoc. 2012;7(6):1193-1206.
58. Liu Z, Li Z, Zhou H, Wei G, Song Y, Wang LI. Imaging DNA molecules on mica surface by atomic force microscopy in air and in liquid. Microsc Res Tech. 2005;66(4):179-185.
59. Lyubchenko YL. Preparation of DNA and nucleoprotein samples for AFM imaging. Micron. 2011;42(2):196-206.
60. Zheng J, Li Z, Wu A, Zhou H. AFM studies of DNA structures on mica in the presence of alkaline earth metal ions. Biophys Chem. 2003;104(1):37-43.
61. Podestà A, Imperadori L, Colnaghi W, Finzi L, Milani P, Dunlap D. Atomic force microscopy study of DNA deposited on poly L-ornithine-coated mica. J Microsc. 2004;215(3):236-240.
62. Buechner CN, Tessmer I. DNA substrate preparation for atomic force microscopy studies of protein-DNA interactions. J Mol Recognit. 2013;26(12):605-617.
63. At R *et al*. Sample preparation procedures for biological atomic force microscopy. J Microsc. 2005;218(3):199-207.
64. Nettikadan S, Tokumasu F, Takeyasu K. Quantitative analysis of the transcription factor AP2 binding to DNA by atomic force microscopy. Biochem Biophys Res Commun. 1996;226:645–649.
65. Lee AJ, Szymonik M, Hobbs JK, Wälti C. Tuning the translational freedom of DNA for high speed AFM. Nano Res. 2015;8(6):1811-1821.
66. Types of Mica | Axim Mica. https://aximmica.com/mica/types-of-mica/
67. Mica - freshly cleaved for AFM, SPM & FIB. https://www.agarscientific.com/mica.
68. Ando T. High-speed atomic force microscopy coming of age. Nanotechnology. 2012;23:062001.
69. John S *et al*. Direct Visualization of Dynamic Protein-DNA Interactions with a Dedicated Atomic Force Microscope. Biophys J. 1998;74(6):2840-2849.
70. Guo F *et al*. Imaging of DNA Molecules by Atomic Force Microscope. ieee. 2017:315-318.
71. Rivetti C, Guthold M, Bustamante C. Scanning force microscopy of DNA deposited onto mica: Equilibration versus Kinetic trapping studied by statistical polymer chain analysis. J Mol Biol. 1996;264(5):919-932.

72. Herrero-Gala E *et al*. Mechanical Identities of RNA and DNA Double Helices Unveiled at the Single-Molecule Level. J Am Chem Soc. 2013;135:122-131.

73. Coury JE, Anderson JR, McFail-Isom L, Williams LD, Bottomley LA. Scanning force microscopy of small ligand− nucleic acid complexes: tris (o-phenanthroline) ruthenium (II) as a test for a new assay. J Am Chem Soc. 1997;119(16):3792-3796.

74. Fang Y, Hoh JH, Spisz TS. Ethanol-induced structural transitions of DNA on mica. Nucleic Acids Res. 1999;27(8):1943-1949.

75. Rivetti C, Codeluppi S. Accurate length determination of DNA molecules visualized by atomic force microscopy: evidence for a partial B- to A-form transition on mica. Ultramicroscopy. 2001;87(1-2):55-66.

76. Hansma HG, Revenko I, Kim K, Laney DE. Atomic force microscopy of long and short double-stranded, single-stranded and triple-stranded nucleic acids. Nucleic Acids Res. 1996;24(4):713-720.

77. Japaridze A *et al*. Toward an Effective Control of DNA's Submolecular Conformation on a Surface. Macromolecules. 2016;49:643−652.

78. Bui V-C, Nguyen T-H. DNA aggregation induced by Mg2+ ions under different conditions. J Mol Recognit. 2018;31(9):e2721.

79. Heenan PR, Perkins TT. Imaging DNA Equilibrated onto Mica in Liquid Using Biochemically Relevant Deposition Conditions. ACS Nano. 2019;13(4):4220-4229.

80. Akpinar B, Haynes PJ, Bell NAW, Brunner K, Pyne ALB, Hoogenboom BW. PEGylated surfaces for the study of DNA-protein interactions by atomic force microscopy. Nanoscale. 2019;11:20072–20080.

81. Poly-L-Ornithine Solution (0.01%) | A-004-C. https://www.merckmillipore.com/GB/en/product/Poly-L-Ornithine-Solution-0.010-0,MM_NF-A-004-C?bd=1#overview.

82. Lyubchenko YL, Shlyakhtenko LS. AFM for analysis of structure and dynamics of DNA and protein-DNA complexes. Methods. 2009;47:206–213.

83. Bensimon A, Simon A, Chiffaudel A, Croquette V, Heslot F, Bensimon D. Alignment and sensitive detection of DNA by a moving interface. Science (80-). 1994;265(5181):2096-2098.

84. Kumar S *et al*. Direct Imaging of Protein Organization in an Intact Bacterial Organelle Using High-Resolution Atomic Force Microscopy. ACS Nano. 2017;11(1):126-133.

85. Uchihashi T, Kodera N, Ando T. Nanovisualization of proteins in action using high-speed AFM. In: Single-Molecule Studies of Proteins. Springer New York; 2013:119-147.

86. Ando T, Uchihashi T, Fukuma T. High-speed atomic force microscopy for nano-visualization of dynamic biomolecular processes. Prog Surf Sci. 2008;83:337-437.

87. Uchihashi T, Ganser C. Recent advances in bioimaging with high-speed atomic force microscopy. Biophys Rev. 2020;12:363–369.

88. Uchihashi T. Real-time nanoscale visualization of biological molecules at work with high-speed atomic force microscopy. 2019 IEEE Int Conf Manip Manuf Meas Nanoscale, 3M-NANO 2019 - Proc. 2019:253-256.

89. Ando T *et al*. High-speed atomic force microscopy for observing dynamic biomolecular processes. J Mol Recognit. 2007;20(6):448-458.

90. Ando T, Uchihashi T, Kodera N. High-Speed AFM and Applications to Biomolecular Systems AFM: atomic force microscopy. Annu Rev Biophys. 2013;42:393–414.

91. Uchihashi T, Watanabe H, Fukuda S, Shibata M, Ando T. Functional extension of high-speed AFM for wider biological applications. Ultramicroscopy. 2016;160:182-196.

92. Ando T. Directly watching biomolecules in action by high-speed atomic force microscopy. Biophys Rev. 2017;9:421-429.

93. Ando T, Kodera N, Takai E, Maruyama D, Saito K, Toda A. A high-speed atomic force microscope for studying biological macromolecules. PNAS. 2001;98:12468–12472.

94. Ando T. High-speed atomic force microscopy and its future prospects. Biophys Rev. 2018;10:285-292.

95. Lyubchenko YL, Shlyakhtenko LS. Imaging of DNA and Protein–DNA Complexes with Atomic Force Microscopy. Crit Rev Eukaryot Gene Expr. 2016;26:63-96.

96. Watanabe H. Development of Wide-area Tip-scanning High- speed Atomic Force Microscopy. IEEE. 2019:12-16.

97. Bennink ML, Nikova DN, Van der Werf KO, Greve J. Dynamic Imaging of Single DNA-Protein Interactions Using Atomic Force Microscopy. Vol 479. Elsevier; 2003:3-15.

98. Leung C *et al*. Real-time visualization of perforin nanopore assembly. Nat Nanotechnol. 2017;12:467-473.

99. Wessels WA, Broekmaat JJ, Beerends RJL, Koster G, Rijnders G. Fast and gentle side approach for atomic force microscopy. Rev Sci Instrum. 2013;84(12):123704.

100. Humphris ADLL, Miles MJ, Hobbs JK. A mechanical microscope: High-speed atomic force microscopy. Appl Phys Lett. 2005;86(3):1-3.

101. Lavrik OI. PARPs' impact on base excision DNA repair. DNA Repair (Amst). 2020;93.

102. Heenan PR, Wang X, Gooding AR, Cech TR, Perkins TT. Bending and looping of long DNA by Polycomb repressive complex 2 revealed by AFM imaging in liquid. Nucleic Acids Res. 2020;48(6):2969-2981.

103. Shibata M *et al*. Real-space and real-Time dynamics of CRISPR-Cas9 visualized by high-speed atomic force microscopy. Nat Commun. 2017;8(1):1-9.

104. Kodera N *et al*. Structural and dynamics analysis of intrinsically disordered proteins by high-speed atomic force microscopy. Nat Nanotechnol. 2021;16:181-189.

105. Chang PI-T, Yu Y-J. Automatic DNA Geometric Analysis from Atomic Force Microscopy Images. 2019.

106. Chang PI, Hsaio M-C. Resolution-Free Accurate DNA Contour Length Estimation from Atomic Force Microscopy Images. hindawi. 2019.

107. Gould SAC *et al*. From atoms to integrated circuit chips, blood cells, and bacteria with the atomic force microscope. J Vac Sci Technol B Microelectron Nanom Struct. 1990;8:369-373.

108. Butt HJ, Wolff EK, Dixon NB, Peterson CM, Hansma PK. Imaging cells with the atomic force microscope. J Struct Biol. 1990;165:54-61.

109. Putman AJ *et al*. Atomic force microscope with integrated optical microscope for biological applications. Rev Sci Instrum. 1992;63:1914-1917.

110. Sokolov I, Dokukin M. Mechanics of biological cells studied with atomic force microscopy. Microsc Microanal. 2014;20(3):2076-2077.

111. Drake B *et al*. Imaging crystals, polymers, and processes in water with the atomic force microscope. Science (80-). 1989;243(4898):1586-1589.

112. Weisenhorn AL *et al*. Immobilized proteins in buffer imaged at molecular resolution by atomic force microscopy. Biophys J. 1990;58(5):1251-1258.

113. Butt HJ, Downing KH, Hansma PK. Imaging the membrane protein bacteriorhodopsin with the atomic force microscope. Biophys J. 1990;58(6):1473-1480.

114. Hoh JH, Lal R, John S, Revel J, Arnsdorf MF. Atomic force microscopy and dissection of gap junctions. Science (80-). 1991;253:1405-1408.

115. Hoh JH, Hansma PK. Atomic force microscopy for high-resolution imaging in cell biology. Trends Cell Biol. 1992;2:208-213.

116. Schneider SW, Larmer J, Henderson RM, Oberleithner H. Molecular weights of individual proteins correlate with molecular volumes measured by atomic force microscopy. Eur J Physiol. 1998;435:362–367.

117. Meyer E, Howald L, Overney RM, *et al*. Molecular-resolution images of Langmuir–Blodgett films using atomic force microscopy. Nature. 1991;349:398-400.

118. Hansma HGG, Laney DEE, Bezanilla M, Sinsheimer RLL, Hansma PKK. Applications for atomic force microscopy of DNA. Biophys J. 1995;68(5):1672-1677.

119. Jiang Y, Ke C, Mieczkowski PA, Marszalek PE. Detecting ultraviolet damage in single DNA molecules by atomic force microscopy. Biophys J. 2007;93(5):1758-1767.

120. Samorí B, Siligardi G, Quagliariello C, Weisenhorn AL, Vesenka J, Bustamante CJ. Chirality of DNA supercoiling assigned by scanning force microscopy. 1993;90(8):3598-3601.

121. Guthold M *et al*. Following the assembly of RNA polymerase-DNA complexes in aqueous solutions with the scanning force microscope. Proc Natl Acad Sci. 1994;91(26):12927-12931.

122. Radmacher M, Fritz M, Hansma HG, Hansma PK. Direct observation of enzyme activity with the atomic force microscope. Science (80-). 1994;265(5178):1577-1579.

123. Edstrom RD, Meinke MH, Yang XR, Yang R, Elings V, Evans DF. Direct visualization of phosphorylase-phosphorylase kinase complexes by scanning tunneling and atomic force microscopy. Biophys J. 1990;58(6):1437-1448.

124. Hansma HG *et al*. Reproducible Imaging and Dissection of Plasmid DNA Under Liquid with the Atomic Force Microscope. Science (80-). 1992;256:1180-1184.

125. Henderson E. Imaging and nanodissection of individual supercoiled plasmids by atomic force microscopy. Nucleic Acids Res. 20(3):445-447.

126. Murray MN *et al*. Atomic force microscopy of biochemically tagged DNA. Proc Natl Acad Sci USA. 1993;90:3811-3814.

127. Rees WA, Keller RW, Vesenka JP, Yang C, Bustamante C. Scanning force microscopy imaging of transcription complexes: evidence for DNA bending in open promoter and elongation complexes. Science (80-). 1993;260:1646-1649.

128. Zenhausern F, Adrian M, Heggeler-Bordier B Ten, Eng LM, Descouts P. DNA and RNA polymerase/DNA complex imaged by scanning force microscopy: Influence of molecular-scale friction. Scanning. 1992;14(4):212-217.

129. Niu L, Shaiu W, Vesenka J, Larson DD, Henderson E. Atomic force microscopy of DNA-colloidal gold and DNA-protein complexes. Adv DNA Seq Technol. 1993;1891:71-77.

130. Rees W, Keller R, Vesenka J, Yang G, Bustamante C. Evidence of DNA bending in transcription complexes imaged by scanning force microscopy. Science (80-). 1993;260(5114):1646-1649.

131. Khalid AAM *et al*. Atomic force microscopy investigation of the interactions between the MCM helicase and DNA. Materials (Basel). 2021;14(3):1-15.

132. Zheng P, Li H, Koussa MA, Sotomayor M, Wong WP, Corey DP. Probing the Unbinding Kinetics of DNA-H-NS-DNA Protein Complexes by a High-Speed and High-Throughput Single-Molecule Pulling Assay using Atomic Force Microscopy. Biophysj. 2014;106(2):386a.

133. Puppels GJ, Putman CAJ, Grooth BG de, Greve J. Raman microspectroscopy and atomic force microscopy of chromosomal banding patterns. Laser Study Macrosc Biosyst. 1993;1922:145-155.

134. Fang Y *et al*. Solid-state DNA sizing by atomic force microscopy. Anal Chem. 1998;70(10):2123-2129.

135. Cary RB, Peterson SR, Wang J, Bear DG, Bradbury EM, Chen DJ. DNA looping by Ku and the DNA-dependent protein kinase. Proc Natl Acad Sci. 1997;94(9):4267-4272.

136. Erie DA, Yang G, Schultz HCH, Bustamante C, Erie DA. DNA bending by Cro protein in specific and nonspecific complexes: implications for protein site recognition and specificity. Science (80-). 1994;266(5190):1562-1566.

137. Han W *et al*. Strained DNA is kinked by low concentrations of Zn^{2+}. Proc Natl Acad Sci. 1997;94(20):10565-10570. www.pnas.org.

138. Rippe K, Guthold M, von Hippel PH, Bustamante C. Transcriptional activation via DNA-looping: visualization of intermediates in the activation pathway of E. coli RNA polymerase· σ54 holoenzyme by scanning force microscopy. J Mol Biol. 1997;270(2):125-138.

139. Guthold M *et al*. Direct observation of one-dimensional diffusion and transcription by Escherichia coli RNA polymerase. Biophys J. 1999;77:2284–2294.

140. Hoitsma NM *et al*. AP-endonuclease 1 sculpts DNA through an anchoring tyrosine residue on the DNA intercalating loop. Nucleic Acids Res. 2020;48(13):7345-7355.

141. Zhang S, Aslan H, Besenbacher F, Dong M. Quantitative biomolecular imaging by dynamic nanomechanical mapping. Chem Soc Rev. 2014;43(21):7412-7429.

142. Senapati S, Lindsay S. Recent Progress in Molecular Recognition Imaging Using Atomic Force Microscopy. Acc Chem Res. 2016;49(3):503-510.

143. Ueno H *et al*. Simple dark-field microscopy with nanometer spatial precision and microsecond temporal resolution. Biophys J. 2010;98:2014–2023.

144. Leake MC. Single-Molecule Cellular Biophysics. Cambridge University Press; 2013.

145. Marks L, Uhlemann S, Müller H, Hartel P, Zach J, Haider M. What Are the Resolution Limits in Electron Microscopes? Phys Rev Lett. 2013;6:2013.

146. Gumpp H, Puchner EM, Zimmermann JL, Gerland U, Gaub HE, Blank K. Triggering enzymatic activity with force. Nano Lett. 2009;9(9):3290-3295.

147. Miller H, Zhou Z, Shepherd J, Wollman AJM, Leake MC. Single-molecule techniques in biophysics: a review of the progress in methods and applications. Rep Prog Phys. 2018;81.

148. Watson JD *et al*. Molecular Biology of the Gene. 7th ed. Cold Spring Harbour Laboratory Press; 2013.

149. Tan YZ, Carragher B. Seeing Atoms: Single-Particle Cryo-EM Breaks the Atomic Barrier. Mol Cell. 2020;80(6):938-939.

150. Nakane T *et al*. Single-particle cryo-EM at atomic resolution. Nature. 2020;587:152-156.

151. Yip KM, Fischer N, Paknia E, Chari A, Stark H. Atomic-resolution protein structure determination by cryo-EM. Nature. 2020;587(7832):157-161.

152. Bai X chen. Seeing Atoms by Single-Particle Cryo-EM. Trends Biochem Sci. 2021;46(4):253-254.

153. Robson A-L *et al*. Advantages and Limitations of Current Imaging Techniques for Characterizing Liposome Morphology. Front Pharmacol. 2018;9:1-8.

154. Endo M, Sugiyama H. Single-molecule imaging of dynamic motions of biomolecules in DNA origami nanostructures using high-speed atomic force microscopy. Acc Chem Res. 2014;47(6):1645-1653.

155. Obataya I, Nakamura C, Han S, Nakamura N, Miyake J. Mechanical sensing of the penetration of various nanoneedles into a living cell using atomic force microscopy. Biosens Bioelectron. 2005;20(8):1652-1655.

156. Guillaume-Gentil O, Potthoff E, Ossola D, Franz CM, Zambelli T, Vorholt JA. Force-controlled manipulation of single cells: from AFM to FluidFM. Trends Biotechnol. 2014;32(7):381-388.

157. Willemsen OH, Snel MMEE, Cambi A, Greve J, De Grooth BG, Figdor CG. Biomolecular interactions measured by atomic force microscopy. Biophys J. 2000;79(6):3267-3281.

158. Dufrêne YF, Evans E, Engel A, Helenius J, Gaub HE, Müller DJ. Five challenges to bringing single-molecule force spectroscopy into living cells. Nat Methods. 2011;8(2):123-127.

159. Togashi Y, Mikhailov AS. Nonlinear relaxation dynamics in elastic networks and design principles of molecular machines. Proc Natl Acad Sci. 2007;104(21):8697-8702.

160. Düttmann M, Togashi Y, Yanagida T, Mikhailov AS. Myosin-V as a mechanical sensor: an elastic network study. Biophys J. 2012;102(3):542-551.

161. Fuentes-Perez ME, Dillingham MS, Moreno-Herrero F. AFM volumetric methods for the characterization of proteins and nucleic acids. Methods. 2013;60(2):113-121.

Chapter 4

Materials and Methods

4.1. Materials

4.1.1. Cell cultures:

E.coli (XL1 blue and K12 that produce *Thermus aquaticus* DNA Polymerase I FEN domain), M13 phage culture (to produce ssDNA for synthesizing flap DNA)

4.1.2. *Plasmids: parS* plasmid (to produce *parS* DNA) and pJONEX with tFENΔ19 insert (to produce overhang DNA by PCR)

4.1.3. *Media:* (for growing bacteria and phage cultures) LB media, 2x YT media, agar

4.1.4. *Reagents for DNA and Protein Production and Purification:*

Solution	Contents
2x SDS-PAGE loading dye (to load the protein)	2% SDS, 20% glycerol, 125 mM TRIS, 0.05% (w/v) bromophenol blue, pH 6.9
2x DNA Native PAGE loading dye (to load DNA)	(20% glycerol, 125 mM TRIS, 0.05% bromophenol blue, pH 6.9)
DNA Native PAGE staining solution (to stain the gel for imaging)	2 mg/mL brilliant blue, 40 % methanol, 10% acetic acid
DNA Native PAGE de-staining solution (to de-stain the gel for imaging)	40% methanol, 10% acetic acid
(NEB) DNA loading dye (6x)	
midoori green staining dye	Used in place of Ethidium bromide to make the DNA visible in UV light

Table 4.1: Solutions for gel electrophoresis.

Reagents	Amount
acrylamide (37.5:1 acrylamide: bisacrylamide) (for SDS-PAGE)	30% (w/v)
APS (for SDS-PAGE)	10%
Deoxycholic acid solution (to prevent frothing of cell extract during protein purification)	10μL/mL
DTT (to stabilize the protein in solutions)	100 mM
EDTA (chelating agent to bind metal ions)	250 mM, pH 7.8
KCl (source of K^+ for buffer solutions)	1 M
$MgCl_2$ (source of Mg^{2+} for buffer solutions)	100 mM
NaCl (source of Na^+ for buffer solutions)	5 M
PEG 4000 (for phosphorylation of oligonucleotides)	50%
PEG 6000 (for precipitating phage particles)	20%
SDS (for SDS-PAGE)	10%
Sigma Type XIV DNA (for UV nuclease assay)	2 mg/mL
sodium azide (to flash freeze protein sample in liquid N_2)	0.1%
Tetracycline (antibiotic to grow bacteria culture)	1.5 mg/mL in ethanol
ammonium sulphate (s)(to salt out protein during purification)	
ethanol (for DNA precipitation)	
glycerol (in protein solutions for freeze storage)	
liquid N_2 (for freeze storage samples)	
PEI (for protein purification)	
TEMED (for SDS PAGE)	

Table 4.2: Stock Solutions for DNA and protein synthesis and purification.

4.1.5. Enzymes:

(New England Biolabs-NEB) DNA ligase (for ligation), *Thermus aquaticus* DNA Polymerase I (for PCR), Klenow fragment (for production of flap DNA from M13 DNA), lysozyme (to lyse the cells during protein extraction), Polynucleotide kinase (PNK) (to phosphorylate the oligonucleotides for annealing and ligation)

4.1.6. DNA Oligonucleotides (for flap and nicked DNA assembly) *and dNTPs*

4.1.7. PCR primers (see appendix 6 for details)

4.1.8. Chromatography columns: SP, Q and Heparin columns for protein purification

4.1.9. Kits for PCR and DNA purification:

4.1.9.1. Monarch PCR clean-up kit (NEB catalogue number T1030),

4.1.9.2. Monarch DNA gel extraction kit (NEB catalogue number T1020)

4.1.9.3. Monarch Plasmid Mini-prep Kit (NEB catalogue number T1010)

4.1.9.4. KAPA HiFi PCR Kit (KR0368 – v13.19)

4.1.9.5. PCR Biosystems PCR kit

4.1.9.6. Fisher Scientific Gene JET PCR purification kit

4.1.10. Buffers:

Stock solutions: HEPES (1 M, pH 8), Phosphate buffer (500 mM, pH 8), Tris-Cl (500 mM, pH 8), 250 mM potassium glycinate buffer (pH 9.3),

Solution	Contents
DNA ligation buffer	with Mg^{2+} and ATP
TE buffer	10 mM TRIS, 1 mM EDTA, pH 8
DNA buffer	20 mM NaCl, 20 mM TRIS-Cl, 1 mM EDTA, pH 8
NTE buffer	100 mM NaCl, 10 mM TRIS, 1 mM EDTA, pH 8
5x resolving buffer	500 mM TRIS, 500 mM bicine, 0.2 mM EDTA, pH 8.3
2x stacking buffer	250 mM TRIS, pH 6.9
1x SDS-PAGE running buffer or 1x DNA Native PAGE running buffer	100 mM TRIS, 100 mM bicine, 2 mM EDTA, 1% SDS
TBG buffer	100 mM TRIS, 100 mM bicine, 10% glycerol
TAE buffer	40 mM TRIS-acetate, 1 mM EDTA
TBE buffer	108 g TRIS, 55 g boric acid, 40 mL 500 mM EDTA, (1 L)
Zymogram reaction buffer	100 mM TRIS, 100 mM Bicine, 10% glycerol, 50 mM NaCl, 1 mM DTT, 10 mM $MgCl_2$, 10 mM KCl

Table 4.3: Buffer Compositions.

4.1.11. Reagents for Sample Immobilization for AFM:

Polyornithine (PLO) (1 mg/mL stock), mica discs, $CaCl_2$ (10 mM stock), KCl (20 mM stock), $MgCl_2$ (100 mM stock), $NiCl_2$ (100 mM stock), HEPES (500 mM stock, pH 7.8), HPLC water.

4.1.12. AFM Cantilever Tips:

(Already discussed in chapter 2) Bruker TESPA-V2, Fastscan A, Fastscan D, Fastscan D-SS, AC-40 (biolever mini) (appendix 2), Nunano SCOUT 350 RAI (see appendix 3 for details)

4.1.13. AFM Instruments:

Dimension 3100 Scanning Force Microscopes with Nanoscope IIIa and IV control units

Bruker Multimode with Nanoscope V controller in homebuilt vibration and acoustic vibration isolation systems

Bruker Fastscan Bio

4.2. Methods:

4.2.1. Production and Purification of 100 bp Flap and Nicked DNA:

The flap DNA structure is not a form of DNA usually found in cells of organisms. It is formed in the cell for a very short period of time during the process of cell division and thus cannot be extracted. Hence, to study the interactions with DNA Polymerase I, it is required to synthesize the DNA *in vitro.*

Synthetic DNA oligonucleotides (see appendix 4 for details of the DNA sequence) varying between 28 to 62 nt were designed (Eurofins) to assemble and form flap DNA and nicked DNA (negative control for the reaction). The oligos were diluted in required volume of TE buffer, (as described in the Oligonucleotide Synthesis Report of the product (see appendix 5)) to obtain a concentration of about 100 µM. The concentration of oligonucleotides was determined using nanodrop spectrophotometer, and their molarity was calculated using the molecular weight mentioned in the oligonucleotide synthesis report.

$$Conc.(Molarity) = \frac{conc.(g/L)}{molecular\ weight(g/mol)} \qquad \textit{(equation 1)}$$

To prepare flap DNA, 10µL each of the oligonucleotides 1,7,4-6 (see appendix 4 for details of the DNA sequence) were added to prepare a 100 µL solution with 50 mM HEPES, 50 mM NaCl and 1 mM EDTA in miliQ water. To prepare nicked DNA, 10µL each of the oligonucleotides 1-6 were added to prepare a 100 µL solution with 50 mM HEPES, 50 mM NaCl and 1 mM EDTA in miliQ water. Since the DNA oligos were already phosphorylated at the 5' ends, no phosphorylation step was required. The above solutions were incubated at 90°C for 5 minutes and allowed to cool to room temperature gradually to allow them to anneal.

ssDNA fragments form H-bonds with complementary strands. This process is called annealing. The annealed DNA segments have a single phosphodiester bond break on the strands which need to be sealed by ligases (see figure 4.1). These enzymes catalyse the formation of phosphodiester bonds between 3'-OH and 5'-PO_4^{3-} ends of DNA or RNA using ATP or NAD^+ as cofactors, depending on the type of ligase [1(chap2)]. If the 5' ends of the DNA don't have PO_4^{3-} attached to them, they are phosphorylated by polynucleotide kinase (PNK) that add PO_4^{3-} onto free 5' terminals [2(chap4)].

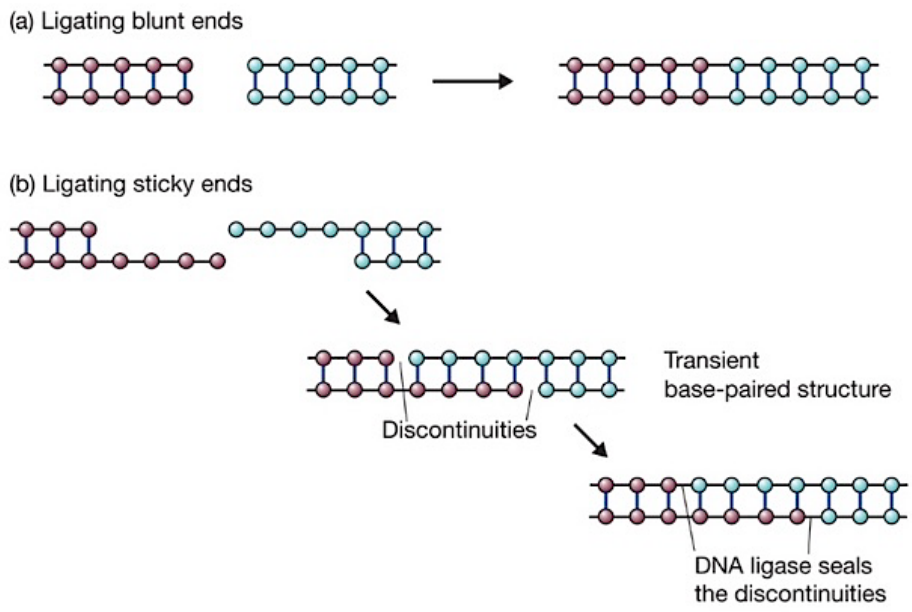


Figure 4.1: Types of ligations of DNA with (a) blunt ends and (b) sticky ends [2(chap4)].

For ligation, the amount of ligase units required was calculated: 1 mL of the enzyme contains 400,000 units, each of which is capable of ligating 50% 5' PO_4^{3-} sites at 0.12 µM concentration in a volume of 20 µL in 30 minutes at 16°C. Hence 4 µL ligase was added into each of the tubes of flap and nicked DNA mixture, along with 20 µL ligase buffer and 76 µL water to dilute the mixture. The reaction mixture was incubated at room temperature for 30 minutes followed by 65°C for 20 minutes to denature the ligase. The DNA samples could be stored at -20°C.

The DNA molecules were then purified by gel electrophoresis. Linear DNA molecules separate *w.r.t.* their size when exposed to an electric field through a gel matrix. DNA being negatively charged, migrates to the anode to a distance inversely proportional to their lengths (see figure 4.2). Two kinds of matrix can be used: agarose and polyacrylamide. Agarose gel electrophoresis can be used to separate DNA fragments up to hundreds of kilobases. Polyacrylamide Gel Electrophoresis (PAGE) without SDS (called Native PAGE) allows high resolution separation of DNA over a narrow size range. In this technique, the cross-linking of flap DNA that might happen during agarose gel electrophoresis is avoided [3(chap7)].

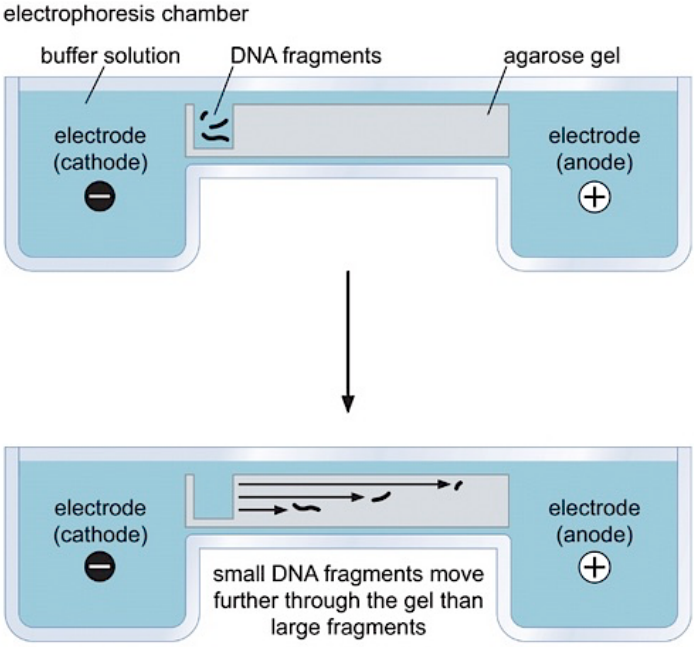


Figure 4.2: DNA separation by gel electrophoresis reproduced from Watson *et al.*, 2013, chap. 7[3].

The DNA fragments prepared above had a lot of oligonucleotides that had not been ligated or annealed. Hence, the DNA fragments of desired length (100 bp) were separated from the shorter oligos (28-62 nt) by DNA Native PAGE followed by gel elution and dissolution of the required gel slice in HEPES buffer (25mM HEPES, 100mM NaCl, pH 8). The standard agarose gel purification kits were not used because the flap DNA cross-linked in the agarose gel. Hence native-PAGE, which has a limited purity with gel purification kits, was performed. The concentration of DNA was determined using a spectrophotometer.

4.2.2. Production and Purification of Overhang DNA Fragments:

PCR or Polymerase Chain Reaction is a technique to form multiple copies of any region of DNA molecule if the sequence at the ends of the targeted region is known. Two short oligonucleotides, usually about 17 bases, called primers, are annealed to the ends of the target DNA sequence (called the DNA template) due to their complementarity to the template sequence. The enzyme *Thermus aquaticus* DNA Polymerase I is used to extend the DNA primers from their 3' ends in the presence of dNTPs and Mg^{2+}. The PCR reaction consists of about 25-30 repeats of the cycle of denaturation of the template DNA strand at ~94°C, followed by cooling to 50-60°C to allow the primers to anneal, then heating to 74°C to allow the polymerisation reaction by *Taq* polymerase (figure 4.3). The duration and temperatures of each the PCR steps varies depending on the sequence of the template DNA and the nature of the DNA primers, *i.e.* length of the primers, amount of complementarity to the DNA template, denaturation temperature for the template DNA. The PCR products can be analysed on an agarose gel [2(chap9)].

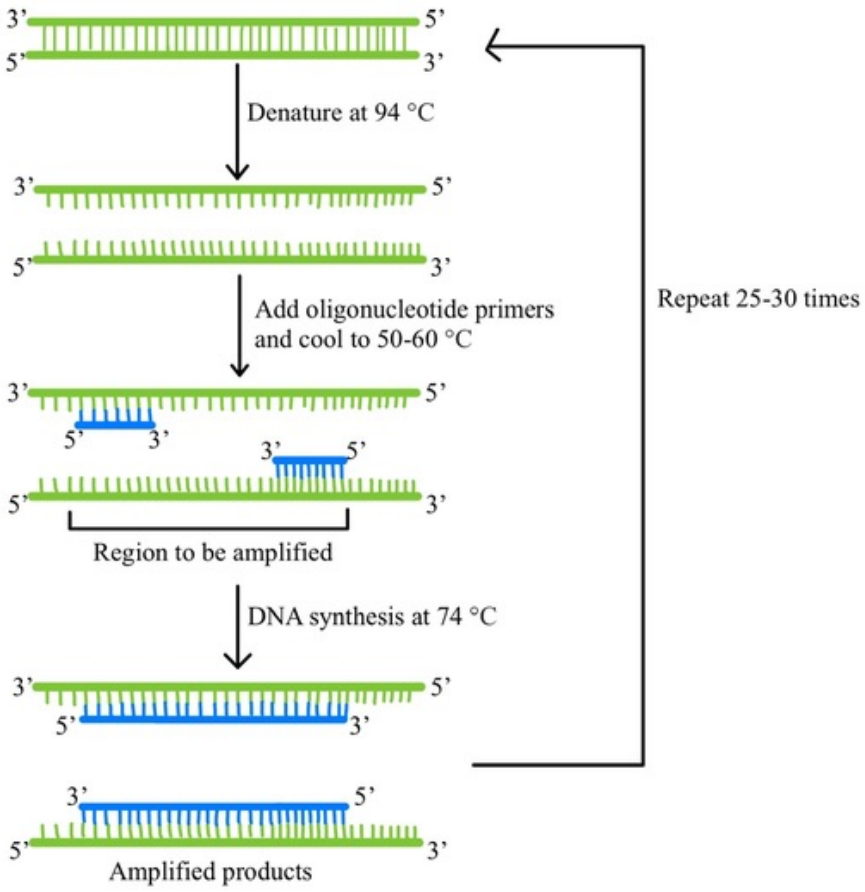


Figure 4.3: Schematic of PCR, reproduced from T. A. Brown (Chapter 9, 6th edition)[2].

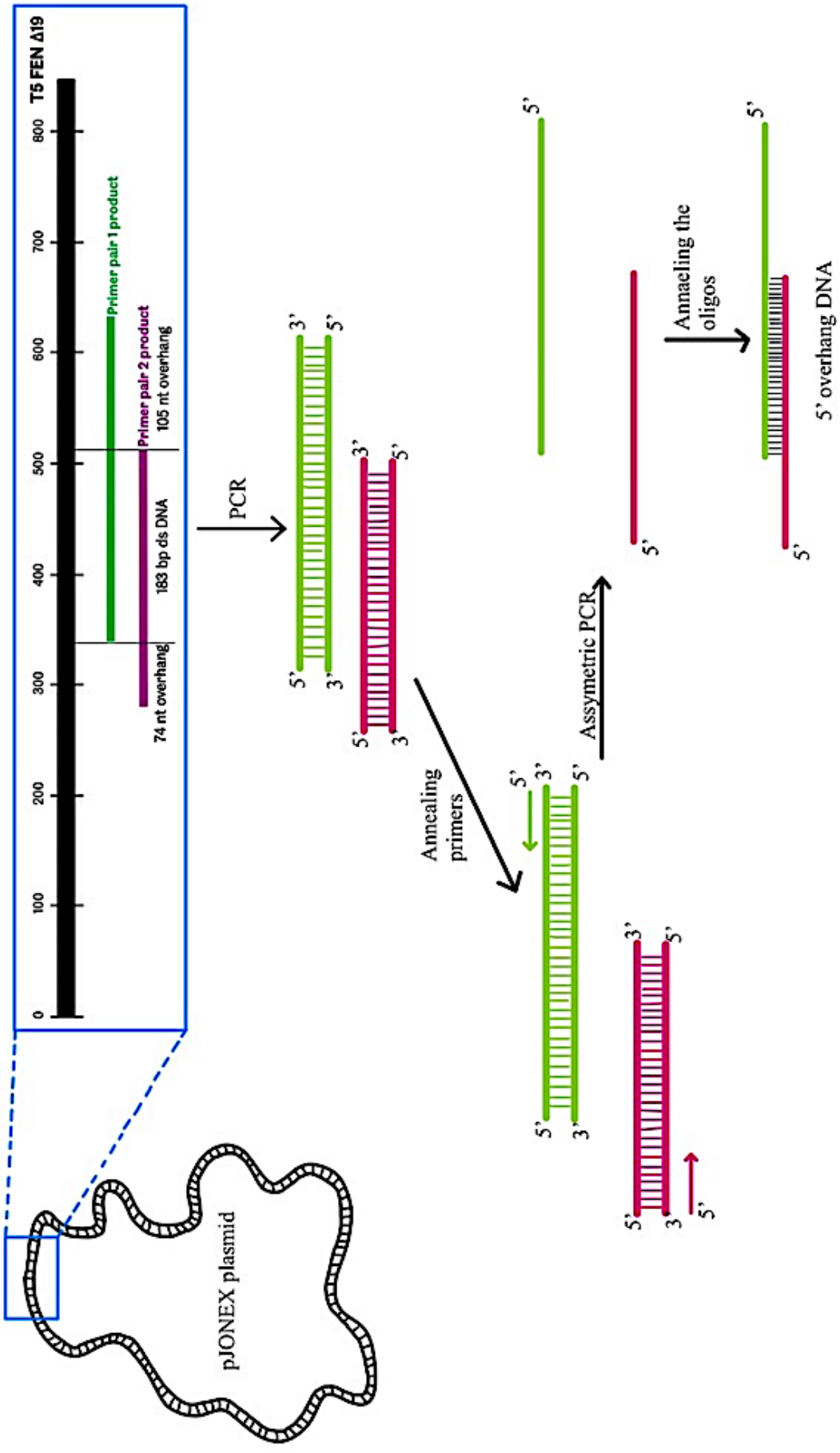


Figure 4.4: Schematic of preparation of overhang DNA using PCR .

The principle behind the construction of overhang DNA was the performance of two PCRs (figure 4.4). The first one aimed to produce 2 PCR products P1 and P2 of lengths 288 bp and 257 bp respectively that overlap for 183 bp, with 74 bp and 105 bp dsDNA left on either side (see appendix 6 for primer details). The template used for this set of PCR was plasmid pJONEX with tFENΔ19 insert. The second PCR stage consisted of 2 separate sets of asymmetric PCRs with P1 and P2 as respective templates and one primer each such that the ssDNA products formed as a result would anneal to leave 5' overhangs of 74 nt and 105 nt on the ends.

Two sets of primers were designed and ordered from Eurofins. Following the 1st PCR, the products were analysed on an agarose gel and purified using Monarch PCR clean-up kit (NEB catalogue number T1030). Asymmetric PCR was then performed with only 1 primer each for the templates P1 and P2 respectively, followed by analysis on an agarose gel and purification using Monarch PCR clean-up kit (T1030). The primers used were carefully selected such that the ssDNA annealed to form 5' overhang and not 3'. The 2 ssDNA fragments thus obtained were allowed to anneal by heating at 95°C for 1 minute, mixing them by gentle pipetting followed by gradual cooling to room temperature. The DNA mixture was then analysed on an agarose gel and purified using Monarch DNA gel extraction kit (NEB catalogue number T1020). The concentration of DNA was determined by a spectrophotometer.

4.2.3. Production and Purification of Flap DNA from M13 DNA Template:

Bacteriophage M13 is a filamentous bacteriophage that consists of a single-stranded circular DNA genome about 6400 nt long (figure 4.5). Filamentous bacteriophages are used in the field of molecular biology to provide single-stranded copies of DNA segments cloned in vectors. These ssDNA fragments can be used as templates for site-directed mutagenesis, DNA sequencing, construction of subtractive cDNA libraries, and synthesis of probes specific to DNA strands [1(chap3)]. In this case, the M13 DNA was used as a template to synthesize dsDNA with a ssDNA flap.

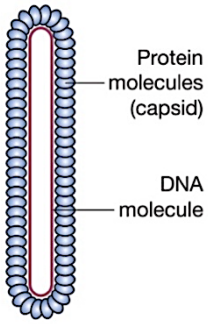


Figure 4.5: Filamentous phage structure [2(chaps2-3)].

During M13 infection into *E.coli* cells, the phage's ssDNA molecule acts as a template for synthesis of a complementary strand to form double-stranded DNA, called the Replicative Form (RF) (figure 4.6). This form behaves like a normal plasmid with a high copy number and can be easily prepared by lysing the infected *E.coli* cell. The ssDNA though, is contained in the M13 phage particles that are secreted from the infected bacteria cells. A large titre value of phage particles can be prepared from a very small volumes of cell cultures (like 5 mL). Since the cells don't need to be lysed, there are no steps required to remove the cell debris. Hence, M13 ssDNA can be prepared by growing small volume of infected *E.coli* cells, centrifugation of bacteria pellet, precipitation of phage particles by PEG, removal of phage protein coat by phenol extraction and lastly ethanol precipitation of DNA (figure 4.7). This process, though straightforward, requires a high extracellular phage titre *w.r.t.* the infected E.coli cells [2(chaps2-3)].

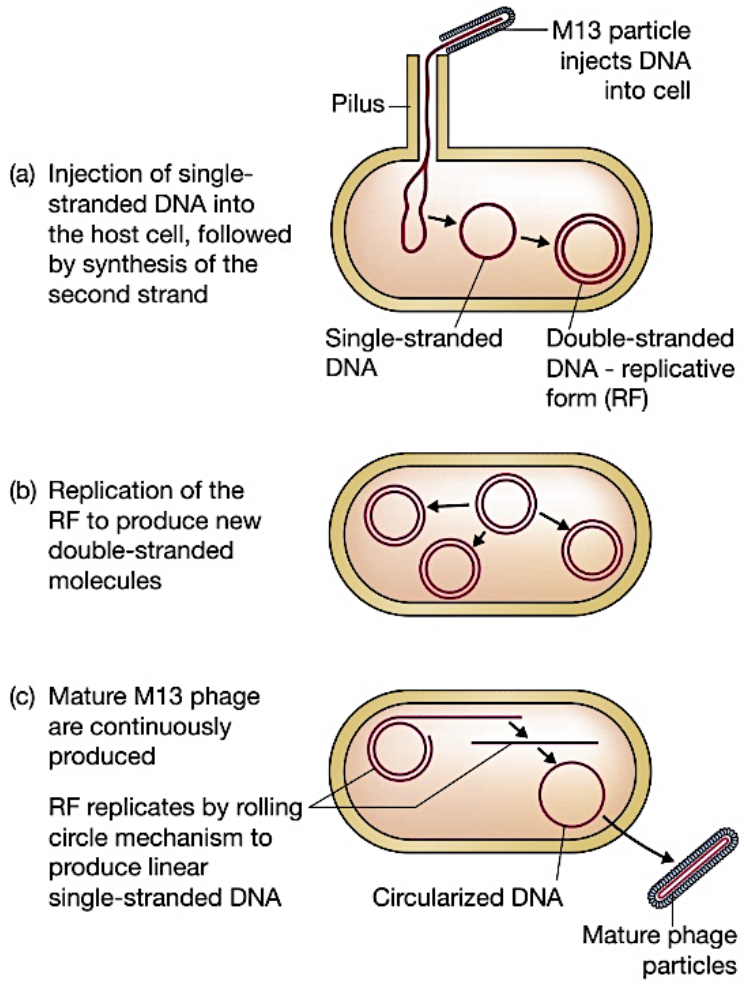


Figure 4.6: Infection cycle of M13 [2(chaps2-3)].

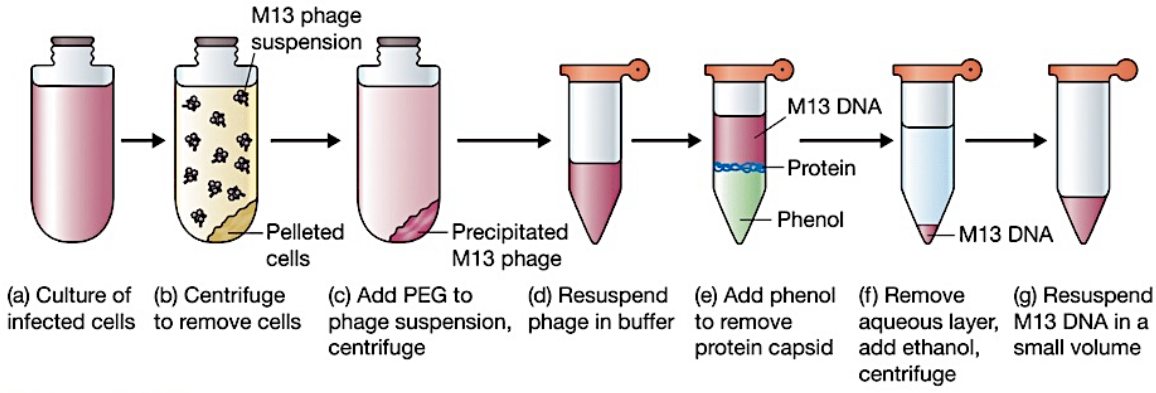


Figure 4.7: Preparation of ss DNA from M13 infected bacterial culture [2(chaps2-3)].

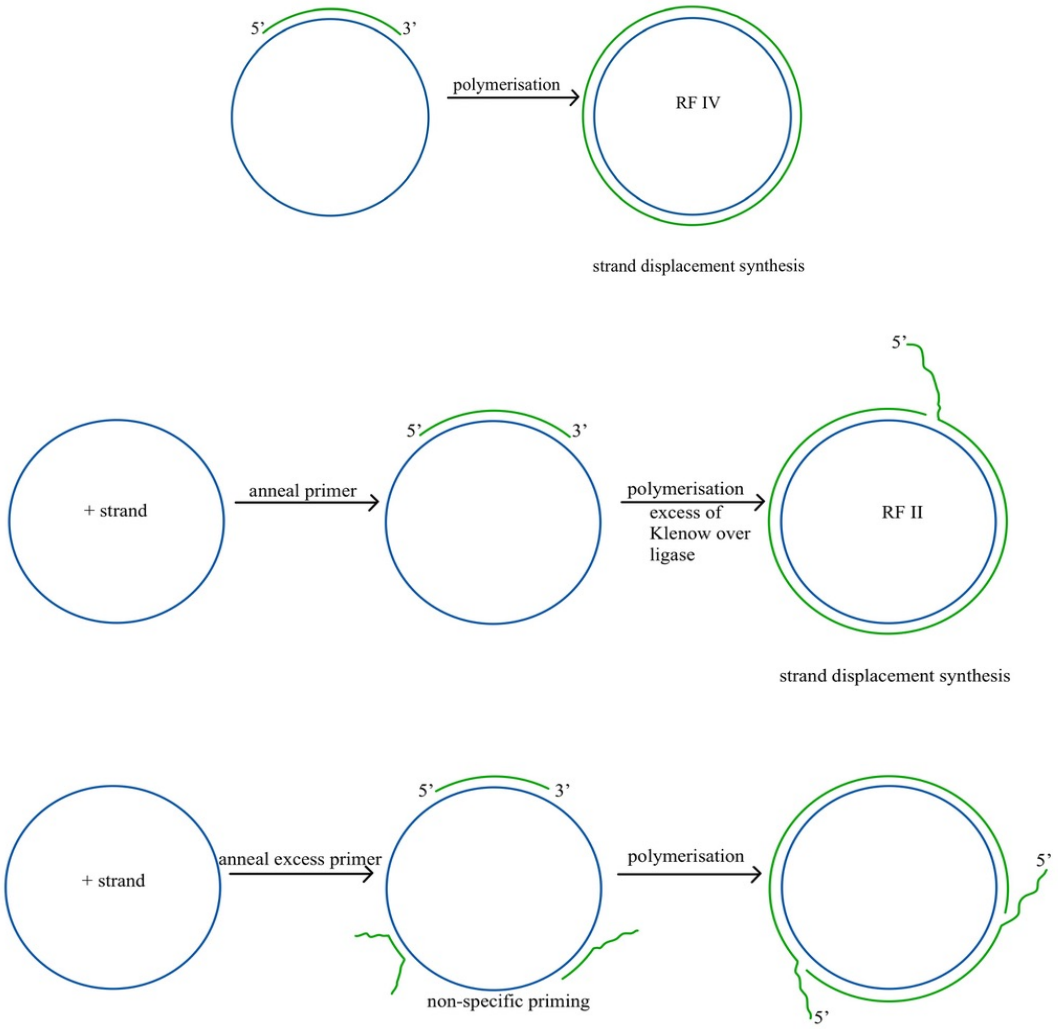


Figure 4.8: Formation of RF IV and RF II from ssDNA in M13 (reproduced from Sayers and Eckstein, 1991) [4].

The ssDNA can be converted to dsDNA (RF IV) by annealing the primer complementary to the target sequence and polymerisation in the presence of Klenow fragment and ligase. But, in the scenario of excess of Klenow *w.r.t.* ligase, the RF II form may be formed (see figure 4.8). As the Klenow has no 5' exonuclease activity, the circular dsDNA has a 5' flap. Furthermore, in the presence of excess primers, the RF II may have multiple polymerisation sites leading to multiple 5' flaps [4]. This RF II form DNA with 5' flap is suitable to study the mechanism of action of FEN using AFM. The process has been summarized in figure 4.9.

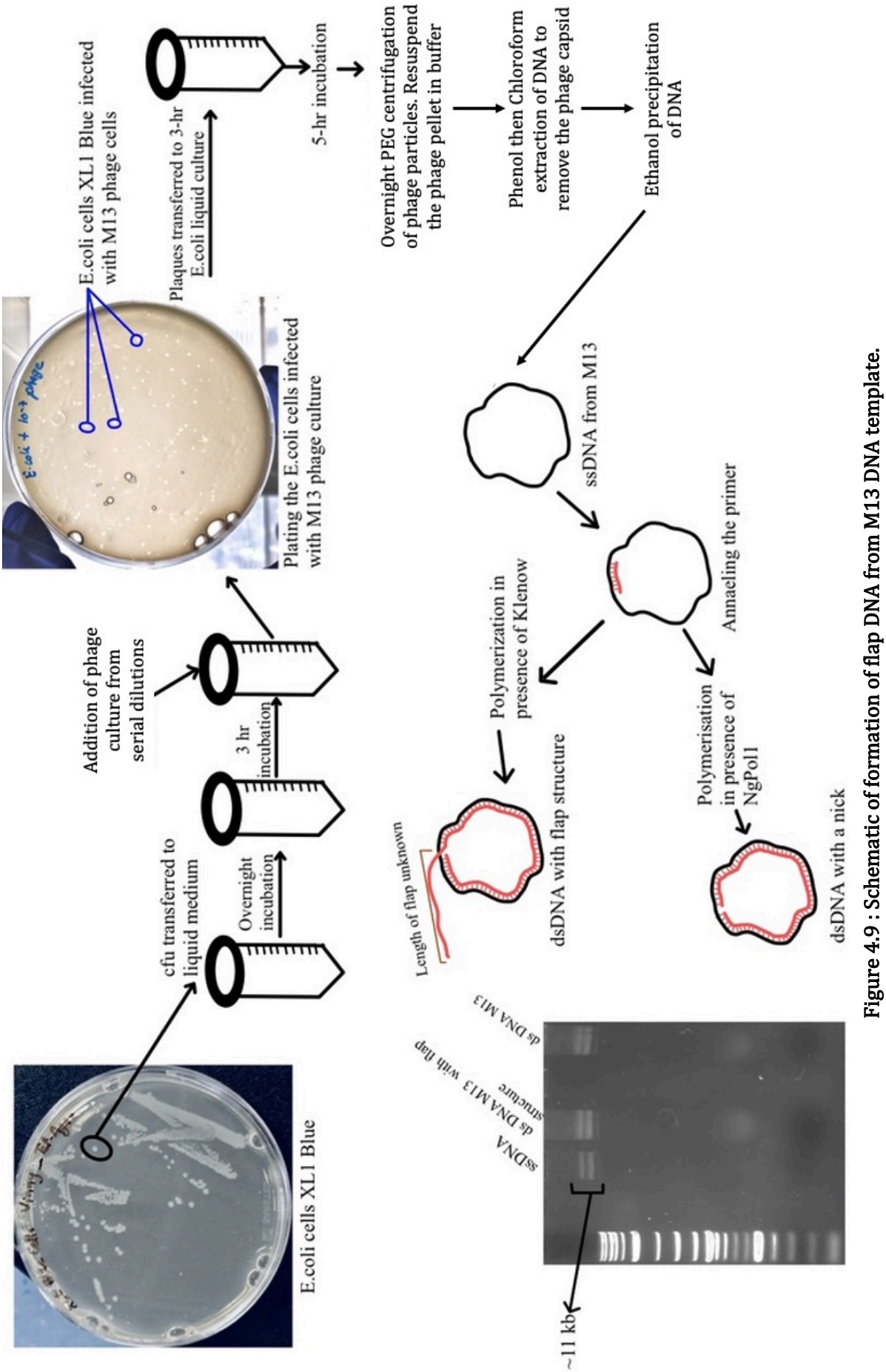


Figure 4.9 : Schematic of formation of flap DNA from M13 DNA template.

4.2.3.1. Cell Culture Preparation:

M13 phage particles require a host cell to grow so as to obtain a sufficient concentration of ssDNA. Competent *E.coli* cells, like XL1-Blue are generally used for this purpose. The XL1-Blue strain allows recombinant plasmid selection by blue-white screening and are tetracycline resistant [5]. The bacteria cells are grown in LB media which allows non-selective cultivation of *E.coli* strains for DNA production, cloning and protein synthesis, and selective cultivation of antibiotic resistant *E.coli.* Upon transfection by M13 phage, the *E.coli* cells are grown in 2X YT media, which has double the concentration of yeast extract compared to LB medium and is suitable for non-selective cultivation of M13 bacteriophage [6].

E.coli cells (XL1 blue) culture was streaked onto LB plates and incubated overnight at 37°C. A single colony was transferred into 3 mL of 2X YT medium and incubated overnight (37°C at 200 rpm). Fresh cells were prepared by adding 1 drop (50 µL) of the overnight culture into 3 mL of fresh media and incubated at 37°C for 3 hr at 200 rpm. 10 aliquots of phage cells in 2X YT medium were prepared as serial dilutions from 10^{-3} to 10^{-12}. 10 µL of each of these aliquots were added into 3 mL fresh cell culture, which were then added to 3 mL LB agar (stabilized at 55°C water-bath). This mixture was poured onto solidified LB agar plates to obtain a layer of soft agar. The culture plates were incubated overnight at 37°C [1(chap2)].

4.2.3.2. Transfection and DNA Extraction:

PEG precipitation of phage particles: Phage particles are very small and need to be pelleted by high-speed centrifugation. They therefore need to be precipitated by PEG [2(chap3)]. PEG or Polyethylene Glycol is a straight chain polymer of monomeric unit $H(OCH_2CH_2)_nOH$. It has various molecular weights based on the number of monomeric units repeats. PEG is used to assist the crowding of molecules in aqueous solutions: precipitation of DNA molecules on the basis of size, precipitation of bacteriophage particles during their purification, and various other molecular biology processes [1].

Phenol-Chloroform Extraction: For purification of nucleic acids, proteins are removed from the aqueous solutions of nucleic acids by phenol followed by chloroform extraction. This process is used to remove and inactivate enzymes that the solution might contain from previous steps, or purifying DNA from cell lysates. The proteins in the solution are first denatured by

phenol (pH equilibrated by hydroxyquinoline) and then by chloroform, which is a more efficient process due to deproteination in two organic solutions [1(see appendix 8)].

Ethanol Precipitation Of DNA: Following deproteination, precipitation of nucleic acids from aqueous solutions is performed using ethanol. Ethanol reduces the hydration layer from nucleic acids and reveals the negatively charged phosphate groups. This facilitates the binding of cations like Na^+ and causes the reduction of repulsive forces between the polynucleotide chains enabling DNA precipitation. Ethanol precipitation is aided by the presence of cations hence the addition of ammonium acetate helps neutralize the exposed phosphate residues. Ammonium acetate also reduces the coprecipitation of unwanted contaminants like dNTPs or oligosaccharides with nucleic acids. The precipitated DNA can be stored in open tube to evaporate the ethanol and then dissolved in an appropriate buffer. The concentration and purity of DNA can then be determined by spectrophotometry. Optical density at 260 nm corresponds to 50 µg/mL dsDNA, 40 µg/mL ssDNA and RNA, and ~33 µg/mL single-stranded oligonucleotides. The ratio OD_{260}/OD_{280} is measured to determine the purity of nucleic acids: a ratio lower than 1.8 indicates protein or phenol contamination requiring repetition of phenol-chloroform extraction and its following steps [1(see Appendix 8)].

A single plaque from the soft agar plates prepared above (section 4.2.3.1) was transferred into 100 µL of fresh cells and incubated overnight at 37°C. 100 mL of 2X YT media was inoculated with 1 mL of fresh overnight cells and grown at 37°C/200 rpm until OD of 0.3 at 600 nm was obtained. To this, the phage culture (in *E.coli* cells mixtures prepared above) was added and continued to incubate for 5 hr. This was transferred to centrifuge tubes (100 mL) and cells were pelleted by centrifugation. The supernatant was decanted immediately and $1/5^{th}$ volume of 20% PEG 6000 (prepared in 2.5 M NaCl) was added. The phage particles were allowed to precipitate overnight at 4°C. Phage pellet was obtained by centrifugation and it was resuspended in 1 mL TE buffer (pH 7.4) and 9 mL water. The mixture was centrifuged again and the phage-containing supernatant was transferred to a clean centrifuge tube.

With the supernatant obtained above, PEG precipitation was performed again and the phage pellet dissolved in NTE buffer (pH 8), followed by phenol-chloroform extraction, removal of chloroform by sodium acetate and lastly ethanol precipitation of DNA. The DNA pellet was resuspended in DNA buffer and its optical density was measured at 260 nm and 280 nm in a 1 mL quartz cuvette. Phenol extraction and its following steps were repeated if the ratio of OD_{260} /OD_{280} was lower than 1.8 [4].

4.2.3.3. Preparation of RF IV DNA [4]:

Annealing of primer to template DNA: A solution of 10 µL Tris-HCl (500mM, pH-8), 10 µL NaCl (500 mM), 6 µL primer and 10 µg ssDNA template (~2-5 µg/µL) was prepared in MiliQ/HPLC water to a final reaction volume of 36 µL. This was incubated at 70°C for 5 min in a water bath, followed by 37°C for 20 min to anneal. The reaction mixture was then transferred onto ice.

Polymerization: The following reagents were added to the template/primer mixture after annealing mixture had been cooled on ice: 2.5 µL each of 10 mM dATP, dGTP, dTTP and dCTPS, 10 µL ATP (10 mM), 10 µL $MgCl_2$ (100 mM), 3 µL Tris-HCl (500 mM, pH 8) and 10 units of Klenow enzyme. The volume of the reaction was adjusted to ~80 µL with water, mixed, spun briefly and then placed in a water bath at 16°C. The solution was incubated for 16-40 h and heat inactivated at 70°C for 10 min. 20 µg dsDNA was thus prepared from 10 µg of template.

4.2.4. Production and Purification of 300 bp Flap DNA:

The same method as the synthesis and purification of 100 bp flap DNA (section 4.2.1) was applied using longer oligonucleotide fragments for annealing and ligation (see appendix 7 for details of the oligonucleotide assembly), with the exception that the oligonucleotide 3 was phosphorylated first. This was done by preparation of 25 µL mix of 2.5 µL T4 DNA ligase buffer (10X), 1 µL T_4 kinase (polynucleotide kinase or PNK), 100 ng DNA (oligo 3), 2.5 µL PEG 4000 (50%) and H_2O. The reaction mix was incubated at 37ºC for 30 minutes followed by heat inactivation of the enzyme at 65ºC for 20 minutes. The mixture was then cooled on ice. The concentration of the DNA was estimated by spectrophotometry.

The DNA was not purified by gel elution as the agarose gel electrophoresis cross linked the DNA and native DNA gel elution left contaminants that were seen as noise during AFM imaging.

4.2.5. Production and Purification of Proteins:

To precipitate the protein out of the cell lysate, salting out by ammonium sulphate followed by polyethyleneimine (PEI) is performed. During salting-out, preferential solvation occurs due to the exclusion of the covalent (salt) from the protein's hydration layer that was important for maintaining the solubility and native conformation of the protein [7]. Ammonium sulphate also stabilizes the protein structure. PEI, a basic cationic polymer, binds the negatively charges nucleic acids or acidic proteins and forms a network that rapidly precipitates. The PEI is then removed by ammonium sulphate[8] which in turn needs to be removed by dialysis. Dialysis is the separation of molecules across a semi-permeable membrane due to a concentration gradient and is limited by the size of the pores of the dialysis membrane relative to the size of the molecules[1(see appendix 8)]. Dialysis membranes are available in a wide range of sizes and are used here to transfer the protein from one buffer to another and/or to remove chemicals from a protein containing buffer.

For purification of proteins, column chromatography is one of the most common techniques. Protein sample is passed through glass columns filled with modified acrylamide or agarose beads. The various versions of this technique separate the proteins on the basis of their different properties [3(chap7)].

In case of *Taq*FEN and *Taq*PolI, Ion-exchange chromatography (IEX) was used, in which the proteins were separated by their surface ionic charge using beads that were modified to be negatively or positively charged (figure 4.10). Proteins are made up of various amino acids containing weak acidic and basic groups, hence their net surface charge depends on the pH environment[9]. This chromatography method makes use of the relationship between the net surface charge and pH unique to proteins[10]. Proteins that interact weakly with the beads elute out by wash with a low-salt buffer, while the proteins that interact strongly need a strong salt buffer to elute. The salt masks the charged region causing the protein to be released from the beads. Since the surface charge of the proteins varies, different proteins are eluted from the column at

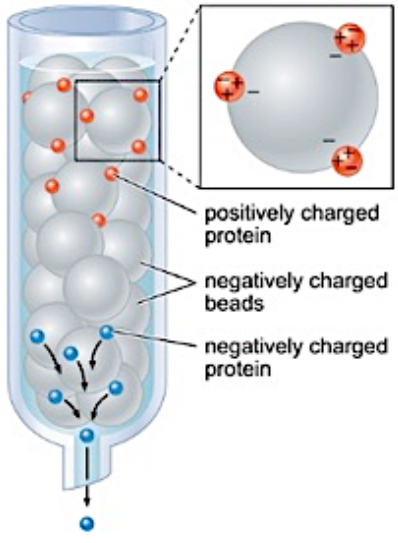


Figure 4.10 : Ion-Exchange chromatography column [3(chap7)].

a specific salt concentration. On gradually increasing the concentration of salt in the elution buffer, the proteins with similar charge characteristics can also be separated into different fractions by elution from the column [3(chap7)].

During IEX, the column is first equilibrated, then the sample is added and allowed to bind the target molecules to the resin, followed by washing to remove the outbound material. Elution is then performed to allow the bound molecules to be released from the ionic exchanger by change in buffer composition, followed by removal of all bound molecules to clean the column. The columns are selected based on the properties of the protein we are targeting. The isoelectric point or the pI (pH at which the protein has no net charge) of a protein is used to determine the kind of column to be used: a positively charged anionic exchanger (*e.g.* Q, DEAE) is used with a buffer of pH higher than pI (in which the protein carries negative charge) while a negatively charged cationic exchanger (*e.g.* S, SP, CM) is used with a buffer pH lower than pI (in which the protein has positive charge). When the pI of the protein is unknown, we usually begin with a strong exchanger (Q,S,SP) that maintains their charge over a wider pH range[11]. In the case of *Taq*PolI and *Taq*FEN, phosphate buffer (buffering range of pH 6.7-7.6) was used with a cationic exchanger SP column and Tris buffer (buffering range of pH 7.5-8.0) was used with an anionic exchanger Q column.

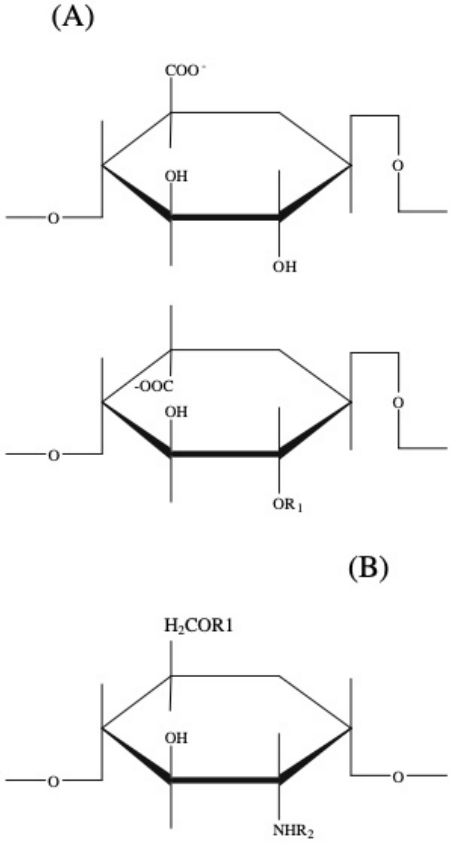


Figure 4.11 : The structure of a heparin polysaccharide, consisting of alternating hexuronic acid (A) and D-glucosamine resides (B). The hexuronic acid can either be D-glucuronic acid (top) or its C-5 epimer, L-iduronic acid. R1=-H or $-SO-3$; R2= $-SO-3$ or $-COCH_3$.

Affinity chromatography is another technique that can be applied to purify proteins based on their affinity to certain molecules, like Heparins. Heparins are negatively charged linear polysaccharides that can bind numerous biomolecules (figure 4.11). In heparin chromatography, the high charge density and distribution of heparin allows it to act as an affinity ligand and an ion exchanger. Heparin chromatography allows the molecules to be

specifically and reversibly adsorbed by heparin immobilized on an insoluble matrix and causes enrichment of low-abundance proteins [12].

Proteins can also be separated on the basis of charge to mass ratios by electrophoresis. If the protein is treated with a strong ionic detergent like SDS (sodium dodecyl sulphate) and a reducing agent like mercaptoethanol, all protein structures higher than secondary are eliminated. SDS coating allows the protein to display characteristics of an unstructured polymer. SDS ions coat the polypeptide chain which gives it negative charge while mercaptoethanol breaks all disulphide bonds between cysteine residues. Hence, electrophoresis can be used to separate the proteins like DNA and RNA [3(chap7)].

SDS-PAGE is performed in a discontinuous buffer system with the buffer having a pH and ionic strength different from the buffer in the gel cast. The SDS-polypeptide complex migrates through the stacking gel of high porosity on the application of current at the electrodes. The complex then deposits as a very thin zone/stack on the surface of resolving gel, thus allowing the sample to concentrate in a very small volume and increase the resolution of the gel. The stacking gel contains Tris-Cl (pH 6.8), the reservoirs contain Tris-glycine (pH 8.3) and the resolving gel contains Tris-Cl (pH 8.8). Cl^- (in the sample and gel) form the leading edge and glycine molecules form the trailing edge of the moving boundary. A low conductivity and steep voltage region is formed between these edges that sweeps the protein and deposits it on the resolving gel margin. Here, the higher pH ionizes the glycine which then migrates through the stacked polypeptides and travels through the resolving gel behind the Cl^-. The SDS-polypeptide complex moves ahead from the moving boundary through the resolving gel region of uniform voltage and pH and separate according to their size by sieving [1(see Appendix 8)].

To check the activity of the protein, Zymogram can be performed. It is a technique to measure the nuclease activity of protein by electrophoresis. SDS-polyacrylamide gel is infused with a protein substrate which is broken by the proteases during migration of proteins in electrophoresis. Coomassie blue staining of the gel shows regions of proteolysis as white bands on a dark blue background. The intensity of the bands can be used to estimate the amount of proteases in the sample solution to a certain extent [13].

UV Nuclease assay can also be used to check the nuclease activity of the protein, which in case of the study of activity of (inactive FEN mutant) *Taq*PolI and *Taq*FEN by AFM should be none, in order to prevent the DNA sample from being digested before it can be detected by

AFM. Type XIV DNA is digested by the protein of interest in a buffer solution and the OD measured at 260 nm at regular intervals of time in minutes. As the nucleases in the protein samples break the DNA to release nucleotides, the absorbance at 260 nm increases until it plateaus. In the absence of nuclease activity, the absorbance remains zero.

4.2.5.1. Thermus aquaticus Flap endonuclease (TaqFEN):

Thermus aquaticus DNA Polymerase I FEN domain (inactive protein with DNA binding ability) producing mutant *Escherichia coli* K12 cells were fermented and centrifuged to obtain a cell pellet. 5mL per gram of the cell pellet was suspended in lysis buffer. Lysozyme and EDTA were added and the suspension incubated at 37°C for 1 hour over a shaker. A few drops of 10 µL/mL deoxycholic acid solution were added to prevent frothing. The cell suspension was sonicated for 10 seconds three times to shear away the nucleic acids and centrifuged at 13000 rpm at 18°C for 30 minutes. This supernatant was incubated over water bath at 70°C for 30 minutes (pasteurization) and centrifuged again. Ammonium sulphate and PEI precipitation of *Taq*FEN was performed, centrifugation was done again and the pellet suspended in 50 mM salt/EDTA/TRIS buffer (pH 8). This solution was dialysed against 50 mM salt/EDTA/TRIS buffer overnight, to remove the ammonium sulphate.

Optical density was measured for the protein solution at this stage to roughly estimate the presence of the protein. Also, a small amount of the sample was kept aside at various steps and electrophoresed by SDS-PAGE to make sure the protein had not been lost during experimentation (see appendix 8 for SDS-PAGE images, and zymogram and UV nuclease assay results for ensuring that the inactive FEN domain did not have any nuclease activity).

IEX on a Q column followed by affinity chromatography on a heparin column was used to separate the protein from other unwanted contaminants. For further purification, IEX on a SP column and then a Heparin column was performed. Since *Taq*FEN does not bind to SP column, the protein was obtained in the flow-through and not the elutions. The protein solutions were assessed by SDS-PAGE and measurement of optical density.

Zymogram and UV nuclease assay were performed to detect the presence of nuclease activity.

4.2.5.2. *Thermus aquaticus* Polymerase I (TaqPolI):

Thermus aquaticus DNA Polymerase I (inactive FEN domain with DNA binding ability) over-producing mutant *Escherichia coli* K12 cells were fermented and centrifuged to obtain a cell pellet. The protein was extracted from the cell pellet in the same way as *Taq*FEN followed by IEX which was different owing to the variation in the amino acid composition of the two proteins.

IEX on a Heparin column was used to separate the protein from the other unwanted proteins. The fractions with the protein were dialysed against 50mM salt/EDTA/ TRIS buffer overnight and purified on the Q column. The fractions with the protein were dialysed against 50 mM salt/EDTA/ HEPES buffer overnight and purified on the SP column, followed by pasteurisation for 30 minutes and centrifugation to remove the unwanted proteins in the fractions. To remove the impurities still present, ammonium sulphate precipitation at varied concentrations of the salt was performed. Since the impurities were still present, purification on SP column in a low pH HEPES buffer (pH 6) and on a Q column in a high pH TRIS buffer (pH 9) was performed.

The protein solutions were assessed by SDS-PAGE and measurement of optical density. Zymogram and UV nuclease assay were performed to detect the presence of nuclease activity (see appendix 9 for SDS-PAGE images, and zymogram and UV nuclease assay results for ensuring that the inactive FEN domain did not have any nuclease activity). Equal volume of glycerol and 0.1% sodium azide were added into the samples to flash freeze in liquid N_2. The samples were then stored at -80°C.

4.2.5.3. T_7 FEN Domain:

Active T_7 FEN protein was purified as above by Prof. Sayers group (The University of Sheffield).

4.2.6. Study of Size and Shape of the *Taq* DNA polymerase I and FEN Domain:

The structure of *Taq* DNA PolI and FEN were studied using online databases like Protein Data Bank, to better estimate the size of the protein when imaging under the AFM.

4.2.7. Production and Purification of ParS site containing DNA Fragments (200-455 bp):

The plasmid DNA containing *parS* sites were synthesized and purified by Dr. Hwang's lab (University of Sheffield). The DNA contained 15 bp *parS* sites that ParB protein bound to. Short DNA fragments that contain 0,1, 2 and 9 copies of the *parS* site were created by performance of PCR[14] with the template of plasmids containing 0,1, 2 and 9 copies of the *parS* sites respectively (see appendices 15-17 for details of the DNA sequence, the plasmid was used as template with M13 primers). This step produced fragments of length 200 bp, 215 bp, 300 bp and 455 bp respectively. The PCR products were purified using Fisher Scientific Gene JET PCR purification kit [15]. The DNA was then analysed on an agarose gel (figure 4.12) and concentration was estimated by a nanodrop spectrophotometer.

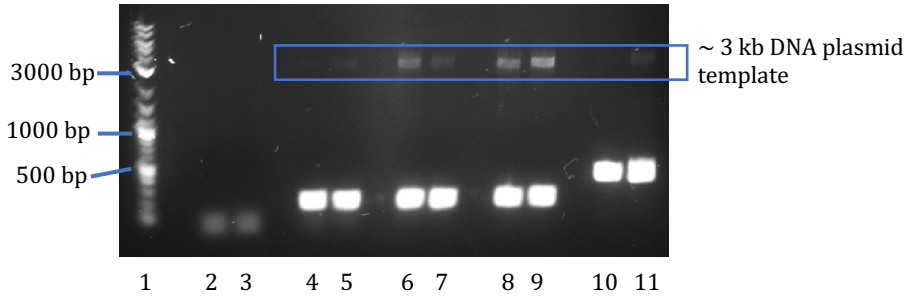


Figure 4.12: Agarose gel electrophoresis result showing the separation of short open DNA containing 0-9 *parS* sites. 1. NEB Quick-load purple 1 kb Plus DNA ladder, 2-3. Negative control (DNA primers), 4-5. 0 *parS* DNA, 6-7. 1 *parS* DNA, 8-9. 2 *parS* DNA, 10-11. 9 *parS* DNA.

4.2.8. Sample Immobilization for AFM:

For imaging the sample using AFM, the sample needs to be immobilized on a comparatively flat surface, preferably one that has atomic flatness. The surface used in the current work was muscovite mica, of which layers were cleaved to expose fresh negatively charged ionic sheet. DNA and protein samples were immobilized using surface charge interactions. DNA, comprising of a negatively charged backbone, could be directly immobilized on the mica surface, hence required neutralization of the negative charge. The protein, on the other hand, generally has some positively charged residues on its surface that allowed it to immobilize directly on the mica surface. In both the cases, experiments were performed to obtain the most suitable procedure to immobilize the sample for AFM imaging.

4.2.8.1. Preparation of the sample dilutions for imaging: Concentrated DNA and protein samples were diluted into centrifuged and filtered (through a 0.2 μm filter) HEPES buffer (pH 8), divided into aliquots, and stored at -20°C. The aliquots can be thawed when required and diluted further to the required concentration for imaging.

4.2.8.2. Preparation of parS-ParB sample dilutions for imaging: DNA sample was diluted with reaction buffer (50 mM TRIS (pH 7.5), 100 mM NaCl and 5 mM $MgCl_2$), and immobilized on PLO treated mica to image the negative control sample. The protein solution (final concentration ~3.7 μM or ~7.4 μM per 30 nM DNA) was added to this reaction mixture to obtain the DNA-protein reaction mixture, which was incubated at 21ºC for 5-10 minutes to allow the protein-DNA interaction. For experiments to image DNA-protein interaction in the presence of CTP in the buffer, the reaction mixture was prepared to obtain CTP final concentration to be either 1 μM or 0.5 μM. This solution was then further diluted to obtain an AFM suitable concentration before immobilization on PLO treated mica.

4.2.8.3. Imaging of DNA sample in air: 50 μL polyornithine (10 μg/mL) was added onto freshly cleaved mica discs and allowed to incubate at room temperature (in this case 21°C) for 5 minutes. This was washed with HPLC/deionised water by repeated pipetting and dried under nitrogen gas. 10 μL DNA sample was added onto the PLO treated mica disc and allowed to incubate for 2-3 minutes followed by washing with HPLC water by repeated pipetting and drying under nitrogen gas.

4.2.8.4. Imaging of DNA sample in liquid: After following the same procedure as imaging in air, 50 μL of suitable buffer or HPLC water was added onto the sample for imaging.

4.2.8.5. Imaging of protein sample in air: The protein sticks onto the mica discs, even without PLO. Hence, 10 μL of protein sample was added onto freshly cleaved mica discs and allowed to incubate at RT for 2-3 minutes. This was washed with HPLC water or HEPES buffer (10-25 mM) by repeated pipetting and dried under nitrogen gas. The protein sample was allowed to dry with/without washing with water. At the same time, immobilization of the protein sample on PLO was also performed in the same way as DNA (section 4.2.8.2)

4.2.8.6. Imaging of protein sample in liquid: After following the same procedure as imaging in air, ~50 μL HEPES buffer was added for imaging. Also, another method that was tried was not drying the water after washing the sample, and imaging in the water/buffer that remains.

4.2.8.7. *Immobilisation of DNA using Cations (Ca^{2+}, Mg^{2+}, Ni^{2+}):*

To immobilize the DNA molecules, divalent cations like Ca^{2+}, Mg^{2+} and Ni^{2+} could be used. Since the FEN and DNAPolI use certain divalent cations (like Mg^{2+}, Mn^{2+}, Co^{2+}, Zn^{2+} and Ni^{2+}) as cofactors for nuclease activity [16], we were limited to Ca^{2+} for the immobilization. Yet, for the estimation of a range of concentrations of divalent cations required for Ca^{2+} immobilisation, Mg^{2+} and Ni^{2+} were experimented with as well. The aim was to immobilize the DNA samples in a way to be detected by the AFM cantilever and at the same time be flexible enough to interact with the protein during dynamic imaging. Divalent cationic chloride solutions were prepared in MiliQ or HPLC water and filtered through 0.2 µm filters followed by centrifugation to remove impurities or precipitates. Solutions ranging from 1.25 nmoles to 100 µmoles were added on ~5 cm^2 of freshly cleaved mica, incubated at 16ºC, washed with H_2O and dried with N_2. DNA solution of concentrations varying from 25 fmoles to 7.5 pmoles was then added, incubated at 16ºC, washed with H_2O and dried with N_2. This sample thus prepared was imaged in dry environment. Imaging in HEPES buffer or water was performed by addition of 50 µL buffer or water.

4.2.8.8. *Immobilisation of 300-2000 bp DNA fragments using Ni^{2+}:*

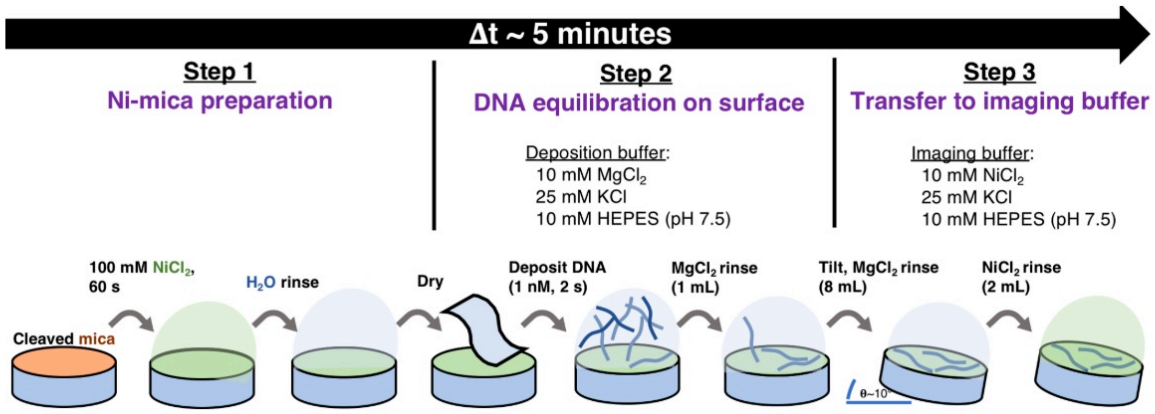


Figure 4.13: Schematic of sample preparation method for AFM by Heenan and Perkins (2019) [17].

The persistent issue with immobilization of long DNA strands was the convolution of DNA to the extent that no conformational details could be visualized. This was addressed by allowing the DNA to gradually transfer from a 3-dimentional solution to a 2-dimentional surface without any sudden change in the buffer conditions of the DNA solution. This was achieved for 200-3000 bp open DNA strands by Heenan and Perkins (2019)[17]. The method was hence tried with

closed plasmid DNA of lengths ~3 kb. The method details are as follows: freshly cleaved mica was coated with Ni^{2+}, followed by addition of DNA in a deposition buffer containing Mg^{2+} and finally immediate washing with an imaging buffer containing Ni^{2+} to seal the DNA molecules (schematic of the method in figure 4.13). The core of this method is the avoidance of drying the surface at any step of the immobilization.

4.2.8.9. Imaging of flap DNA and FEN protein together on PLO, imaging in air and liquid:

Flap DNA was mixed with protein in a suitable concentration in HEPES buffer (25 mM) containing Ca^{2+} (2 mM) and K^{+} (5mM). The reaction mixture was incubated at a range of temperatures (16ºC - 60ºC) for numerous time periods (5-60 s) and added onto PLO treated mica. The sample was then washed with 25 mM HEPES buffer (without salts) and dried under N_2.

4.2.8.10. Dynamic imaging of FEN or DNAPolI interacting with DNA:

DNA solution was prepared in HEPES buffer (25 mM HEPES, pH ~7.5) containing 2 mM Ca^{2+} and 5 mM K^{+}. In either case of active and inactive FEN, the sample was immobilized on PLO treated mica, washed with HPLC water and dried under N_2. To image the interaction with inactive FEN, the sample prepared above was imaged in the same buffer as the DNA solution. In case of interaction with active FEN, the same buffer solution but with Mg^{2+} in place of Ca^{2+} and an additional 10 mM DTT was used for imaging. During the course of imaging, protein solution (in HEPES buffer) was added (~10-20 µL) onto the sample with caution to not disturb the imaging in process. Continuous images were taken to observe the conformation changes.

4.2.9. Softwares for Image Analysis:

4.2.9.1. Bruker Nanoscope Analysis 1.7: used to flatten the AFM images and obtain the cross-section of DNA or protein molecules to observe the variation in height.

4.2.9.2. Image J (Fiji) with plugin NeuronJ: used to trace the lengths of DNA in order to observe any change in lengths due to binding of protein, to characterize the molecules on the basis of their lengths and to verify the efficiency of DNA ligation.

4.2.9.3. Microsoft Excel: used to plot the histograms and box and whisker plots of lengths of DNA, and line graphs to plot the motion of DNA in dynamic AFM images.

4.2.9.4. Adobe Photoshop: to assemble dynamic AFM images into videos (image time-lapse).

4.2.9.5. Adobe Illustrator: to segment the DNA strands in the dynamic AFM images to be tracked separately.

4.2.9.6. Tracker (Video Analysis and Modelling Tool): The software was used to track the segments of DNA prepared from Adobe Illustrator, to quantify the motion of DNA due to protein interaction.

Bibliography

1. Sambrook R. Molecular Cloning - Sambrook & Russel - Vol. 1, 2, 3.pdf. 2001. doi:10.1002/polb.23247
2. Brown TA. Gene Cloning and DNA Analysis. 6th ed.; 2010.
3. Watson JD *et al*. Molecular Biology of the Gene. 7th ed. Cold Spring Harbour Laboratory Press; 2013.
4. Sayers JR, Eckstein F. Phosphorothioate-based site-directed mutagenesis for single-stranded vectors. In: McPherson MJ, ed. Directed Mutagenesis: A Practical Approach (Practical Approach Series) 1st Edition. first. Oxford University Press; 1991:49-68.
5. Stratagene. XL1-Blue Competent Cells. www.stratagene.com. Published 2004.
6. Aldrich S. Introduction to Microbial Media | Sigma-Aldrich. https://www.sigmaaldrich.com/technical-documents/articles/biology/microbial-media.html.
7. Wingfield PT. Protein Precipitation Using Ammonium Sulfate HHS Public Access. Curr Protoc Protein Sci. 2001;3:3.
8. Burgess RR. Chapter 20 Protein Precipitation Techniques. Vol 463. 1st ed. Elsevier Inc.; 2009.
9. GE Healthcare. Ion Exchange Columns and Media Selection Guide.
10. GE Healthcare. Ion Exchange Chromatography Columns and Resins Selection Guide.
11. Ion Exchange Chromatography | LSR | Bio-Rad. https://www.bio-rad.com/en-uk/applications-technologies/ion-exchange-chromatography?ID=MWHAY9ESH.
12. Xiong S, Zhang L, He QY. Fractionation of proteins by heparin chromatography. Methods Mol Biol. 2008;424:213-221.
13. Leber TM, Balkwill FR. Zymography: A Single-Step Staining Method for Quantitation of Proteolytic Activity on Substrate Gels. Vol 249.; 1997.
14. PCR Biosystems. Product Description: PCRBIO HiFi Polymerase Uses the Latest Developments in Polymerase Technology. http://frodo.wi.mit.edu/primer3/.
15. Fisher Scientific UK. GeneJET PCR Purification Kit.
16. Feng M *et al*. Roles of divalent metal ions in flap endonuclease-substrate interactions. Nat Struct Mol Biol. 2004;11:450-456.
17. Heenan PR, Perkins TT. Imaging DNA Equilibrated onto Mica in Liquid Using Biochemically Relevant Deposition Conditions. ACS Nano. 2019;13(4):4220-4229.

Chapter 5

Immobilization Techniques for 100 bp DNA fragments

5.1. Introduction

For AFM imaging, the sample to be imaged should ideally be immobilized onto an atomically flat surface such that the cantilever can map the topography of the sample features against the backdrop of a relatively flatter surface. The immobilization can be done on bare or chemically modified mica, supported lipid bilayers (SLB) on mica, 2D crystals of streptavidin grown on planar lipid bilayers containing biotin lipid, highly oriented pyrolytic graphite (HOPG), glutaraldehyde, DNA origami tiles placed on a mica surface, glass coverslips, plain or silanized graphite (HOPG) or gold, *etc*. The substrate can also be derivatized if required [1,2].

The B form of DNA has ~34 Å (3.4 nm) helical pitch with 10 base pairs per pitch and a width of ~24 Å (2.4 nm) (see table 5.1)[3]. The length of 100 bp long flap and nicked DNA used in the current work can thus be estimated to be ~34 nm. To image DNA molecules, the surface on which the sample is immobilized needs to have a lower roughness than the DNA molecules, *i.e.* a height << 2.4 nm. Hence, an atomically flat surface like mica can be used for this purpose (see figure 5.1). As already described in section 3.6.1, layers of mica can be cleaved to expose negatively charged K^+ ions. Thus electrostatic adsorption can be used to fix charged samples on to the mica surface[4]. Negatively charged molecules like DNA can be adsorbed on mica by either the addition of divalent cations (Mg^{2+}, Ca^{2+}, Ni^{2+}, or Zn^{2+}) in the sample buffer [1,5,6] or treatment of mica surface with these cations or positively charged molecules like poly-L-ornithine (PLO) or poly-L-lysine (PLL). Different binding capacities can be applied to alter the DNA surface adhesion by regulating the concentrations and ratio of monovalent and divalent cations, and their exposure time [4,7].

	A	B	Z
Overall proportions	Short and broad	Longer and thinner	Elongated and slim
Rise per base pair	2.3 Å	3.32 Å	3.8 Å
Helix-packing diameter	25.5 Å	23.7 Å	18.4 Å
Helix-rotation sense	Right-handed	Right-handed	Left-handed
Base pairs per helix repeat	1	1	2
Base pairs per turn of repeat	~11	~10	12
Pitch per turn of helix	24.6 Å	33.2 Å	45.6Å

Table 5.1: Comparison of characteristics of A, B and Z form of DNA, reproduced from *Watson (2014)* [3].

In this chapter, the immobilization of 100 bp long DNA by neutralization of the negative charge of mica using poly-L-ornithine and divalent cations has been explored to determine which method is most suitable to allow the dynamic imaging of DNA-protein interactions. The following sections will explore the various concentrations of DNA on PLO and cation treated mica and a range of concentrations of divalent cations that allow efficient immobilization.

5.2. Immobilization using Polyornithine

Poly-L-Ornithine is a synthetic positively charged amino acid that is used for the attachment of molecules and cells on surfaces like plastic and glass [8].

To study protein-DNA interactions, the first step was to standardize the parameters and conditions for imaging the DNA alone by AFM. As already mentioned in section 2.2.3, FEN recognises the end of the ssDNA branching out from the dsDNA (figure 5.1 e) and does not depend on sequence specific recognition of flap DNA. With the aim to understand this protein DNA interaction, any DNA strand with a flap structure of any sequence and length could be used in a biological setup. For AFM, the concentration required is far less than that required for biological applications and assays. Hence, flap DNA and nicked DNA structures (figure 5.1 d-e) could be synthesized (see section 4.2.1) by annealing and ligation in low concentrations, with nicked DNA being used as a negative control to examine DNA-protein interactions in the absence of flap. The oligonucleotides were ordered from Eurofins and the size of the DNA construct was limited to 100 bp. The important aspect of the DNA synthesis was the removal of impurities of any kind as they would be visible on an AFM image as noise.

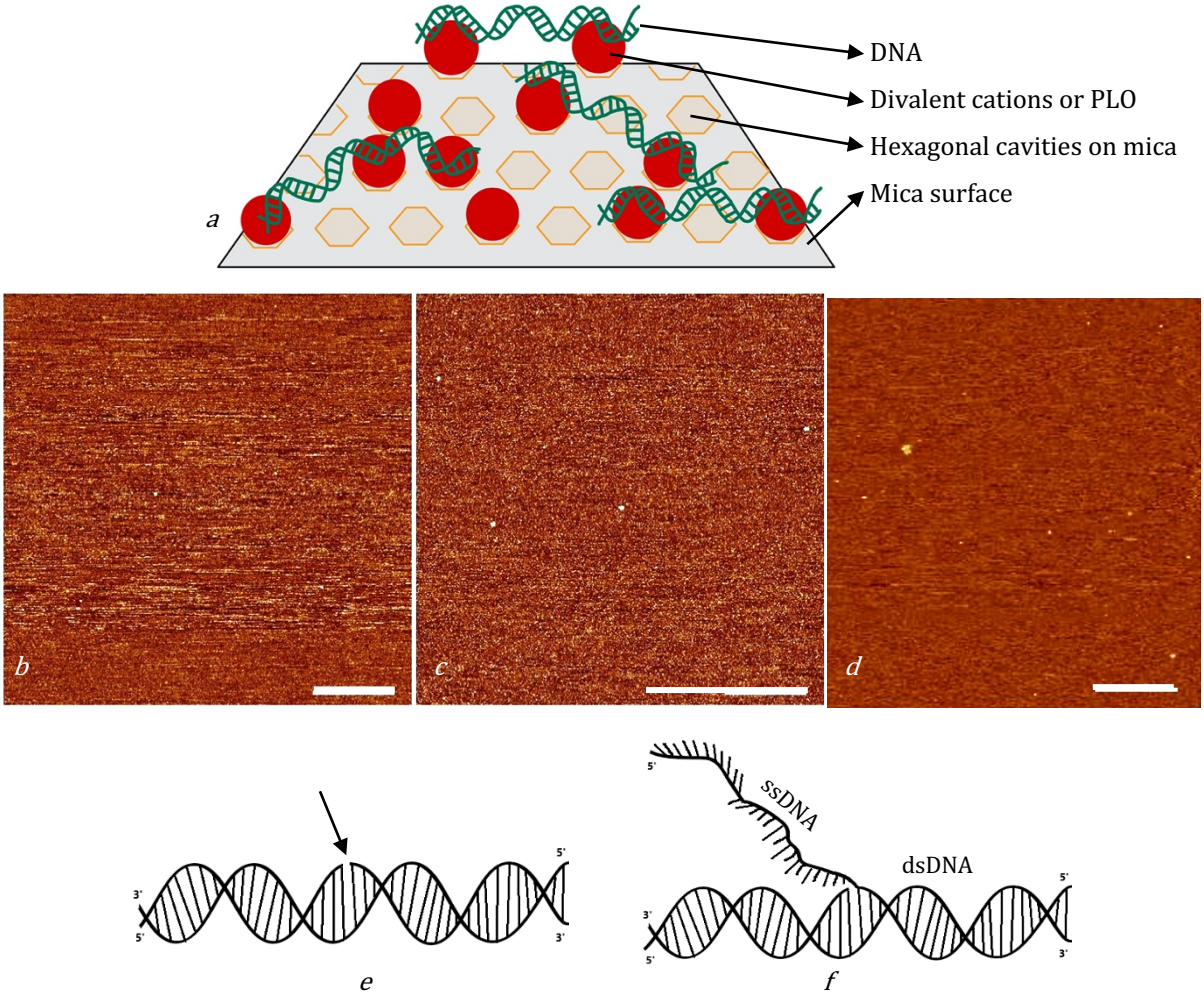


Figure 5.1: a. Schematic of the mica surface treated with divalent cations or PLO with DNA immobilized on it. b-c: Mica surface imaged in water and 5 mM HEPES (pH 8) respectively. Scale bar 400 nm, height 2 nm. d: PLO treated mica imaged in air: scale bar 1µm, height 1 nm e: Diagrammatic representation of Nicked DNA (arrow shows the single break in the phosphodiester bond) and f: Flap DNA consisting of dsDNA with ssDNA branch.

As described in section 4.2.8.1, 4.2.8.3 and 4.2.8.4, the DNA sample was imaged on PLO treated mica. As negative controls, PLO-free and PLO-treated mica were imaged in air, water, & HEPES (pH-8) to observe the surface for DNA or protein immobilization (see figure 5.1). The surface appeared to be very flat and no topographical features of notable height were observed.

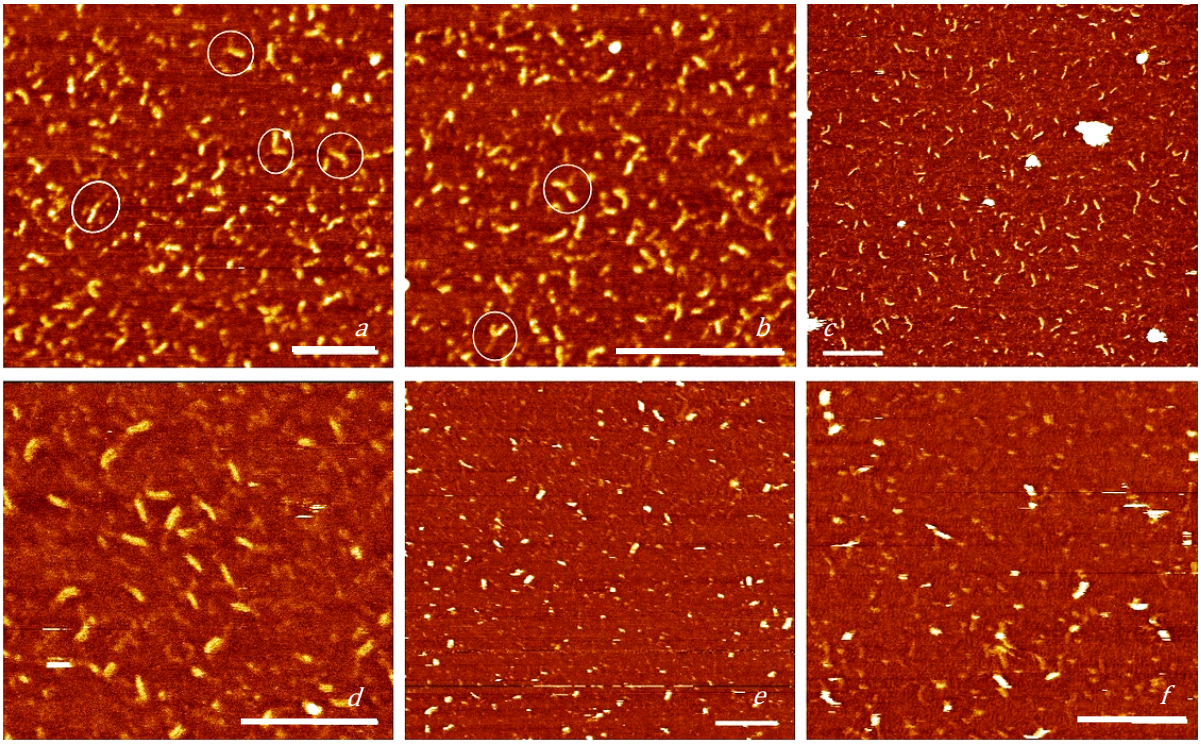


Figure 5.2. 0.5 pmole flap DNA on PLO: images in air a-b, height 1.4 nm, scale bar 100 nm.
images in water: c-d, height 6 nm, scale bar 100 nm.
images in buffer (5 mM HEPES, pH 7) e-f, height 6 nm, scale bar 100 nm.
Flap DNA in air (encircled in images a-b) show the presence of ssDNA branch next to the dsDNA, but the images in liquid look like nicked DNA with almost no molecule showing flap. Many molecules appear as overhang DNA either because of mobility of DNA at the ends or inefficient ligation step causing the formation of an overhang instead of flap DNA.

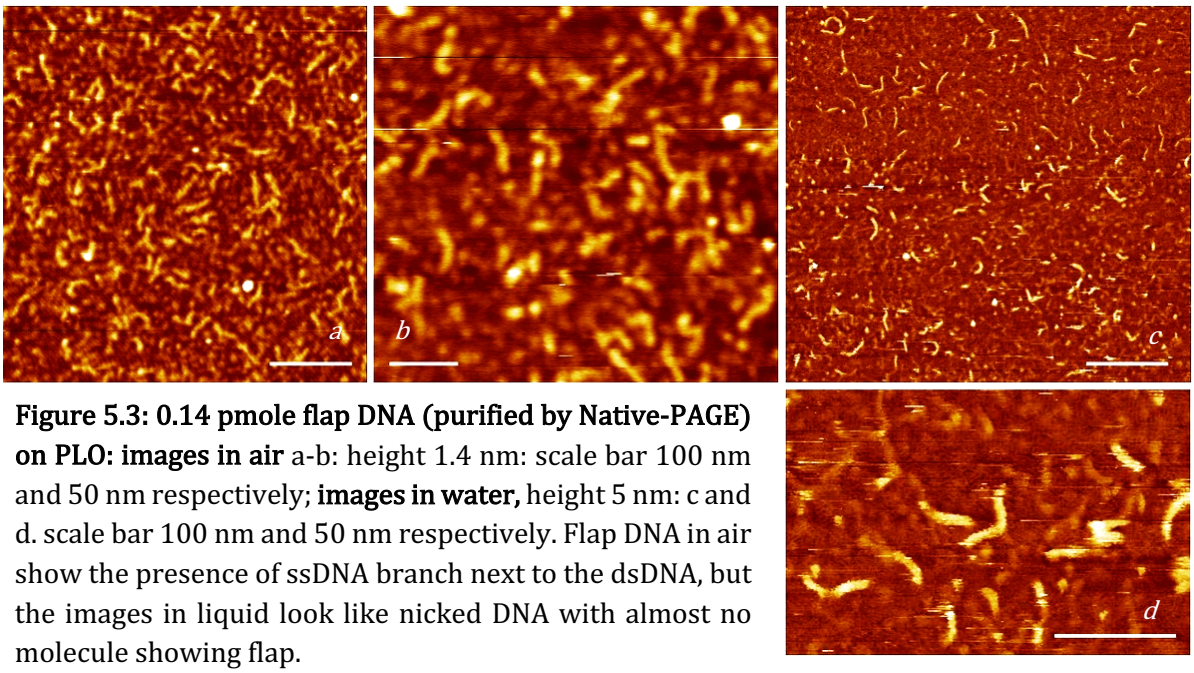


Figure 5.3: 0.14 pmole flap DNA (purified by Native-PAGE) on PLO: images in air a-b: height 1.4 nm: scale bar 100 nm and 50 nm respectively; **images in water,** height 5 nm: c and d. scale bar 100 nm and 50 nm respectively. Flap DNA in air show the presence of ssDNA branch next to the dsDNA, but the images in liquid look like nicked DNA with almost no molecule showing flap.

0.5 pmole (20 µL, 25 nM) of flap DNA and nicked DNA sample were immobilized on PLO treated mica discs of area ~30 mm^2. The amount of DNA to be immobilized was estimated through trials of various sample concentrations. During sample preparation, the DNA was treated by various methods to obtain the optimized conditions for immobilization on PLO coated mica with least amount of noise: the DNA sample when added onto the PLO treated mica was either washed with ~150 µL H_2O and then dried with N_2, or washed and allowed to air dry, or not washed at all and allowed to air dry. In the first 2 cases, the images displayed DNA molecules on a clean surface but, when the images had not been washed with H_2O at all, the images displayed a considerable amount of noise. Thereby, the method of washing and N_2 drying method was adhered to.

The Flap DNA imaged in air (see figure 5.2 a-b and 5.3 a-b) revealed a Y-shaped flap structure (observed in multiple molecules seen in images produced during repeated experiments spanning several days) of height ~300 pm compared to the dsDNA section of height 400-500 pm (figure 5.4). The length of the flap structure was ~12 nm long while the double-stranded section was ~ 28 nm long and ~6 nm wide. The flap was absent in some of the DNA molecules and there were some smaller fragments as well, thereby indicating that the ligation and annealing reactions had not been 100% efficient, which is the case with ligase enzyme catalysed reactions. The images in air, though showing the distinction between the dsDNA and the ssDNA branch, did not reveal the helical pitch of the DNA molecules and there were some very rare instances where the "beadiness" of the DNA strand could represent the helix. Here, the cross section of DNA fragment revealed peaks ~4 nm or its whole number multiple apart, which represented the pitch of the DNA helix.

When imaged in liquid, DNA molecules revealed some very faint (*i.e.* of a lower height) strands branching from the middle of the dsDNA, which could be flaps. But these were very rare occurrences (see figure 5.2 c-f, 5.3 (c-d) and 5.5) and most of the flap DNA imaged in liquid did not show the presence of the flaps that were seen during imaging in air. The assumed ssDNA flaps seen could have very easily been some unannealed ssDNA fragments randomly immobilized close to the dsDNA seen as background noise. When in folded conformation, it would have been difficult to distinguish these from the protein molecules that looked like "blobs" as well. The short length of the flaps was possibly due to the folding of flap structure, mobilization of flap so as to not be detected by AFM tip, or, as later revealed (see section 6.4), ssDNA having a height very low in comparison to the adjacent dsDNA (~ 400 pm height of flap vs the ~1500 pm height of dsDNA) (see figure 5.5) which made it inconspicuous. The

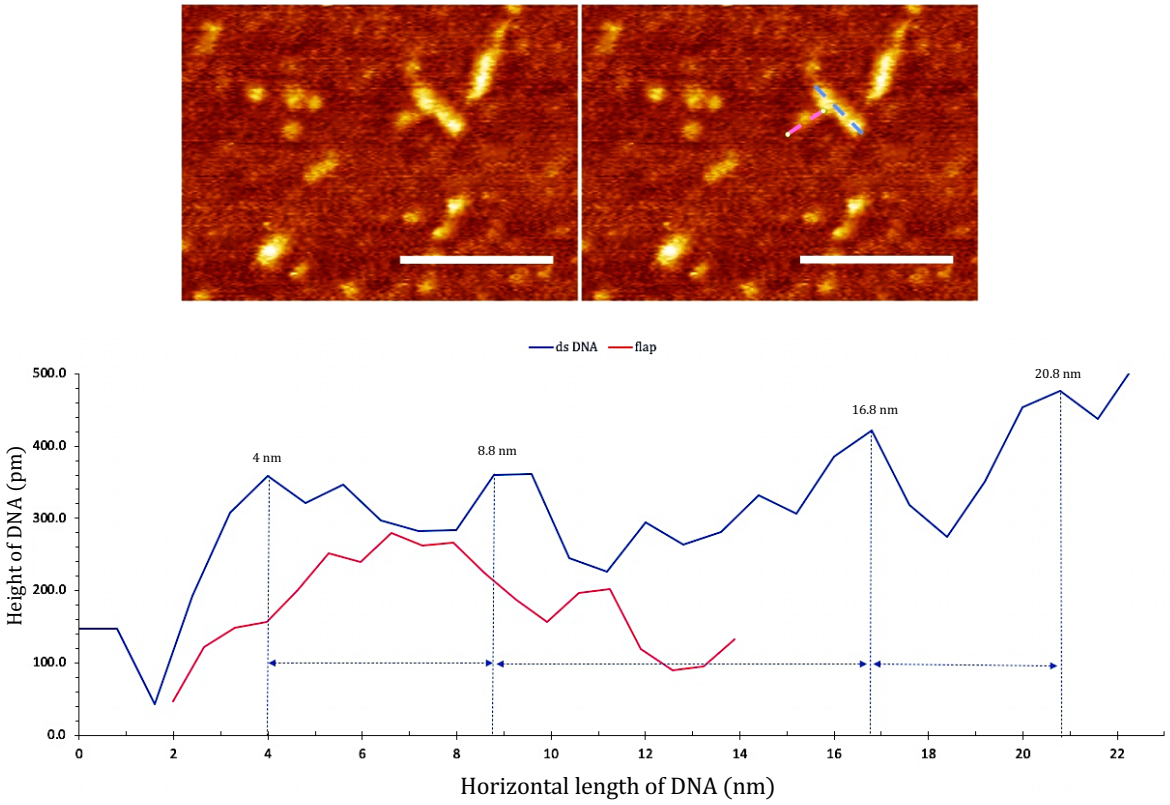


Figure 5.4: 0.5 pmole flap DNA on PLO: images in air: height 1.4 nm, scale bar 60 nm. **Cross section of double-stranded part and single-stranded flap of DNA molecule.** The numbers on the plot indicate the distance in nm of the peaks in the cross section of DNA. The distance between the peaks on the dsDNA average at 4 nm.

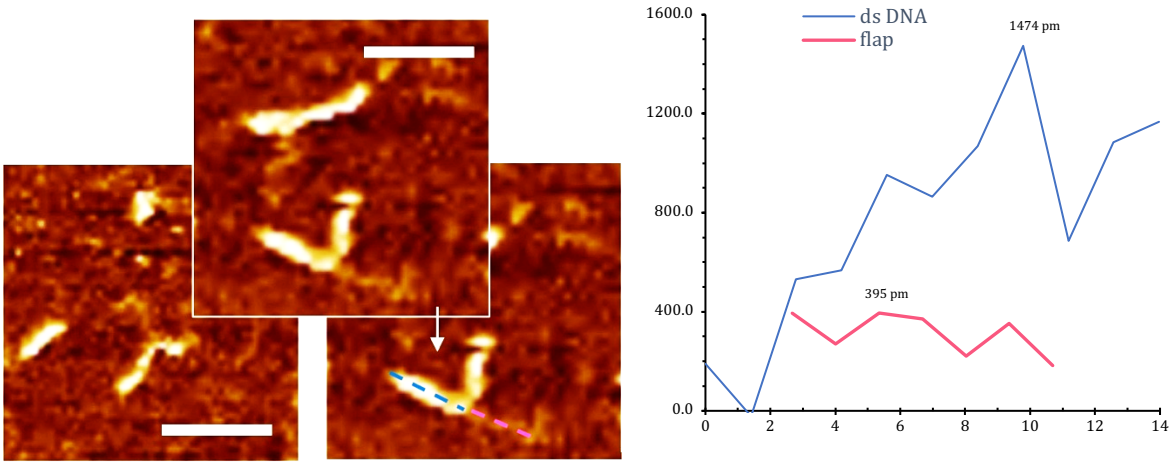


Figure 5.5: 0.5 pmole flap DNA on PLO: images in water: height 4 nm, scale bar 20 nm and 25 nm respectively. **Cross section of double-stranded part and single-stranded flap of DNA molecule**. The numbers on the plot indicate the height in pm of the peaks in the cross section of DNA. The dsDNA has a large height compared to the ssDNA, hence making it difficult to observe the ssDNA when the length of the strand was very short.

situation was worsened by the mesh-like structure of PLO prominently seen in liquid imaging which made it very difficult to distinguish the short 50 nucleotides (nt) ssDNA from the PLO. This predicament could be resolved by either immobilization by a better method such that the

flap sticks well to the surface even during imaging in liquid (see section 5.3), or the increase in length of flap (chapter 6) for easier immobilization and detection. Attempts were made to try to locate the exact same scan site such that precise DNA molecules could be imaged in air, water and buffer. But, owing to the small size and lack of a characteristic shape of the individual DNA molecules, the attempts weren't successful.

The height of dsDNA was higher when imaged in liquid (~2 nm) than during imaging in air (~ 800 pm) (see figure 5.4-5). This was possibly either due to the cantilever tip also imaging the hydration layer formed around the DNA molecules in liquid imaging or the higher electrostatic repulsions between the DNA and tip[9] when imaging in liquid. The works of Santos *et al*[10] have allowed to explain the variation in height by considering the force between the sample and the tip to spread out during imaging interactions with pressure distribution. This happens when the sample features that are to be measured are smaller than the effective area of interaction between the tip and the sample surface, and gives the measured height of the sample as a convolution between the height of the sample and the surface on which the sample is immobilized.

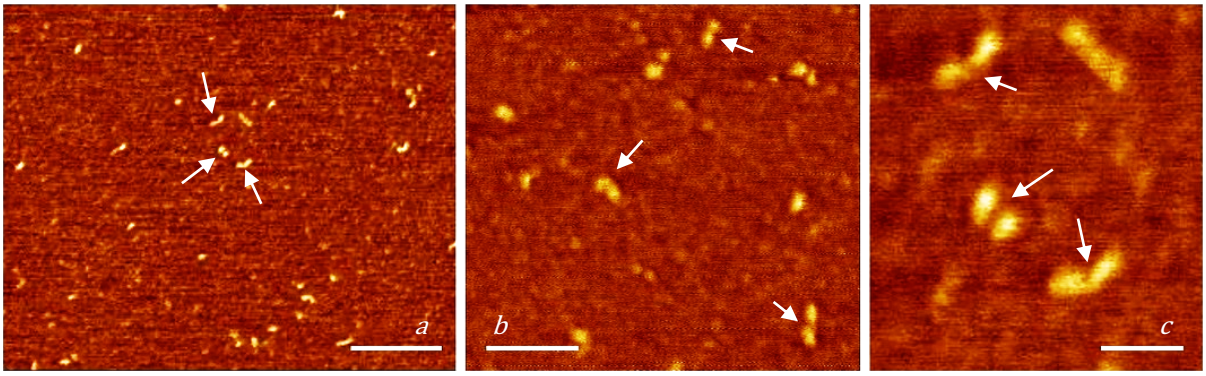


Figure 5.6: Nicked DNA on PLO, images in air: a. height 1 nm, scale bar 200 nm. b. height 2 nm, scale bar 100 nm, c. height 1.2 nm, scale bar 40 nm. The arrows in a and b indicate the nicked DNA with a distinct dumbbell shape. The arrows in figure c mark the phosphodiester bond break on the DNA, that appear as the neck of the dumbbell.

Nicked DNA imaged in air revealed a prominent single phosphodiester bond break as the neck of a 'dumbbell' with each of the arms ~15 nm long (see figure 5.6). However, in comparison, nicked DNA shape changed to strand shape in the later liquid imaging conditions (when immobilized using divalent ions, see section 5.3.2, 5.3.3 and 5.3.5) without any visible break seen on the DNA, though there were bends visible. The prominence of the gap in the dsDNA during imaging in air could be explained by the constriction of the 'arms' of the nicked DNA to widen the phosphodiester bond break, or the lack of shielding of DNA by the absent

hydration layer. The latter, though, seemed less plausible as the 'gap' appeared to be more pronounced which would not have been the case if the tip was detecting the break as it was. In liquid imaging, on the other hand, the DNA molecules were more relaxed causing the 'gap' to assume its normal size. Since the single bond break would be very short, (shorter than the 0.332 nm difference between two consecutive nucleotides in B-form DNA, see table 5.1) the AFM tip (~ 5 nm tip radius for Fastscan-D) could not detect it most of the time. Hence the DNA appeared as continuous strings and not as a dumbbell.

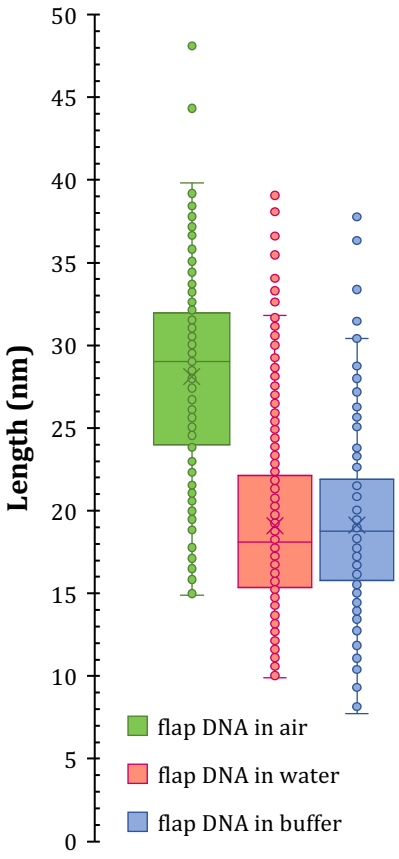


Figure 5.7: Comparison of lengths of 100 bp flap DNA structures when imaged in different environment. The lengths were measured from the images using NeuronJ.

A dilemma of imaging in water and buffer was the short length of the DNA fragments (figure 5.7) compared to the longer DNA observed during imaging in liquid. The DNA molecules in air had an average length (measured using Image J plugin NeuronJ) of ~ 28 nm compared to ~19 nm in water and buffer, while the theoretical length is ~ 34 nm. This could be due to multiple reasons. Firstly, the DNA might not be efficiently immobilized, especially at the ends, during the addition of liquid owing to the small size of DNA molecules limiting the efficiency of immobilization. Secondly, the molecules might have folded over in the z-plane which could be corroborated by an increased height of the DNA molecules. But this was not the case here as the DNA height averaged at ~1.5 nm, not incomparable to the theoretical width of the DNA double helix (~ 2.4 nm), considering the force of tip-sample interaction might have reduced the height of the DNA. But this was possibly the case for immobilization of DNA on divalent cations (see section 5.3) where the height of the DNA was > 4 nm.

Thirdly, the DNA could be breaking in the presence of liquid, specifically liquids at low pH (<5). The pH of the water used could have been lowered by the atmospheric CO_2, hence during subsequent experiments, water was verified to be pH ~7 and was used from freshly opened bottles. Yet, the short length persisted, even while imaging in buffer conditions of pH ~ 8 (HEPES 5 mM). The breakage of DNA in liquid could also be attributed to the sharper tips

used for imaging in liquid, but this has been disregarded as the cause of damage in previously reported experiments [9].

Furthermore, the reduction in length could be due to the change in DNA conformation from the longer B-form to a shorter A-form during immobilization. However, past research has shown that the DNA conformation of the immobilized molecules is determined by the conformation of DNA molecule in the aqueous environment in which it bound to mica, and that subsequent dehydration doesn't shorten the DNA to A-form. Nevertheless, DNA molecules when washed with buffer and dried to image in air, displayed shorter lengths than washing with water (like it was done in these experiments) due to the dehydration assisted conversion of DNA to A form[9].

It was possible that the DNA was very short hence the dimensions were almost spherical, *i.e.* the width of DNA was almost same as the length imaged using AFM. Lastly, there was also the possibility that some denaturation of dsDNA ends occurred that gave the DNA a globular shape[11].

Due to the above explained observations of difficulty in visualizing ssDNA flap and shortening of DNA lengths, the method of immobilization needed to be improved. Therefore, immobilization using divalent cations was experimented with.

5.3. Immobilization using Divalent Cations

DNA contained in a divalent cation buffer can be immobilized on negatively charged mica surface strongly enough to be imaged using AFM. It has been reported [4,12–15] that DNA immobilization is dependent on the radius of transition metal cations, Ni^{2+} being the most adequate with an ionic radius between 0.69-0.74 Å. The dependency is speculated to be due to the ease with which smaller cations can fit into the helical grooves to neutralize the negative charge of the DNA phosphate backbone (section 3.6.2).

The divalent cations that could be used for immobilization of flap DNA to study the interactions with FEN are limited by the fact that numerous cations act as cofactors for the protein to perform exonuclease activity and cleave the DNA strand[16]. The enzyme in the presence of cations would cut the DNA strand too quickly for the AFM to observe any activity when the protein is added onto the immobilized DNA sample. On the contrary, if the protein is allowed to react with DNA in a vial before immobilization for AFM, FEN would cleave the flap off in the presence of divalent cations and the flap would not be seen at all. Ni^{2+} and Mg^{2+} have been shown to be suitable for immobilization of short DNA without protein, as short as 300 bp [15]. These were experimented with to check if they do prove effective for immobilization of DNA lengths of 100 bp with the prospect of varying the concentrations of these cations to slow the rate of protein interactions with DNA to a level that the AFM can successfully observe it. In chapter 7, Mg^{2+} in the reaction buffer were used to allow the interaction of active FEN protein with DNA that has been immobilized on PLO treated mica.

Various concentrations of Ca^{2+} were also experimented with to immobilize DNA molecules. Ca^{2+} act as cofactors for the binding of FEN to flap DNA but don't facilitate protein activity. Hence it could be used as a control experiment to observe DNA-protein interactions with the assurance that the flap would not be chopped off. Chloride salt solutions were used to treat the mica surface (see section 4.2.8.7) with the concentration of DNA used based on the amount of DNA immobilized on PLO.

5.3.1. Flap DNA on Ca^{2+} treated mica:

When the 0.5 pmole of flap DNA was immobilized on 25 nmole $CaCl_2$ treated mica (area 50 mm^3), there were no DNA molecules visible (see figure 5.8). This indicated that there was an absence of charge interactions between the mica surface and DNA, or the concentration of DNA was low for immobilization by Ca^{2+}. Hence, the amount of $CaCl_2$ was increased to 50 nmole and the DNA sample was allowed to dry on Ca^{2+} treated mica without the use of N_2, as was done before. Here, though the amount of DNA visible was significantly more, the molecules tended to move around during consecutive scans (see figure 5.9). This was either because the concentration of Ca^{2+} was still low to allow stable immobilization or the imaging force was very high such that the DNA molecules were displaced by the cantilever. When the DNA quantity was doubled to 1 pmole on the same Ca^{2+} concentration (50 nmole) (figure 5.10), the images obtained were similar to figure 5.8 showing no DNA. This demonstrated that the effectiveness of immobilization depended on the ratio of Ca^{2+} to DNA and not on their respective concentrations.

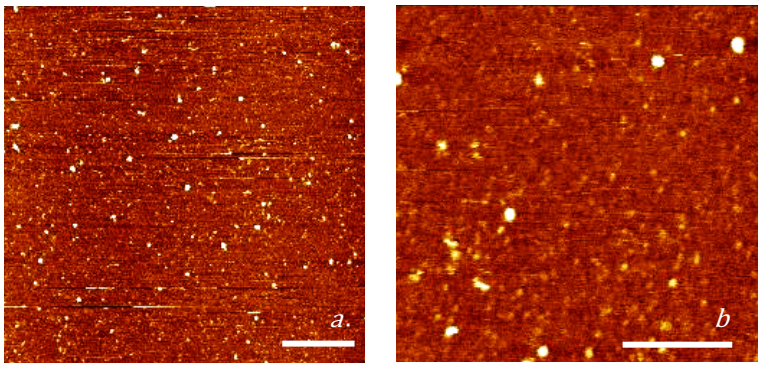


Figure 5.8: 25 nmole $CaCl_2$, 0.5 pmole flap DNA, imaging in air: a. scale bar 200 nm, height 1 nm; b. scale bar 100 nm, height 1 nm. No flap DNA could be seen as the DNA had not immobilized.

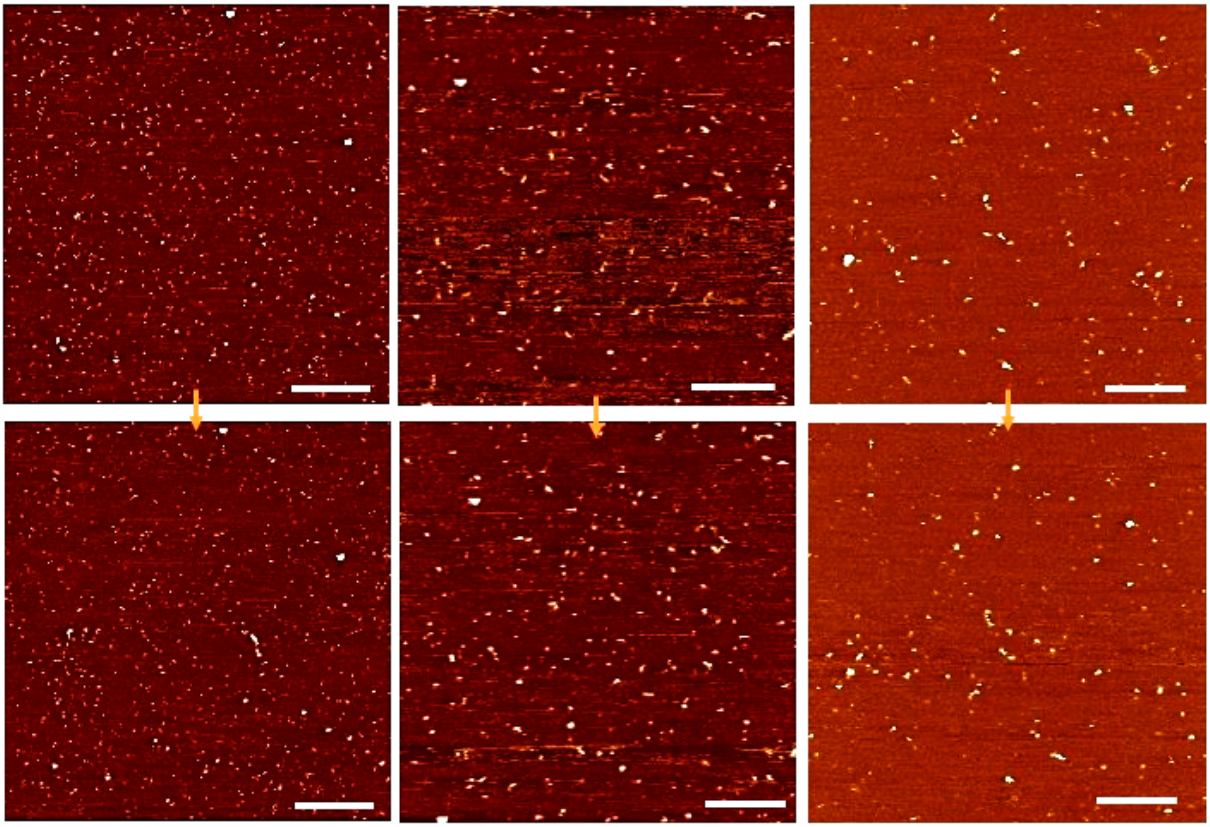


Figure 5.9: 50 nmole $CaCl_2$, 0.5 pmole flap DNA, imaging in water:
a. scale bar 400 nm, height 5 nm, b. scale bar 200 nm, height 5 nm, c. scale bar 120 nm, height 4 nm. The consecutive images, shown by arrows, show movement of particles. No flap DNA could be seen as the DNA had not immobilized.

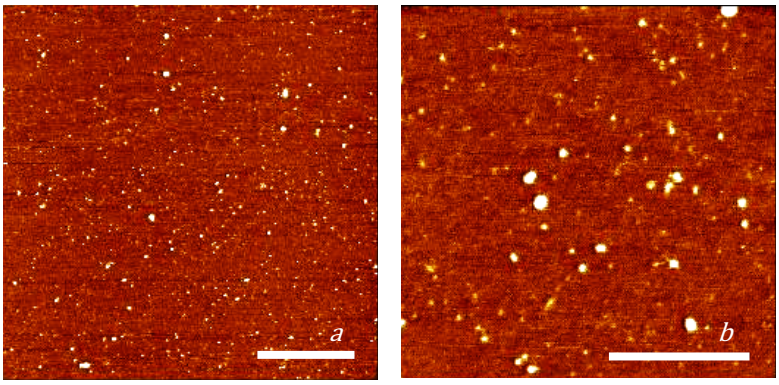


Figure 5.10: 50 nmole $CaCl_2$, 1.0 pmole flap DNA, imaging in air: a. scale bar 200 nm, height 1 nm; b. scale bar 100 nm, height 1 nm. No flap DNA could be seen as the DNA had not immobilized.

At higher amounts of Ca^{2+} (100 nmole), 0.25 pmole of DNA immobilized strongly enough to obtain good spatial resolution where the major grooves of the DNA were prominent (figures 5.11-12). There was very little movement of DNA fragments thereby indicating that the amount of Ca^{2+} was adequate for immobilization of dsDNA. At the same time, the lengths of DNA, measured using NeuronJ Fiji plugin, were shorter (~25 nm, see figure 5.11f) than the 35 nm length expected for 100 bp DNA. This indicated that the DNA had either broken off, had folded over or were less immobilized at the ends. The DNA breaking could be highly unlikely, but the DNA folding should have given an increase in DNA height or breadth at the ends. Since the DNA height was substantially high (>3 nm) compared to the expected 2.37 nm theoretical width of DNA (see table 5.1) and the height of DNA when immobilized on PLO (~1.5 nm) (see section 5.2), a possible explanation was DNA folding in the z-plane.

It was possible that the ineffective immobilization of DNA was owing to the insufficient cations or shielding of DNA preventing the detection of DNA ends by the cantilever. There were very rare instances of faint DNA strands (*i.e.* DNA at a lower height than dsDNA) visible, which might have been single-stranded DNA fragments (see figure 5.11 c and e). This rare occurrence depicted the insufficiency of divalent cations to immobilize single-stranded DNA, because a large amount of ssDNA should have been visible. This might be due to the reduced anchoring of the single phosphate backbone in ssDNA as compared to 2 negatively charged phosphate-sugar chains in dsDNA. Some ends of the DNA strands were incompletely immobilised and tended to move during consecutive scans (figure 5.12), though the segments of DNA that did mobilize well depicted helical resolution with a pitch consistent with the B-form of DNA (figure 5.12 c). The average of the distances between prominent peaks on the cross-section of the DNA segment was 3.22 nm, which was close to the helical pitch of B-DNA. Therefore, it was concluded that this amount of Ca^{2+} could be used for immobilization of dsDNA sample as short as 100 bp, though it was not efficient for flap DNA as the single-stranded DNA was not immobilized. Hence, a higher concentration of Ca^{2+} needed to be used.

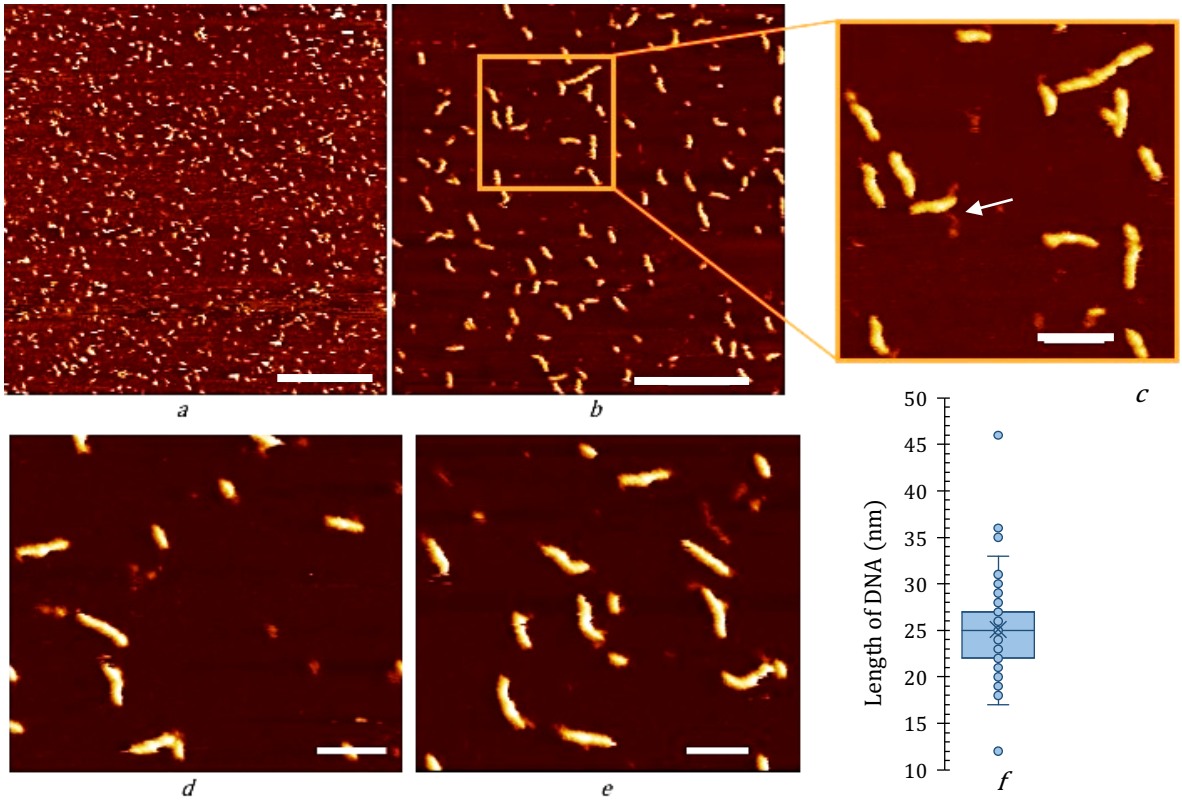


Figure 5.11: 100 nmole $CaCl_2$, 0.25 pmole flap DNA, imaging in water: a. scale bar 500 nm, height 7 nm; b. scale bar 200 nm, height 8 nm; c. scale bar 50 nm, height 8 nm; d and e. scale bar 50 nm, height 6 nm; f. length (nm) distribution of dsDNA section of flap DNA in water. Flap DNA in liquid looked like nicked DNA with almost no molecule showing flap. Some very rare molecules seen that do show the presence of ssDNA, which may or may not be flap DNA, like the ones marked by arrow in figure c.

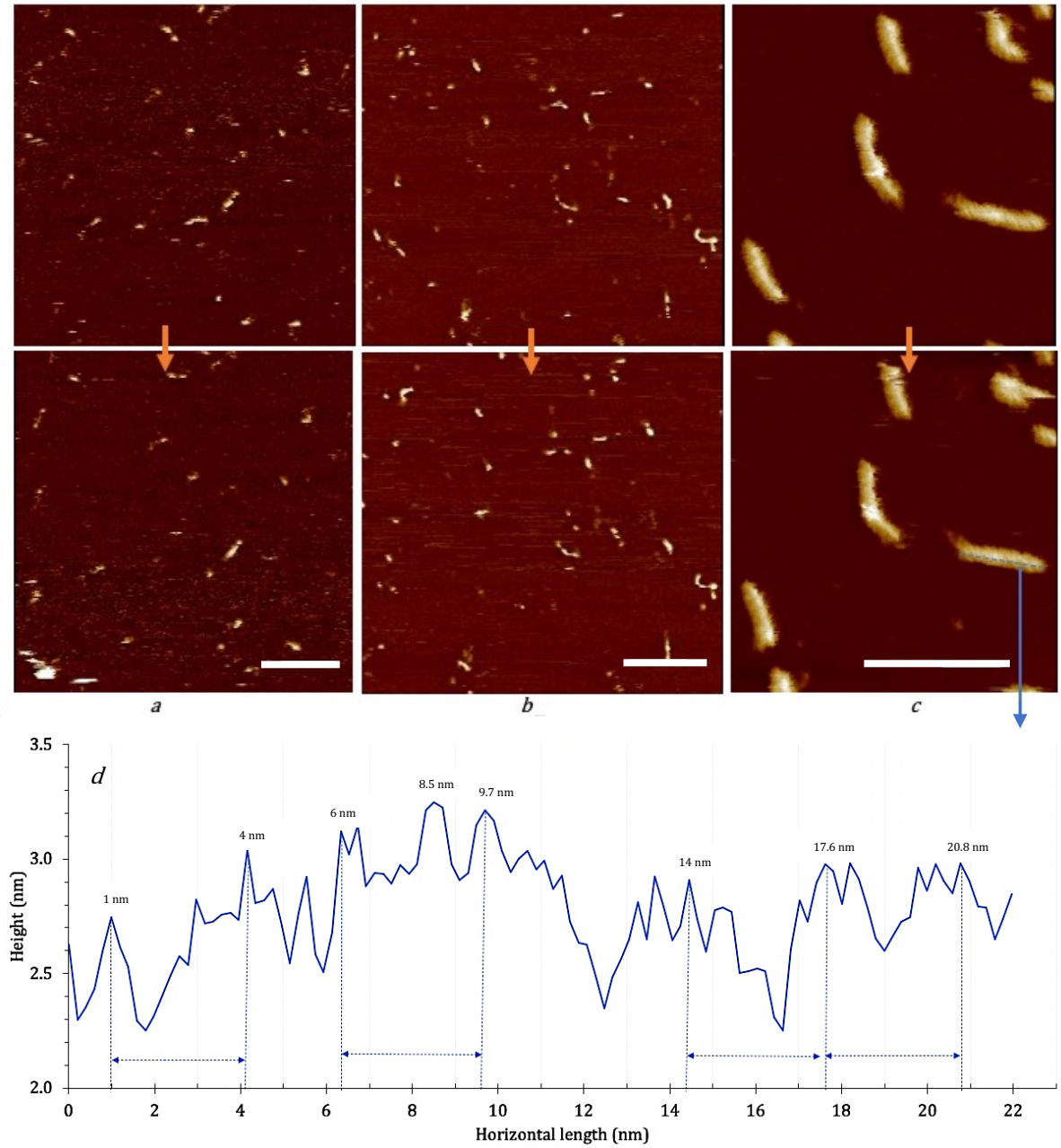


Figure 5.12: 100 nmole $CaCl_2$, 0.25 pmole flap DNA, imaging in water: a. scale bar 100 nm, height 6 nm; b. scale bar 100 nm, height 5 nm; c. scale bar 50 nm, height 6 nm. d. Cross-section of one of the DNA fragments from figure c. The numbers on the plot indicate the distance in nm between the peaks in the cross-section of DNA. The vertical lines indicate the difference between the consecutive peaks of the DNA corresponding to the phosphate backbone of the DNA in the double helix bordering the grooves.

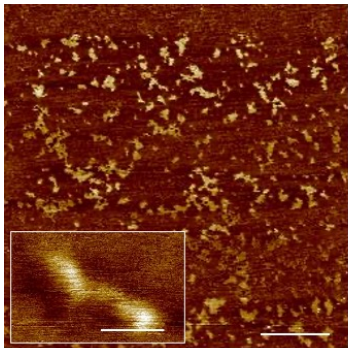


Figure 5.13: 200 nmole $CaCl_2$, 0.1 pmole flap DNA, imaging in air: scale bar 200 nm, height 1.1 nm; inside: scale bar 20 nm, height 1.2 nm. The DNA appear to be clumped, making it difficult to resolve the details. Very rare DNA molecules seen individually, like the one in the inside frame.

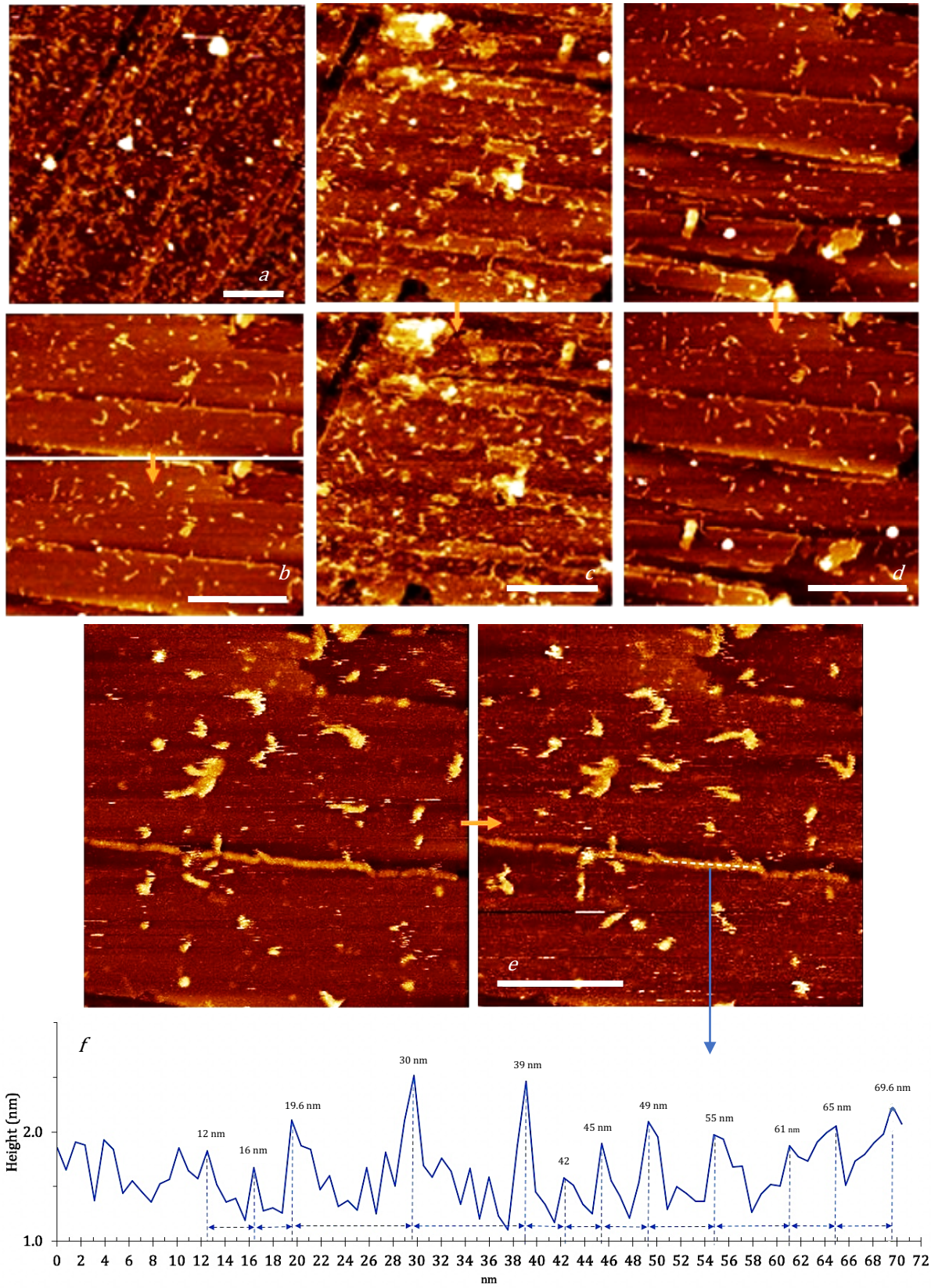


Figure 5.14: 200 nmole $CaCl_2$, 0.5 pmole flap DNA, imaging in water: a. scale bar 200 nm, height 12 nm; b. scale bar 200 nm, height 9 nm; c. scale bar 200 nm, height 10 nm; d. scale bar 200 nm, height 10 nm; e. scale bar 100 nm, height 9 nm. c and d are consecutive frames showing no movement of DNA. The images show gradients on the surface, that are probably the Ca^{2+} deposits along which the DNA had immobilized. f. Cross section of the DNA strands showing peaks at

distances corresponding to the pitch of B form of DNA. The peaks have a separation of the whole number multiples of the pitch value. The average of separation is calculated as 3.47 nm which is consistent with the B-form of DNA. The numbers on the plot indicate the difference in nm between the peaks in the cross-section of DNA.

When the amount of Ca^{2+} was increased to 200 nmole, 0.1 and 0.5 pmole of DNA were immobilized to study the effect of variation in the divalent cation to DNA ratio on the immobilization efficiency. Firstly, at this concentration, the Ca^{2+} could be seen deposited on the mica surface (figure 5.14) and contributed to background noise, thus setting the limit for amount of Ca^{2+} used for immobilization. This sample was dried under N_2 and would have probably been denser and more aggregated if allowed to dry in air. Hence, for dsDNA fragments of length ~ 100 bp, the Ca^{2+} should range above 100 nmole and well below 200 nmole for ~50 mm^3 mica surface. 0.1 pmole DNA tended to clump in the presence of such high amount of Ca^{2+} (figure 5.13) hence no structural features could be distinguished. There were only some very rare isolated DNA molecules seen. 0.5 pmole DNA, on the other hand, formed clumps in certain regions of the mica surface (figure 5.14 a) while in others showed deposits of Ca^{2+} along whose boundaries immobilized DNA fragments were aligned (figure 5.14 b-e). The fragments that were not aligned appeared to move in consecutive frames (figure 5.14 c-d), thus indicating inefficient immobilization in those regions. Also, the DNA along the boundary of Ca^{2+} deposits had a height slightly lower than the other fragments indicating stronger immobilization. The resolution in these cases was also better as the cantilever tip could detect the features of DNA without moving them. Some molecules could be resolved to show the major grooves of DNA with a pitch of 3.47 nm consistent with B-DNA, thereby signifying that the strands seen were indeed DNA molecules aligned close to their ends. Nonetheless, the ends of individual DNA fragments could not be distinguished and there was neither a flap structure nor a single-stranded DNA fragment seen.

Thus, the ~100 bp DNA could be immobilized on Ca^{2+} treated mica with the ion quantity >100 nmole and < 200 nmole, but this sample immobilization method could not be applied to ssDNA fragments as short as 50 nt. Since Ni^{2+} and Mg^{2+} have been used confidently to immobilize longer DNA (~ 300 bp)[15], immobilization by these cations was tried in the following experiments (section 5.3.3-5). The aim here forth was to compare and standardize the concentration of divalent cations required to immobilize 50 nt long ssDNA branch next to 100 bp dsDNA fragment, *i.e.* the flap DNA of interest.

5.3.2. Nicked DNA on Ca^{2+} treated mica

The experimental results obtained above with flap DNA were attempted to be repeated with nicked DNA. The mica discs used had an area of ~50 mm^3. When 0.25 pmole of nicked DNA was immobilized on 25 nmole Ca^{2+} treated mica, the amount of DNA immobilized was very low. The DNA that did immobilize was not very efficiently stuck and displayed movements during consecutive scans (see figure 5.15). The DNA length observed was also not long enough and limited to ~ 25 nm.

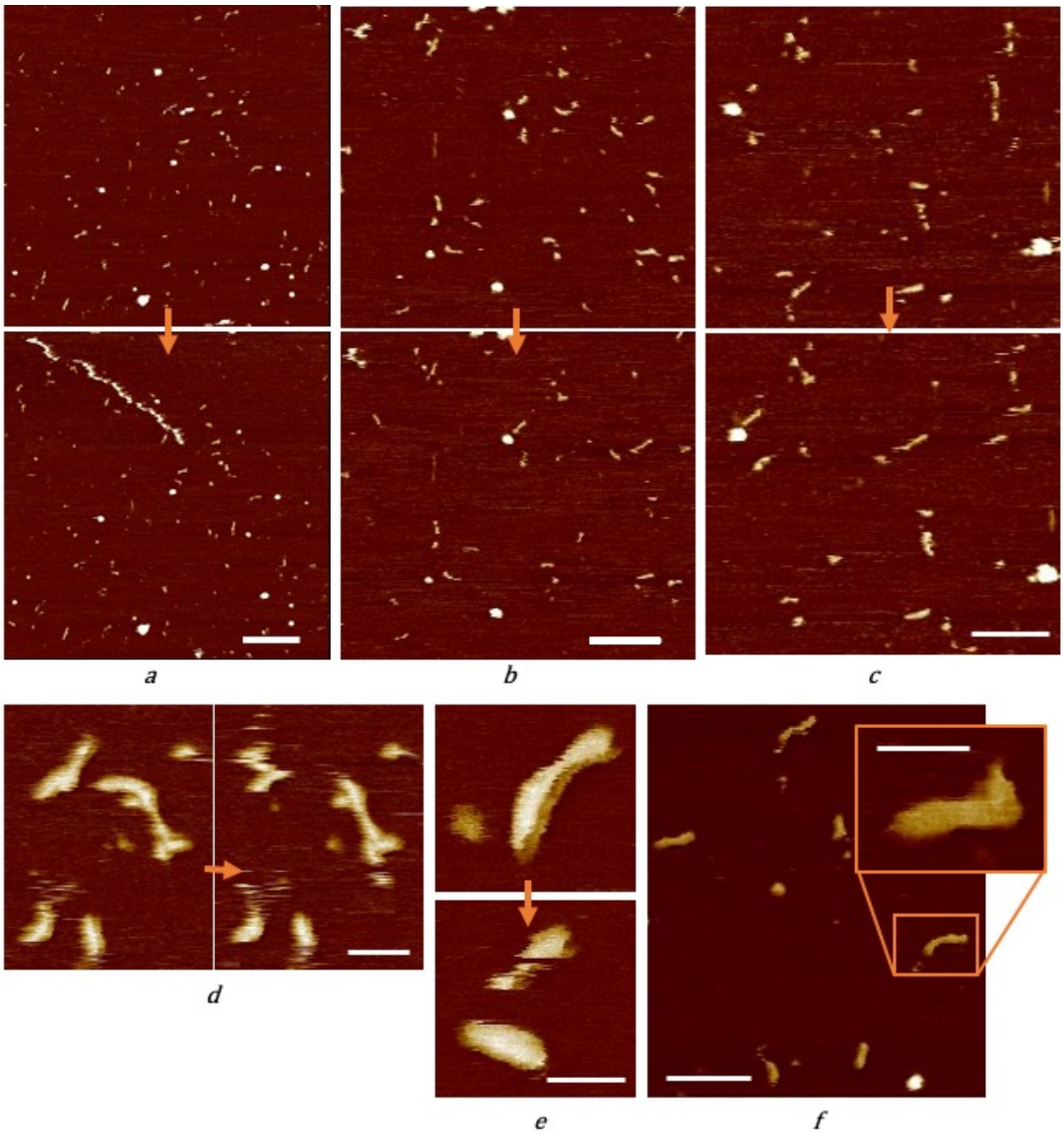


Figure 5.15: 25 nmole $CaCl_2$, 0.25 pmole nicked DNA (imaging in water): a. scale bar 200 nm, height 5 nm; b. scale bar 200 nm, height 5 nm; c. scale bar 100 nm, height 5 nm; d. scale bar 50 nm, height 5 nm; e. scale bar 20 nm, height 5 nm; f. scale bar 100 nm, inside scale: 20 nm, height 6 nm. figure e shows movement of DNA in the consecutive frames. In f, the DNA in the inset shows a zoom in of the DNA strand indicated.

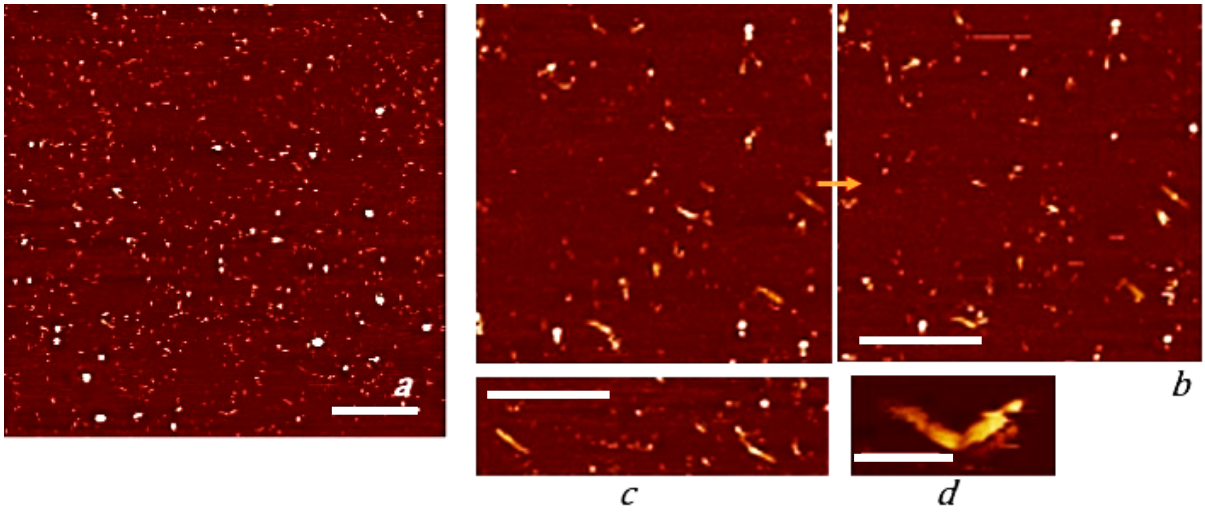


Figure 5.16: 50 nmole $CaCl_2$, 0.25 pmole nicked DNA (imaging in water): a. scale bar 400 nm; b. scale bar 200 nm; c. scale bar 200 nm; all heights 5 nm; d. scale bar 50 nm, height 6 nm.

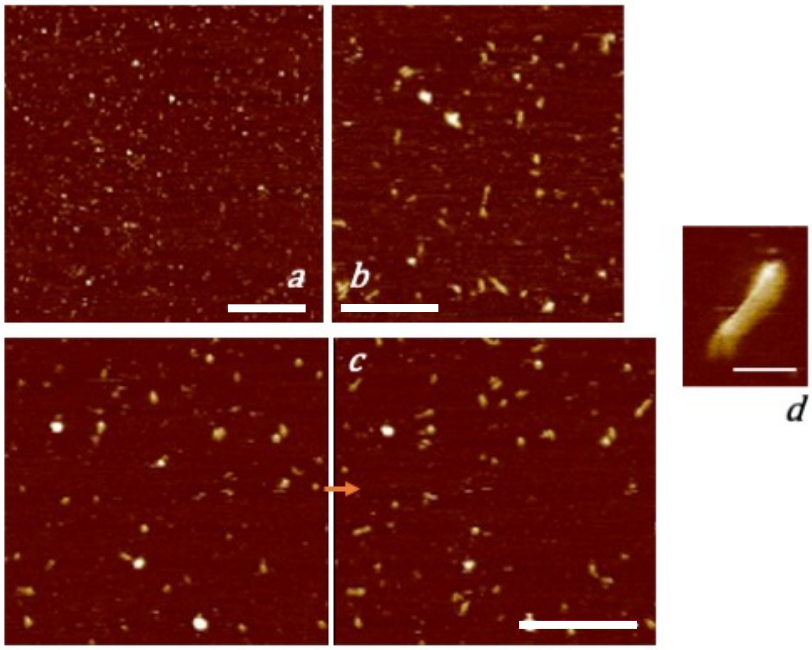


Figure 5.17: 100 nmole $CaCl_2$, 0.25 pmole nicked DNA (imaging in water): a. scale bar 400 nm; b. scale bar 200 nm; c. scale bar 200 nm; d. scale bar 20 nm. all heights 5 nm.

As already mentioned in section 5.3.1, this might be due to inefficient immobilization of DNA at the ends. This indicated that the Ca^{2+} quantity was not sufficient, hence was increased to 50 nmole (figure 5.16). In this case as well, the number of DNA immobilized were very few, moved in consecutive scans and displayed low resolution due to movement of the sample while being scanned by the cantilever. Hence, the concentration was increased further to 100 nmole (figure 5.17) and 150 nmole (figure 5.18) of Ca^{2+} but that too remained ineffective for the same above-mentioned reasons.

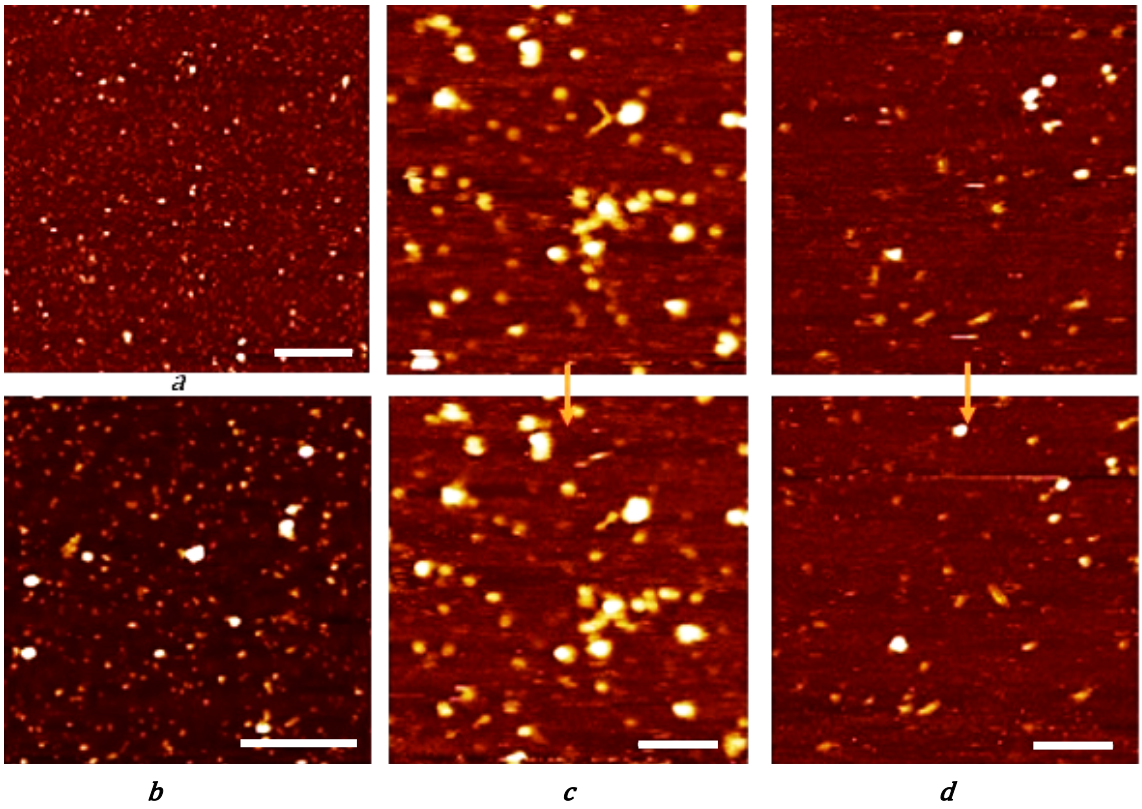


Figure 5.18 : 150 nmole $CaCl_2$, 0.25 pmole nicked DNA (imaging in water): a. scale bar 400 nm, height 10 nm; b. scale bar 200 nm, height 5 nm; c. scale bar 100 nm, height 5 nm; d. scale bar 50 nm, height 7 nm. A lot of noise observed, which is probably the excessive Ca^{2+}, and extremely few DNA molecules seen.

Additionally, in experiments with flap DNA and nicked DNA immobilization on Ca^{2+}(section 5.3.1-2), the height of the DNA fragments imaged in water was larger (> 3 nm) than the height of DNA immobilized on PLO (height of <1.5 nm (see section 5.2)) and the theoretical width of the dsDNA[3] (see table 5.1). This, along with the short lengths of DNA (~25 nm vs the expected ~35 nm, figure 5.11f), indicated that the DNA had folded in the z plane. Also, for 150 nmole of Ca^{2+} (figure 5.18), there were a lot of 'blobs' seen on the surface which were either DNA molecules coiled into globules or Ca^{2+} precipitates with no DNA present. The latter could be due to the shielding of DNA molecules in the presence of high salt concentration which caused the DNA to not stick to the surface at all. In both the cases, the high amount of Ca^{2+} posed hinderance to the imaging of DNA.

Hence, it was concluded that Ca^{2+} were ineffective to immobilize flap and nicked DNA because at low quantities of the salt, the DNA was not immobilized or was partially immobilized so as to compromise on the spatial resolution of the images. At high quantities of the salt, flap DNA was either clumping together or aligning end to end along prominent deposits of Ca^{2+} ions,

while the nicked DNA was either rolling into 'blobs' or not immobilizing at all. In both the samples, flap and nicked DNA, the height of the DNA was more than the theoretical value or the height obtained by immobilization on PLO. Ca^{2+} between 150 nmole and 200 nmole could be assumed to be effective for the imaging but further experimentation is required to establish the 'sweet-spot' of Ca^{2+} suitable for efficient immobilization of dsDNA of length 100 bp.

5.3.3. Nicked DNA on Mg^{2+} treated mica

Second to Ni^{2+}, Mg^{2+} are the most commonly used ions for immobilization of DNA molecules[9,17–19]. Building on the experiments with Ca^{2+}, nicked DNA molecules were immobilized on Mg^{2+} ranging between 200 and 2000 nmole. 0.5 pmole DNA immobilized on 200 nmole Mg^{2+} (figure 5.19 a-b) did not result in any DNA immobilization on the surface, while 1 pmole of DNA did lead to very scarce DNA molecule immobilization seen (figure 5.19 e-h). The molecules, when present at all, moved through consecutive scans (figure 5.19 g-h) and displayed lengths very short compared to the expected lengths of 100 bp DNA. This was possibly due to inefficiency of Mg^{2+} concentration for immobilization. At 400 nmole of Mg^{2+} as well (figure 5.19 c-d), the DNA molecules remained absent from the surface. There were scantily immobilized molecules on the surface treated with 1000-1820 nmole of Mg^{2+} as well (figures 5.20-21), were not completely immobilized, moved during scans, and displayed sizes shorter than the theoretical value of 35 nm.

Hence, it can be concluded that Mg^{2+} was not suitable for immobilization of 100 bp DNA for AFM imaging. To explore if such short DNA fragments could be immobilized using divalent cations at all, Ni^{2+} was experimented with next, because of their success in immobilizing DNA fragments at least 200 bp long.

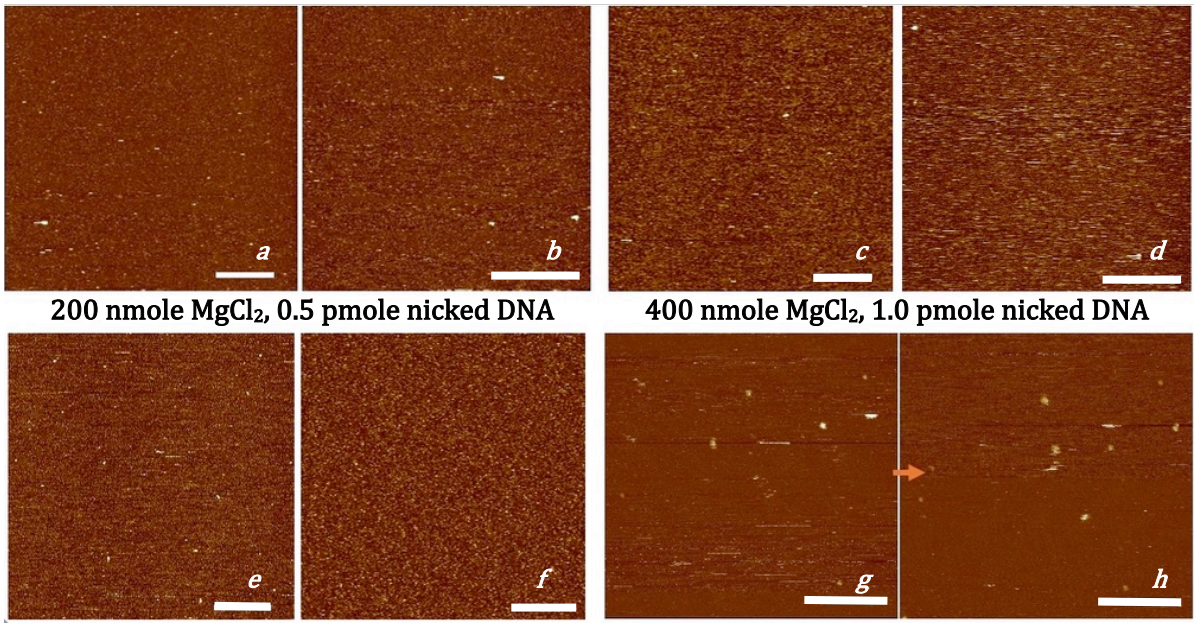


Figure 5.19 (imaging in water) : a and b. scale bar: 400 nm and 200 nm respectively, height 4 nm; c. scale bar 400 nm, height 2 nm; d.: scale bar 300 nm, height 4 nm; e. scale bar 400 nm, height 2 nm; f. scale bar 50 nm, height 6 nm; g and h. scale bar 100 nm, height 4 nm. No DNA molecules could be seen immobilized on the surface.

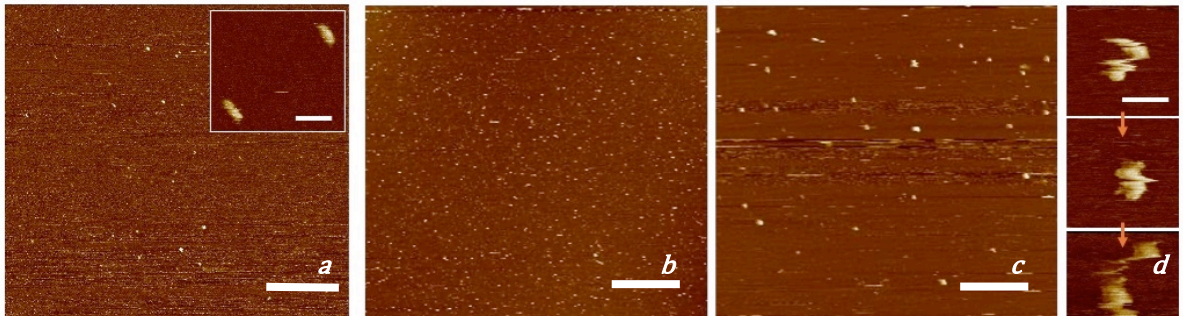
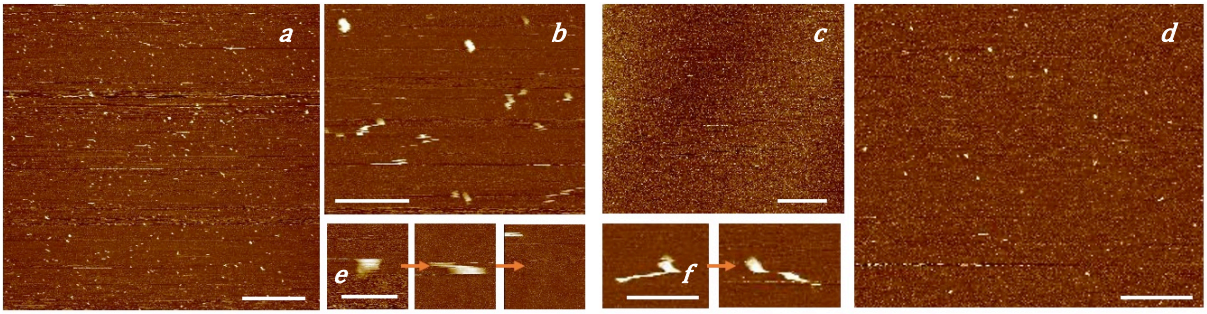
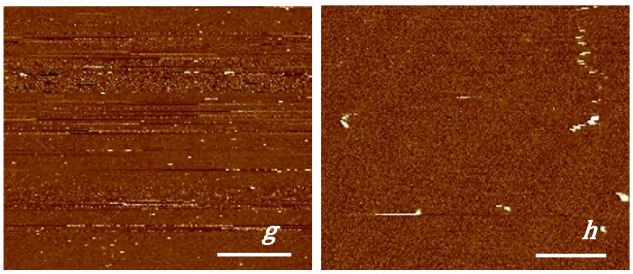


Figure 5.20 (imaging in water) : a. 1000 nmole $MgCl_2$, 1 pmole nicked DNA: scale bar: 700 nm, height 6 nm; in the inset: scale bar 50 nm, height 7 nm; 1000 nmole $MgCl_2$, 2 pmole nicked DNA: b. scale bar 2 µm, height 8 nm; c. scale bar 400 nm, height 8 nm; d. scale bar 30 nm, height 7 nm. No DNA molecules could be seen immobilized on the surface.



1000 nmole $MgCl_2$, 4.0 pmole nicked DNA **1365 nmole $MgCl_2$, 5.0 pmole nicked DNA**



1820 nmole $MgCl_2$, 7.5 pmole nicked DNA

Figure 5.21 (imaging in water):

1000 nmole $MgCl_2$, 4.0 pmole nicked DNA: a. scale bar 700 nm, height 8 nm; b. scale bar 200 nm, height 8 nm; e. scale bar 20 nm, height 6 nm;

1365 nmole $MgCl_2$, 5.0 pmole nicked DNA: all height 2 nm; c. scale bar 2 µm; d. scale bar 400 nm; e. scale bar 60 nm;

1820 nmole $MgCl_2$, 7.5 pmole nicked DNA: scale bar 200 nm, g.., height 4 nm; h. height 2 nm.

Images e and f show the consecutive frames depicting the motion of DNA during continuous imaging. No DNA molecules could be seen immobilized on the surface.

5.3.4. Flap DNA on Ni^{2+} treated mica

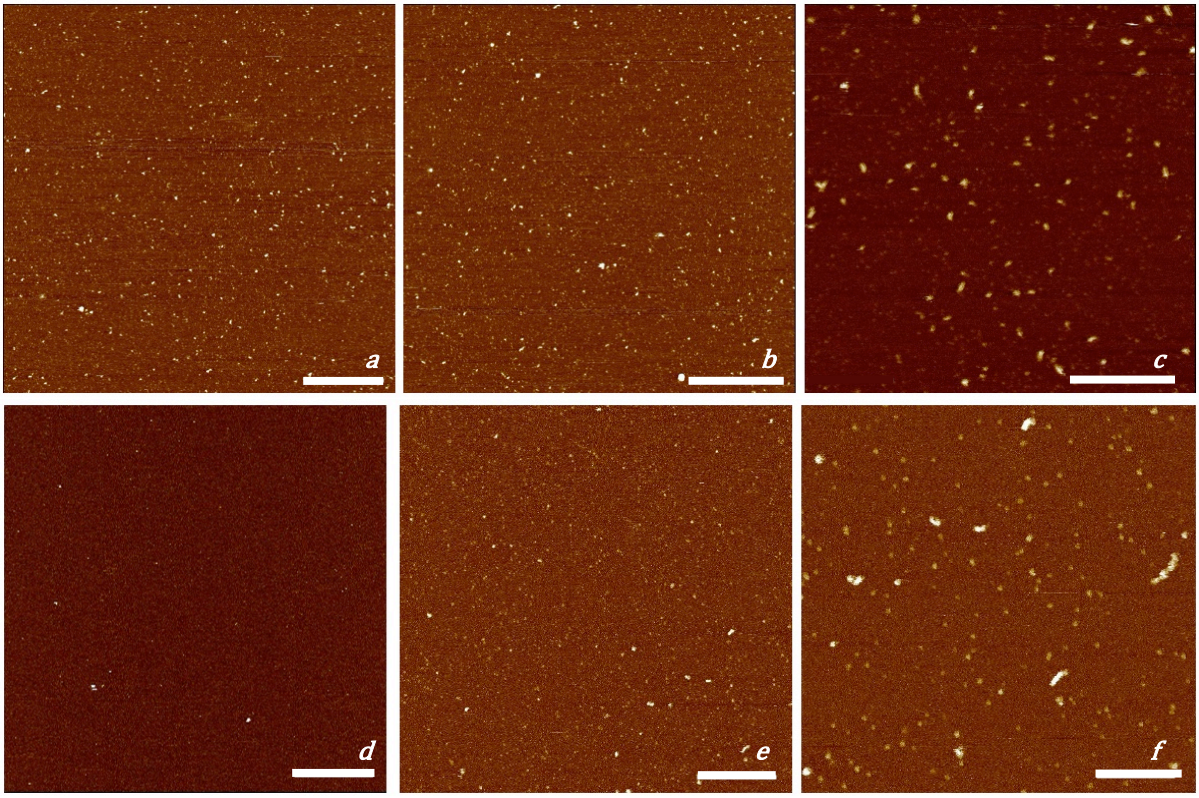


Figure 5.22 : 2 µmole $NiCl_2$, 0.5 nmole flap DNA (imaging in water): a. scale bar 400 nm, height 4 nm; b. scale bar 400 nm, height 4 nm; c. scale bar 200 nm, height 5 nm; d. scale bar 700 nm, height 5 nm; e. scale bar 200 nm, height 4 nm; f. scale bar 100 nm, height 4 nm. No DNA molecules could be seen immobilized on the surface.

Imaging of DNA sample immobilized on Ni^{2+} is a well-established methodology for AFM[15]. The method has been successfully implemented for DNA molecules at least 300 bp long, while the DNA sample in our case was 100 bp long. Hence, the efficiency of the immobilization method needed to be tested. Expanding on the results of immobilization using Ca^{2+} and Mg^{2+} in the previous sections, the quantity of Ni^{2+} used was 2 and 4 µmole with 0.5 nmole of DNA, which was already very high compared to the amount used for PLO, Ca^{2+} or Mg^{2+} aided immobilization. Yet the amount of DNA immobilized was very scarce.

With 2 µmole of Ni^{2+}, the quantity of DNA molecules seen was very low (figure 5.22), considering the high concentration of DNA used. The consecutive scans did not show any movement of immobilized molecules, thereby implying that the immobilization, though scarce, was efficient. The DNA was shorter than the expected length of 35 nm (figure 5.22 f) with no flap or ssDNA segments seen at all.

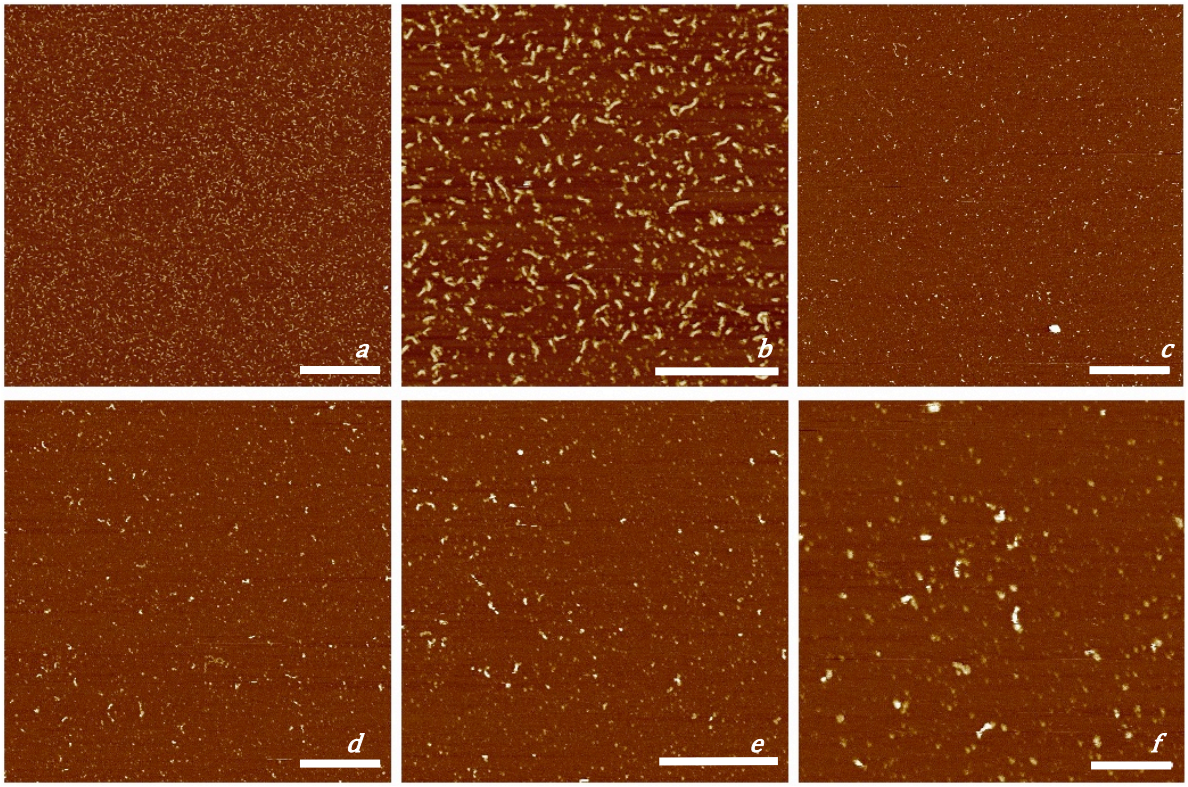


Figure 5.23 : 4 µmole $NiCl_2$, 0.5 nmole flap DNA (imaging in water): a. scale bar 400 nm, height 6 nm; b. scale bar 200 nm, height 6nm; c. scale bar 400 nm, height 4 nm; d. scale bar 200 nm, height 4 nm; e. scale bar 200 nm, height 4 nm; f. scale bar 20 nm, height 4 nm.

At a higher Ni^{2+} quantity of 4 µmole, 0.5 nmole of DNA were immobilized and a good concentration of DNA could be seen (figure 5.23 a-b). The lengths of DNA were similar to the theoretical value of lengths and the molecules did not move around during consecutive scans. Hence, this method could be considered suitable for imaging DNA. Unfortunately, the image results could not be reproduced during the following experiments (figure 5.23 c-f) where the quantity of DNA was very low, even after numerous attempts. This could have been because of some fortunate error in the immobilization method in the first case that caused perfect immobilization, that could not be reproduced in the later images. Yet, comparably higher DNA number were immobilized on this concentration of Ni^{2+} than immobilization with 2 µmole of Ni^{2+}.

Hence, immobilization using Ni^{2+} was ineffective for flap and nicked DNA sample, probably due to the short length of DNA. A higher quantity of Ni^{2+} were nevertheless experimented with using nicked DNA samples.

5.3.5. Nicked DNA on Ni^{2+} treated mica

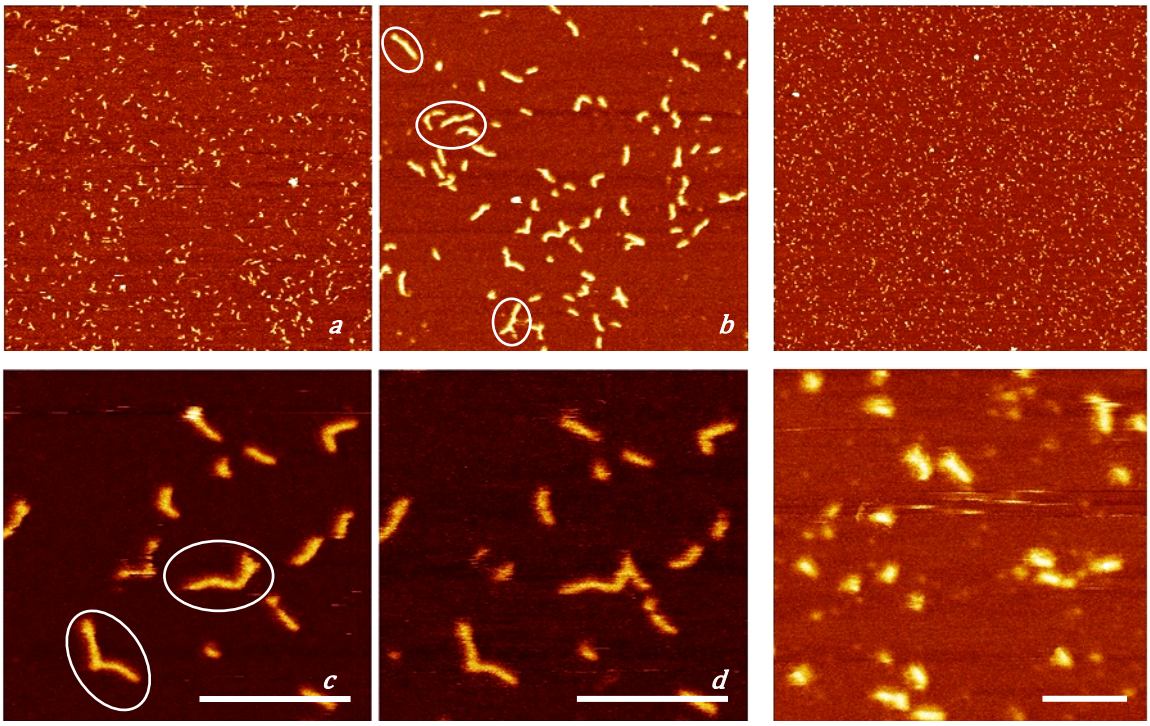


Figure 5.24 : 2 µmole $NiCl_2$, 0.5 nmole nicked DNA (imaging in water): height 6 nm, scale bar: a. 400 nm, b, c and d. 100 nm.

Figure 5.25 (imaging in water): 20 µmole $NiCl_2$, 0.5 nmole nicked DNA: height 6 nm. scale bars 400 and 50 nm.

The experimental results obtained with flap DNA were attempted to be replicated with nicked DNA. The mica disc used here was bigger, hence the amount of $Ni^{2+}$and DNA was increased to accommodate the size of the mica disc. In this case, like flap DNA, the quantity of DNA immobilized was adequate for 0.5 nmole of DNA on 2 µmole (see figure 5.24) and 20 µmole (see figure 5.25) of Ni^{2+}. On 2 µmole of Ni^{2+}, few molecules of DNA tended to move during consecutive scans, while a large percentage of DNA remained immobilized (figure 5.24 c and d). Also, in these images, the resolution was good enough to faintly reveal grooves on the DNA. The lengths of some of the DNA molecules, though, was greater than the expected 35 nm length of 100 bp DNA (encircled in figure 5.24). This could be because of the molecules being very close to each other such that it became difficult to distinguish them apart.

At a higher concentration of Ni^{2+}(figure 5.25) however, lengths of the DNA fragments were smaller than the expected/theoretical lengths indicating that the DNA either tended to break or were partially immobilized. The DNA molecules didn't display any movement in the consecutive scans hence the shortness of the lengths was probably not because of partial

immobilization, but due to molecules breaking. Also, there were no DNA helical grooves resolved. These could be due to the shielding of the sample by the high amount of Ni^{2+} present and/or the repulsive forces between the cantilever tip and the ionised sample surface. These reasons might also explain the shortness of the DNA strands *i.e.* the tip not being able to detect a segment of the molecule.

Hence, Ni^{2+} appeared to be the most feasible divalent cation for immobilization of DNA strands as short as 100 bp with good helical resolution and no mobility of DNA molecules during consecutive scans. The quantity of Ni^{2+} could be ~ 2 µmole with 0.5-1.0 nmole of DNA sample. However, Ni^{2+} are not the right choice for flap DNA as the ssDNA flap was not visible (section 5.3.4) and are also hazardous to work with. Nevertheless, these experiments demonstrated that immobilization by Ni^{2+} is possible for 100 bp DNA imaging alone, even if these are not the preferred choice for DNA-protein interaction study. If the immobilization of DNA had been effective using Mg^{2+} and Ca^{2+}, these would have been the more desired choice for DNA immobilization to study the interaction with protein because they provide the additional benefit of acting as cofactors for the binding and/or activity of most of the enzymes and would thus ensure efficient interaction.

5.4. Conclusion

The above experimental results revealed that 100 bp long dsDNA molecules could be immobilized on PLO treated mica and produced good resolution AFM images with helical grooves visible during imaging in liquid. When imaging in air, flap DNA could be seen as a Y-shaped structure with a higher strand and fainter ssDNA. Yet, PLO treated mica as surface for immobilization was not adequate for visualization of such short ssDNA fragments branching from dsDNA. In liquid imaging conditions the height of the dsDNA was about 3 times the height of the ssDNA. When the latter was very short (as was this case, 50 nt ssDNA), it could not be resolved next to the very distinct dsDNA. The single strand appeared as a very faint strand more or less resembling a blob in most cases that could be mistaken for the protein or background noise during the future dynamic imaging experiments. The imaging would possibly be made simpler, more efficient and fruitful if the length of the DNA strands (both the double-stranded section and the ssDNA flap) was greater, at least 300 bp for dsDNA and ~ 100 nt for ssDNA, consistent with the results reported elsewhere[15].

In case of nicked DNA fragments, the images in air revealed the dumbbell shape of DNA. The neck indicated the break in one phosphodiester bond on one of the DNA strands of the double helix. When these molecules were imaged in liquid, the dumbbell shape was not seen but the DNA appeared as a continuous strand. This was probably because of widening of the bond break in the dehydrated form of DNA where the gap appeared to be more prominent due to the constrained conformation of the DNA strand that was broken. It could be reasoned that the DNA, though immobilized in the B conformation as in the solution in which it was immobilized, was free to conform to the wider A-form (the preferred conformation for dehydrated DNA) for a few bp close to the DNA strand break on the strand of DNA containing the nick. This allowed the DNA strands neighbouring the nick to constrain, hence widening the bond-break and giving the 'dumbbell' shape. In case of imaging in liquid, however, the DNA was more relaxed due to it being in the hydrated B-conformation, hence the broken strand of the DNA was not pulled apart to reveal the gap.

The lengths of DNA immobilized on PLO treated mica varied when imaging in air vs liquid due to the possible inefficient immobilization of the molecule at the ends. When immobilized on divalent cation treated mica, the molecules appeared to be shorter due to folding of DNA molecules in the z-plane or immobilization at the ends. The DNA molecules did not immobilize at lower concentration of divalent cations and could be seen as blobs or strands that moved around in the consecutive frames. Yet, some instances where the DNA were immobilized well, a good helical resolution could be observed where the molecules showed the major grooves of DNA. But, even in the images of the best dsDNA resolution, there were no or very scarce ssDNA seen, signifying the immobilization technique was not suitable for flap DNA. On the contrary, in a high salt concentration, the ions precipitated on the mica surface causing a large amount of background noise that would interfere with locating the protein in forthcoming experiments aimed to study DNA-protein interactions. Ca^{2+} and Ni^{2+} proved to be more effective than Mg^{2+} for immobilization of dsDNA, with Ni^{2+} showing the least amount of mobility of DNA in the consecutive frames. Further experiments could be pursued to achieve the 'sweet spot' of ion concentration for immobilization of DNA molecules as short as 100 bp, while longer strand length could make the experimental results more feasible.

Consequently, even though short DNA offered the advantage of easy manipulation of molecular configuration, this methodology was disregarded for future dynamic imaging experiments, as longer DNA with longer ssDNA flap would be easier to anchor to the surface

and observe due to their greater prominence of longer strands. The next set of experiments were therefore aimed to construct and image DNA molecules with a suitable flap length that could interact with FEN while allowing easy AFM imaging to locate the flap and resolve its conformation changes during protein interaction.

Bibliography

1. Ando T, Uchihashi T, Kodera N. High-Speed AFM and Applications to Biomolecular Systems. Annu Rev Biophys. 2013.
2. Ando T, Uchihashi T, Scheuring S. Filming biomolecular processes by high-speed atomic force microscopy. Chem Rev. 2014;114:3120−3188.
3. James D. Watson *et al*. The Structure of DNA. In: Molecular Biology of the Gene. 7th ed. Pearson; 2014:77-105.
4. Lee AJ, Szymonik M, Hobbs JK, Wälti C. Tuning the translational freedom of DNA for high speed AFM. Nano Res. 2015;8(6):1811-1821.
5. Gaczynska M, Osmulski PA. AFM of biological complexes: What can we learn? Curr Opin Colloid Interface Sci. 2008;13:351–367.
6. Ando T. High-speed atomic force microscopy coming of age. Nanotechnology. 2012;23:062001.
7. Japaridze A *et al*. Toward an Effective Control of DNA's Submolecular Conformation on a Surface. Macromolecules. 2016;49:643−652.
8. Poly-L-Ornithine Solution (0.01%) | A-004-C. https://www.merckmillipore.com/GB/en/product/Poly-L-Ornithine-Solution-0.010-0,MM_NF-A-004-C?bd=1#overview.
9. Hansma HG, Bezanilla M, Zenhausern F, Adrian M, Sinsheimer RL. Atomic force microscopy of DNA in aqueous solutions. Nucleic Acids Res. 1993;21(3):505-512.
10. Santos S, Barcons V, Christenson HK, Font J, Thomson NH. The Intrinsic Resolution Limit in the Atomic Force Microscope: Implications for Heights of Nano-Scale Features.
11. Hansma HG, Revenko I, Kim K, Laney DE. Atomic force microscopy of long and short double-stranded, single-stranded and triple-stranded nucleic acids. Nucleic Acids Res. 1996;24(4):713-720.
12. John S *et al*. Direct Visualization of Dynamic Protein-DNA Interactions with a Dedicated Atomic Force Microscope. Biophys J. 1998;74(6):2840-2849.
13. Guo F *et al*. Imaging of DNA Molecules by Atomic Force Microscope. ieee. 2017:315-318.
14. Zheng J, Li Z, Wu A, Zhou H. AFM studies of DNA structures on mica in the presence of alkaline earth metal ions. Biophys Chem. 2003;104(1):37-43.
15. Hansma HG, Laney DE. DNA Binding to Mica Correlates with Cationic Radius: Assay by Atomic Force Microscopy. Biophys J. 1996;70(4):1933-1939.
16. Feng M *et al*. Roles of divalent metal ions in flap endonuclease-substrate interactions. Nat Struct Mol Biol. 2004;11:450-456.
17. Vesenka J, Guthold M, Tang CL, Keller D, Delaine E, Bustamante C. Substrate preparation for reliable imaging of DNA molecules with the scanning force microscope. Ultramicroscopy. 1992:1243-1249.
18. Coury JE, Anderson JR, McFail-Isom L, Williams LD, Bottomley LA. Scanning force microscopy of small ligand− nucleic acid complexes: tris (o-phenanthroline) ruthenium (II) as a test for a new assay. J Am Chem Soc. 1997;119(16):3792-3796.
19. Fang Y, Hoh JH, Spisz TS. Ethanol-induced structural transitions of DNA on mica. Nucleic Acids Res. 1999;27(8):1943-1949.

Chapter 6

AFM of Longer Flap/Overhang DNA

6.1. Introduction

The flap DNA molecules imaged so far were possibly too short to be immobilized well for AFM imaging. They posed the problems of firstly, being difficult to resolve the DNA helix and secondly, having the single-stranded flap very short to be immobilized at all. Hence, it was necessary to explore the possibility of improvement of DNA immobilization by using longer DNA. For this, three potential methods were applied:

i. synthesis of ~ 300 bp long overhang DNA fragments by assembly of PCR products,

ii. synthesis of single-stranded M13 DNA followed by polymerization in the absence of the exonuclease fragment of DNA polymerase (*i.e.* the klenow segment alone) to produce ~3 kb closed dsDNA with a single-stranded flap of unknown length, and

iii. assembly of long DNA oligonucleotides (80-100 nt) to produce 300 bp dsDNA with 130 nt single-stranded flap.

In this chapter, the above-mentioned methods have been discussed as a necessary step to explore the DNA synthesis techniques for obtaining a suitable substrate for FEN interaction using AFM.

6.2. Overhang DNA from Assembly of PCR Products

Flap endonuclease recognises the 5' end of a ssDNA strand to bind and perform nuclease activity. The activity of the enzyme has been known to vary with respect to the kind of 5' end present. DNA with a 5' flap competes 4-fold more efficiently than a DNA with 5' overhang for FEN binding. Also, the efficiency of FEN binding to the 5' end of ssDNA was affected by the proximity of the 3' DNA at the branching point/elbow of flap DNA. This means that the FEN would not be very efficient in cutting flap DNA if there is a large gap on the dsDNA at the branch point[1]. But there would be binding and activity seen. Hence it would be possible to demonstrate FEN activity on dsDNA having a 5' overhang. Two ssDNA fragments of the required length could be synthesized by asymmetric PCR such that they are complementary in the middle and leave 5' overhangs of desired length at both ends when annealed.

As described in section 4.2.2 and figure 4.4, two sets of primers were designed for template pJONEX with tFENΔ19 insert. The dsDNA produced as a result of polymerization were partially complementary to each other and would produce overhangs when allowed to cross anneal. PCR was performed to produce dsDNA fragments of 288 bp and 257 bp length, followed by 2 sets of asymmetric PCR using single primers such that the ssDNA produced as a result could anneal to produce 5' overhangs and not 3'. The overhang DNA sample thus produced was electrophoresed on an agarose gel, which revealed a very low concentration of the desired overhang product despite a good concentration of the double-stranded PCR products formed (see figure 6.1). Additionally, a low amount of long DNA product (~ 1 kb) was produced probably as a result of random annealing of the ssDNA.

The DNA product when analysed using AFM (figure 6.2) presented DNA molecules that were longer than the expected 183 bp length of dsDNA (~65 nm length) and most importantly, did not have any ssDNA overhangs visible. This indicated that the efficiency of asymmetric PCR was very low and there was a substantial percentage of ssDNA utilized to produce the unrequired longer dsDNA product (~ 1 kb). Also, there was a considerable amount of noise in the AFM images, probably due to the presence of buffers used for DNA purification. Modification of the PCR parameters and concentrations did not improve the yield or quality of the overhang DNA (figure 6.3) as numerous complex arrangements of ssDNA to produce a variation of dsDNA structures were observed. A very small percentage of DNA molecules observed could have been of the conformation required, but in those cases, it was not possible

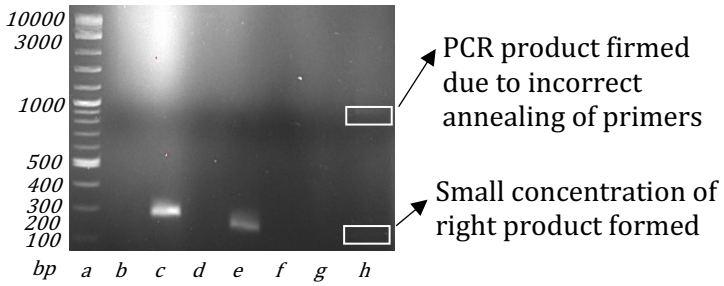


Figure 6.1: DNA electrophoresed on agarose gel: a. 2 log DNA ladder, b. -ve control for PCR for 288 bp fragment, c. PCR product for same, d. -ve control for PCR for 257 bp fragment, e. PCR product for same, f. -ve control for asymmetric PCR for 288 nt fragment, g. -ve control for asymmetric PCR for 257 nt fragment, h. DNA fragments after annealing of asymmetric PCR products.

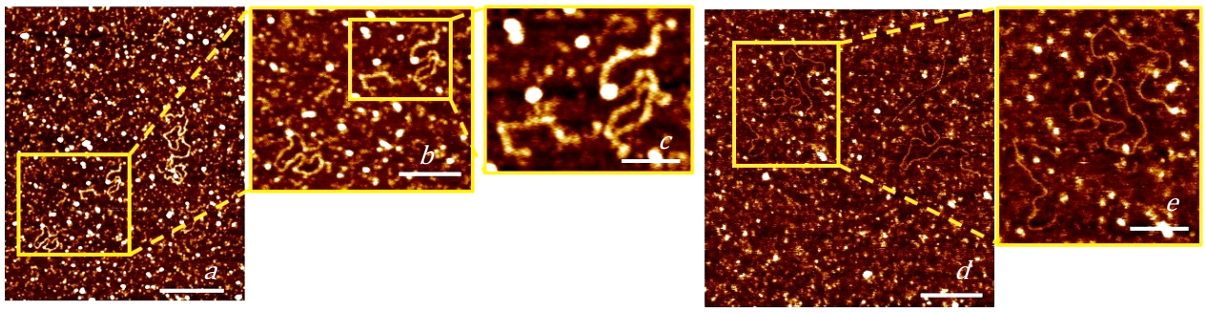


Figure 6.2: Overhang DNA :
imaging in air: a. scale bar 200 nm, height 1 nm, b. scale bar 100 nm, height 1.2 nm, c. scale bar 50 nm, height 1 nm.
imaging in water (height 2.5 nm): d and e: scale bar 100 nm and 50 nm.
The DNA were assembled into complex structures that did not have any clear ssDNA overhang.

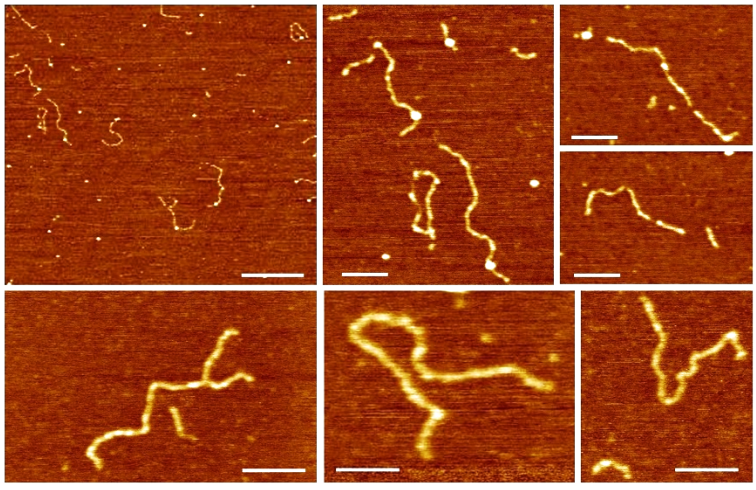


Figure 6.3: Overhang DNA, imaging in air, height 1 nm: a. scale bar 400. b-g scale bar 100 nm. Various arrangements of ssDNA produce variation of complex dsDNA structures.

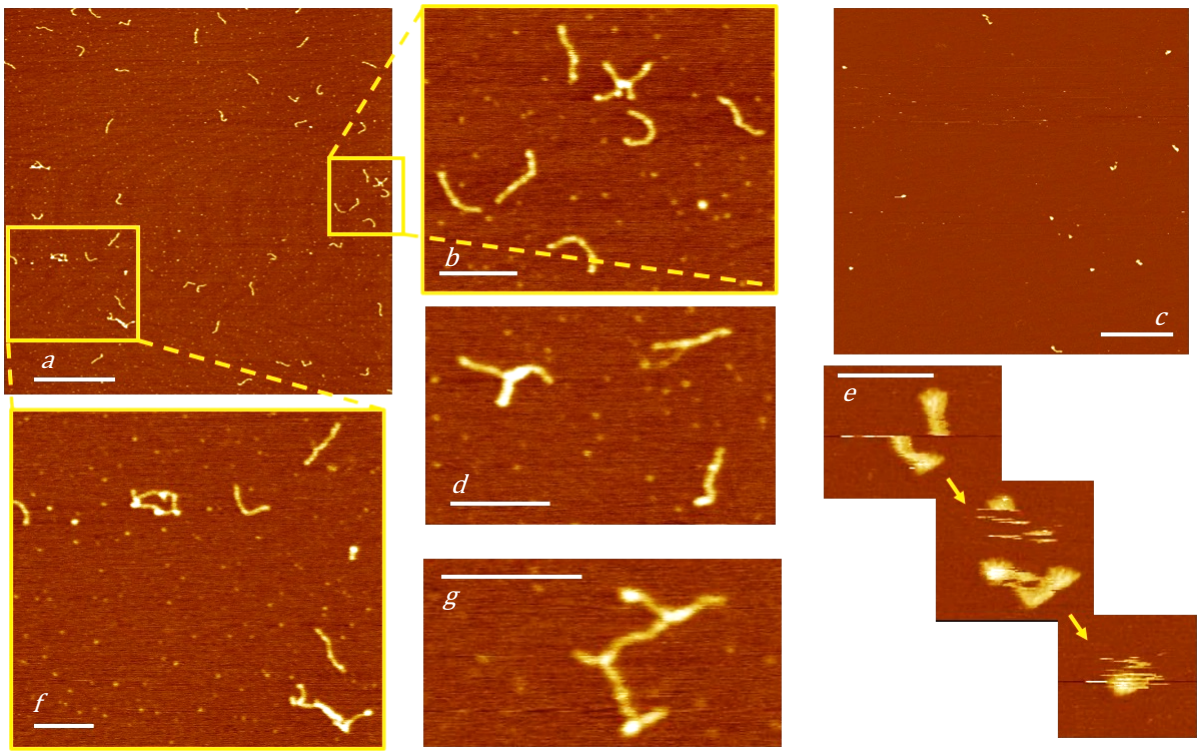


Figure 6.4: Overhang DNA: imaging in air, a,b,d,f,g: height 2 nm, scale bar a: 400 nm. b,d,f,g: 100 nm. **imaging in water**: height 8 nm, scale bar c: 300 nm, e: 50 nm, the images are consecutive and show the movement of DNA with continuous scans.
Various arrangements of ssDNA produce variation of complex dsDNA structures.

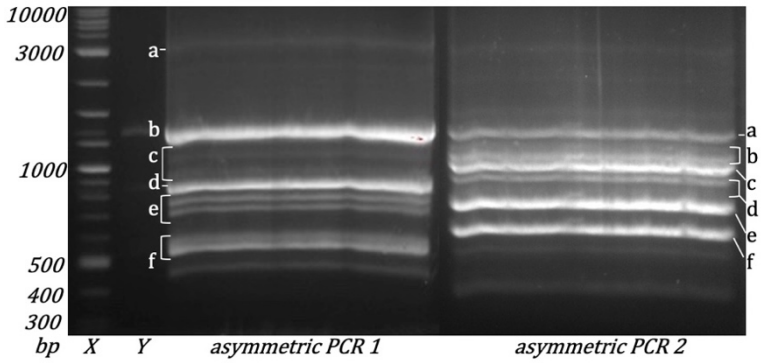


Figure 6.5: DNA electrophoresed on agarose gel: X. 2 log DNA ladder, Y. +ve control for PCR for asymmetric PCR, 1. asymmetric PCR for 288 nt DNA, 2. asymmetric PCR for 257 nt DNA.

to separate the various DNA arrangements in a narrow length/size range by electrophoresis. The gel elution technique too, posed the problem of producing a large amount of noise on the AFM images as the smallest of unwanted particles in the DNA sample were detected. Hence, the production of overhang DNA was attempted without the application of asymmetric PCR with the aim to increase the yield and percentage of overhang DNA among the undesired DNA structures. The observations made are discussed below.

The dsDNA products polymerised by PCR were mixed, denatured (95ºC/10 min), annealed (95ºC/10 min followed by cooling to 37ºC for 1 hr on a heat block/water bath) and immobilized onto PLO treated mica for AFM imaging. Though the yield of DNA was improved, the required DNA structure was still not achieved, and ssDNA cross linked to form complex DNA conformations (figure 6.4). Most importantly, a fraction of linear dsDNA conformations that were produced did not have the ssDNA overhang that was required. Hence, another method was attempted that included regulation of the ssDNA allowed to anneal.

To increase the amount of ssDNA that would anneal, dsDNA products produced after PCR were used to synthesize ssDNA by asymmetric PCR and purified on an agarose gel (figure 6.5). The gel images revealed that non-specific binding of primers had occurred that was the probable cause of the numerous conformations of dsDNA. The bands A-F of the 2 PCR products were eluted using a gel elution kit and mixed in such a way that their lengths varied and produced the annealed product that had an overhang because of the variation in the lengths of the ssDNA:

mix P=1_A+2_D, mix Q=1_B+2_E, mix R=1_C+2_F, mix S=1_D+2_A, mix T=1_E+2_B and mix U=1_F+2_C.

The mixtures P-U were heated to 95ºC for 10 minutes and cooled to 37ºC for 1 hour to allow annealing. The samples were then imaged using AFM (figures 6.6 and 6.7).

The AFM images revealed strands of dsDNA that yet again did not have any ssDNA overhangs, possibly due to the formation of loops at the DNA ends. This argument could be supported by the presence of blobs (of more height than the strands) at the ends of the DNA strands (figure 6.6 b, d, e, f encircled in white). Moreover, the DNA strands, in which the overhangs of different molecules had possibly aligned to form a long molecule, were seen as long strands with ‘beads’ in the middle indicating the cross annealing with the excess ssDNA looping or coiling to form the ‘bead’ (figure 6.6 c, f, i encircled in blue). Additionally, the molecules moved when attempted to image in water (figure 6.7). This made it difficult to observe the cross-linking of DNA molecules (observed in the images in air) with a helical resolution. Thus, the method did not prove to be successful to synthesize the desired DNA molecules with a flap or overhang structure.

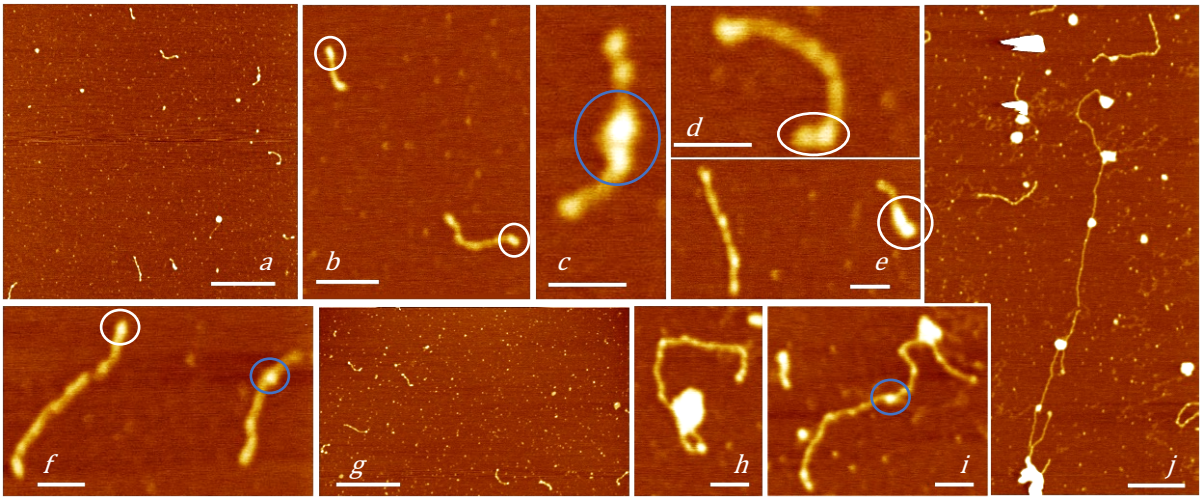


Figure 6.6: Overhang DNA: imaging in air: various combinations of gel elution product were mixed together and allowed to anneal to produce overhang DNA of varied length of ssDNA. The encircled segments of DNA show the section where the DNA had either overlapped or folded on top of one another.

mix P=1A+2D, mix Q=1B+2E, mix R=1C+2F, mix S=1D+2A, mix T=1E+2B and mix U=1F+2C

sample R (a-f): a height 2 nm, scale bar 400 nm, b. height 3 nm, scale bar 100 nm, c-f. height 3 nm, scale bar 50 nm. g.

sample S: height 3 nm, scale bar 400 nm.

sample T: height 3 nm, scale bar: h-i. 50 nm, j. 400 nm.

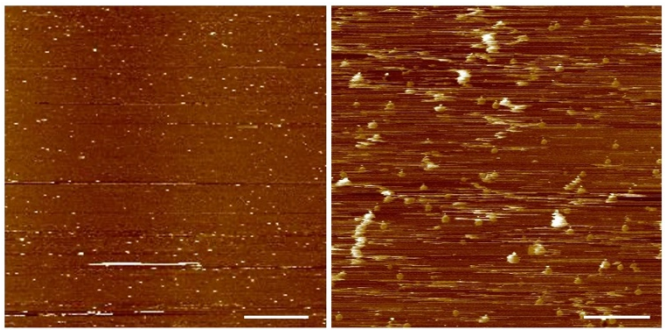


Figure 6.7: Overhang DNA imaging in water, height 4 nm, scale bar: a. Sample R- 1 µm, b. Sample S- 400 nm. The molecules mobilize when imaged in water.

6.3. M13 DNA fragments

Bacteriophage M13 contain single-stranded circular DNA of length ~6400 nt. The ssDNA can be converted to dsDNA by polymerisation in the presence of a Klenow fragment and ligase. But, when the Klenow is present in concentrations more than the ligase, a ssDNA flap is formed branching from the closed dsDNA with an unknown length[2]. This process could be used to synthesize the desired flap DNA structure, as described in section 4.2.3. The ssDNA could be polymerised by Klenow to produce closed dsDNA with a ssDNA flap and by NgPolI to produce closed dsDNA without the flap, the latter being used as a negative control for FEN activity on AFM.

6.3.1. M13 ssDNA

Initially, to obtain control images of ssDNA, M13 ssDNA was purified from the transfected *E.coli* cells and immobilized onto PLO treated mica for AFM imaging (figure 6.8). The images revealed closed DNA strands with prominent 'beads' all along its length and a height of about 1 nm while imaging in air. The 'bead' could be explained either by intermittent folding of DNA along the entire length or by supercoiling of closed DNA. In both the situations, the length of the ssDNA would be much shorter than the length of 6400 nt (nucleotide) DNA. The length of DNA molecules was traced using the NeuronJ plugin of ImageJ and averaged at ~1100 nm (figure 6.8 and 9).

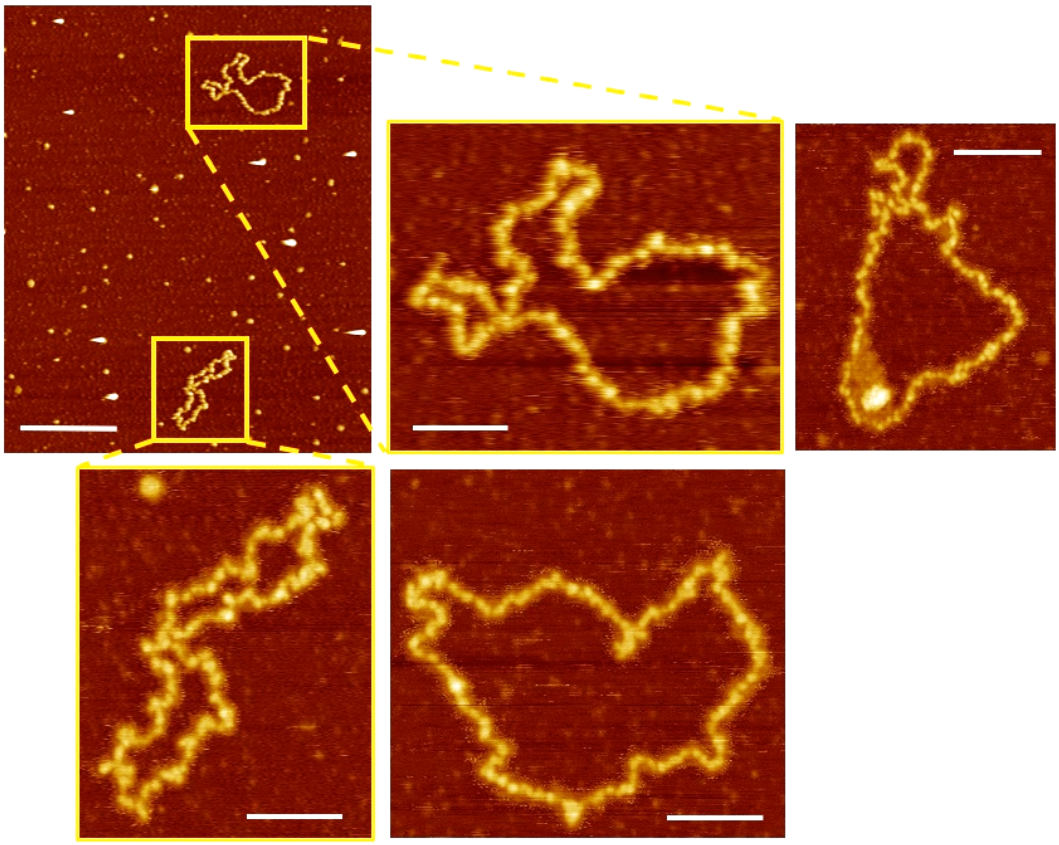


Figure 6.8: M13 ssDNA, imaging in air, height 4.5 nm, scale bar: a. 400 nm, b-e. 100 nm.

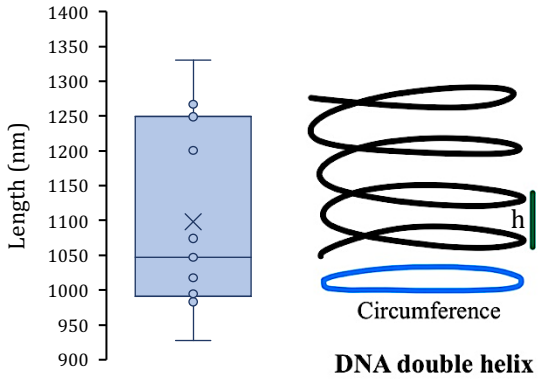


Figure 6.9: Plot of length of ssDNA traced from DNA images.

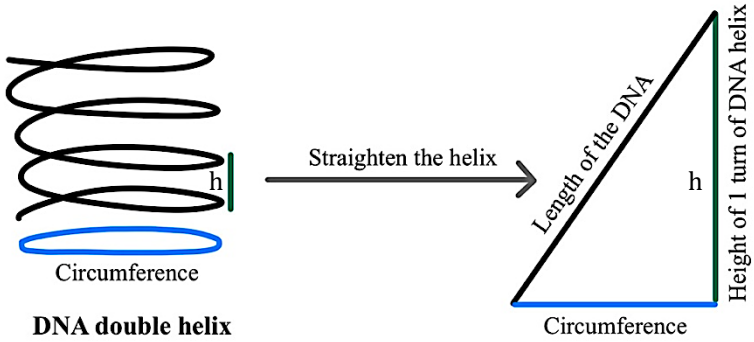


Figure 6.10: Diagrammatic representation of DNA to calculate the length of DNA when straightened and not in helix.

M13 DNA being ssDNA does not form a double helix hence would be longer than its dsDNA helix counterpart. The number of nucleotides (~6400 nt) were hence used to calculate the length of ssDNA that would be observed on the AFM. If one of the turns of a DNA helix is straightened (much like straightening a spring coil), the Pythagorean theorem could be used to

calculate the length of the DNA strand (figure 6.10) when the height of the turn (h), number of nt in a turn (n) and circumference (c) of the helical turn is known.

$$L^2=h^2+c^2 \qquad \textit{(equation 6.1)}$$

hence, length of DNA with radius about 1 nm and height of turn of 10 bp 3.4 nm can be calculated as

$$L^2=(3.4)^2+(2\pi \text{ x } 1)^2$$

$$\text{or } L=7.15 \text{ nm per turn} \qquad \textit{(equation 6.2)}$$

DNA molecules contain 10 bp per turn hence n bp would constitute n/10 turns. Therefore, the length of DNA containing n turns would be n x L or 7.15 x n nm. Thus, for M13 DNA with 6400 nt, the length of DNA if it were double-stranded and stretched straight instead of being a double helix, would be

$$\frac{6400}{10} \, X \, 7.15 = 4576 \text{ nm}$$

The lengths of the ssDNA molecules were traced from AFM images and found out to be averaging at ~1050 nm (figure 6.9), which was about 4 times shorter than the theoretical length expected, thereby indicating that the molecules had folded intermittently or had supercoiled.

The samples imaged in water revealed that the molecules were breaking at the phosphodiester bonds (figure 6.11) either due to low pH of the water used for imaging (due to dissolution of atmospheric CO_2) or a large force of cantilever oscillation. Even when each of these situations were removed, the DNA molecules disintegrated during imaging in liquid.

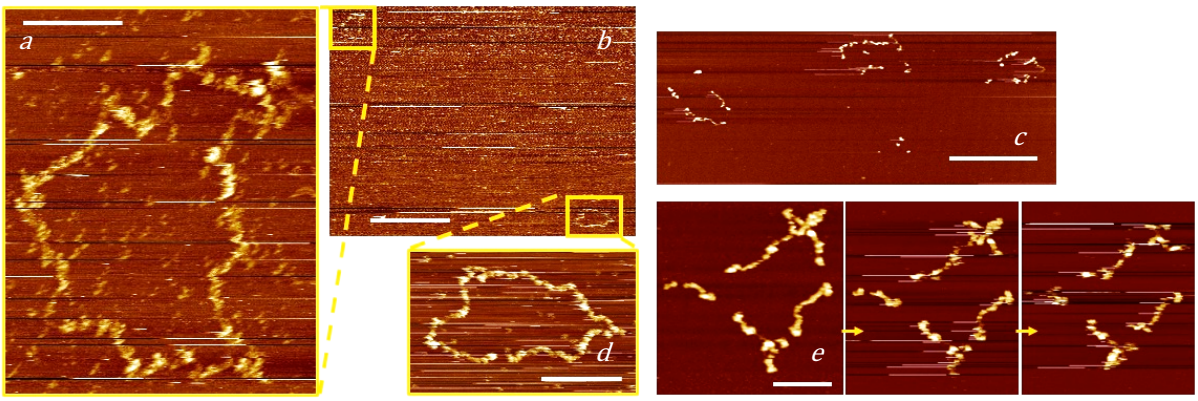


Figure 6.11: M13 ssDNA, imaging in water: a. height 8 nm, scale bar 100 nm, b. height 4 nm, scale bar 1 µm, c. height 4.5 nm, scale bar 400 nm, d. height 8 nm, scale bar 200 nm, e. height 5 nm, scale bar 100 nm. DNA appeared to have 'beady' appearance indicating that the DNA was folded intermittently or supercoiled.

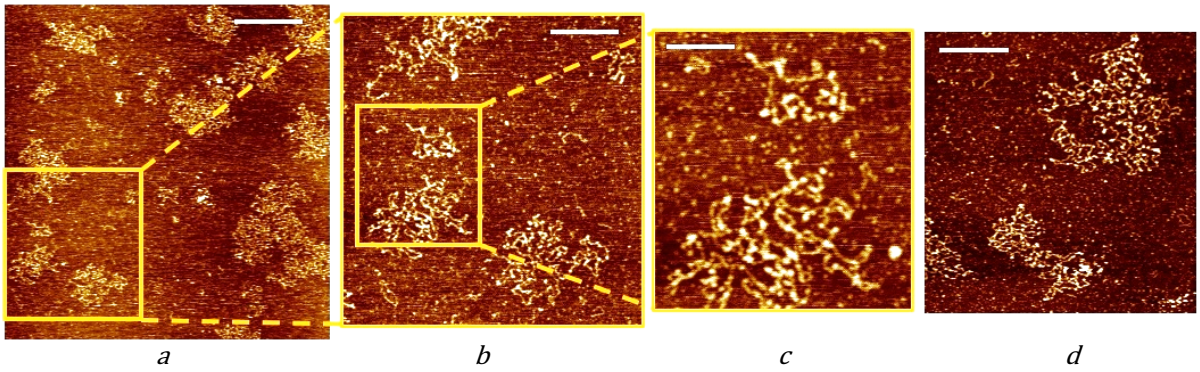


Figure 6.12: M13 ssDNA, imaging in air: a. height 1 nm, scale bar 400 nm, b. height 1 nm, scale bar 200 nm, c. height 1.2 nm, scale bar 100 nm, d. height 1 nm, scale bar 200 nm.

Also, the images by this sample preparation method did not have any reproducibility. When the experiments were repeated (figure 6.12), similar images could not be obtained, but the DNA molecules appeared to be highly convoluted, making it difficult to either resolve the double helix and conformations of DNA during imaging in liquid or to measure the length of DNA. Nevertheless, the molecules appeared to be much longer than the images previously obtained (figure 6.12 vs the previous figures 6.8 and 6.11), even though the sample preparation was the same, indicating that the molecules had supercoiled or folded intermittently previously but did not in the subsequent experiments. This lack of reproducibility could not be explained. The conclusion from these images and the following images of dsDNA and flap M13 DNA (as seen in the sections below) was that the molecules were too long to be suitable to study the conformation of DNA using AFM.

6.3.2. M13 DNA Polymerised by NgPolI

A negative control for flap DNA molecules was closed dsDNA without a flap. This was synthesized by purification of ssDNA from M13 infected *E.coli* cells (see section 4.2.3 for details) and polymerization of the ssDNA using NgPol1 enzyme (see section 4.2.3.3) in the absence of ligases. Polymerase would convert the ssDNA to dsDNA and because the enzyme contains a 5'-exonuclease domain, the ssDNA flap synthesized at the end of polymerization reaction would be digested. This dsDNA was immobilized on PLO and imaged using AFM (figure 6.13-14). The images revealed highly coiled DNA, similar to ssDNA images (figure 6.12), and posed the difficulty of being unable to resolve the conformation and helical pitch of the molecules. Hence, as mentioned above, the preferred sample would be a shorter DNA molecule that does not have DNA convolutions.

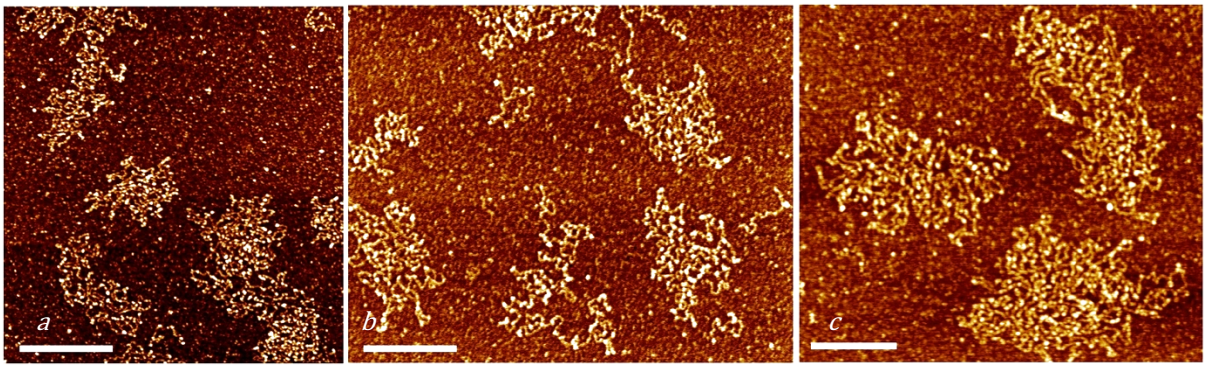


Figure 6.13: M13 dsDNA, imaging in air: a. height 1.2 nm, scale bar 300 nm, b. height 1 nm, scale bar 240 nm, c. height 2 nm, scale bar 200 nm. The DNA molecules were very convoluted and it was not possible to resolve the individual strands.

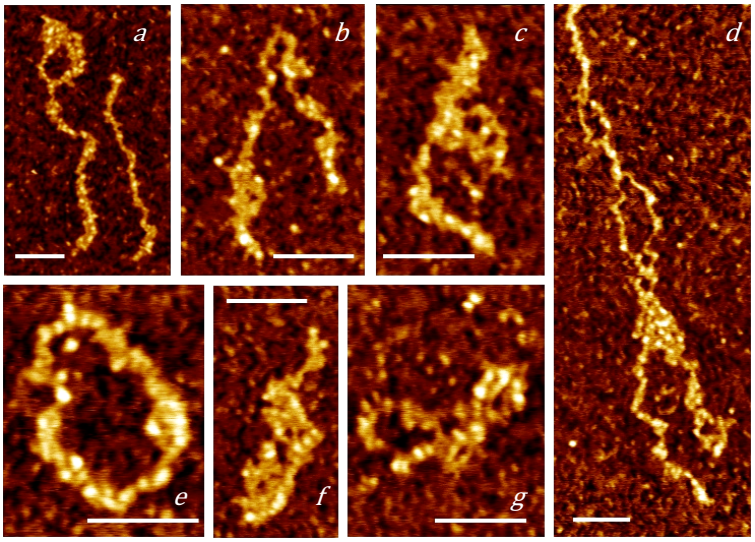


Figure 6.14: M13 dsDNA, imaging in air: a. height 3 nm, scale bar 100 nm, b-g. height 2.5 nm, scale bar 100 nm. The DNA appeared to be very convoluted and it was not possible to resolve any details of DNA.

6.3.3. M13 DNA Polymerised by Klenow: Flap DNA

ssDNA can be polymerised in the presence of Klenow fragment of DNA PolI to produce dsDNA with a flap structure of unknown length. The sample thus prepared and imaged using AFM (figure 6.15) showed the DNA molecules to be highly convoluted, as was the case described above in case of ssDNA and ds DNA molecules.

From the images of dsDNA (with or without flap) (figures 6.13-6.15), it was observed that the DNA resembled the images of ssDNA (figure 6.12) and it was hence possible that the polymerization had not occurred at all. Moreover, since the images did not show the DNA well resolved, the dilemma could not be solved. Therefore, it was worth considering that the process of polymerization of M13 DNA could not be used to synthesize flap DNA as the reaction was probably not efficient.

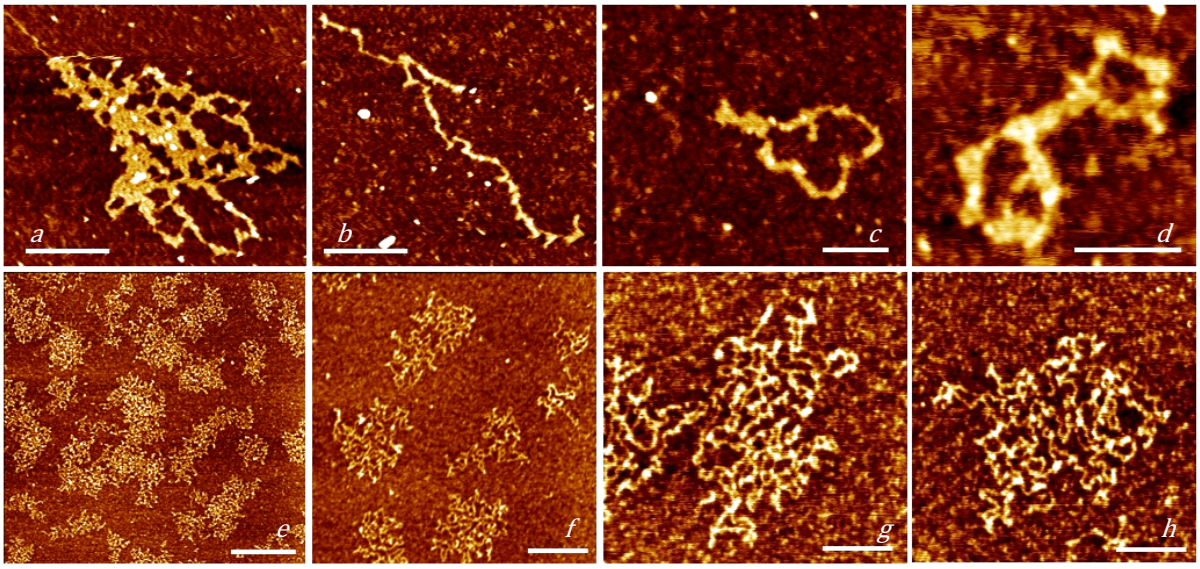


Figure 6.15: M13 flap DNA, imaging in air: a. height 2.25 nm, scale bar 200 nm, b. height 2.5 nm, scale bar 200 nm, c. height 3 nm, scale bar 100 nm, d. height 2.9 nm, scale bar 100 nm, e. height 1 nm, scale bar 500 nm, f. height 1 nm, scale bar 200 nm, g. height 1 nm, scale bar 100 nm, h. height 1 nm, scale bar 100 nm.

From the experiments with M13 DNA, it could be concluded that closed DNA molecules of length ~7 kb immobilized on PLO treated mica to form convolutions that would make it difficult to resolve the conformations of DNA. Hence it was preferable to design molecules of a length longer than the 100 bp DNA that we were working with initially (chapter 5) but not long enough for the DNA to coil or fold over one another. Hence, a length of 300 bp (based on the lengths of DNA that most of the research groups have worked on [3–5]) could be a good estimate and was experimented with in the following sections.

6.4. Assembly of long oligonucleotides into 300 bp Flap DNA

So far, flap DNA had been attempted to be synthesized by annealing short oligonucleotides and by error-prone dimerization of single-stranded M13 DNA to form circular dsDNA with a ssDNA extending as a flap of unknown length in the absence of an exonuclease in the dimerization reaction buffer. The former method had produced flaps successfully, but the lengths of DNA were too short for recognition of the ssDNA flap next to the prominent dsDNA during AFM imaging (chapter 5). The second method had produced long DNA molecules that were too convoluted to observe if the formation of flap DNA had even been successful or not. Hence, the method of DNA synthesis with oligonucleotide assembly was attempted, but with longer oligos to produce DNA at least 300 bp long and a flap of about 100 nt.

The basic principle of flap DNA synthesis method was the same as for 100 bp flap/nicked DNA synthesis described in chapter 5. The oligonucleotides were ordered from Eurofins but instead of ~20 nt length that we were working with initially, oligos varying from 50-150 nt were ordered. The assembly would theoretically lead to a flap DNA molecule (for details, see appendix 7) having a 128 nt flap branching from dsDNA with 178 bp at one end of the branch point and 146 bp at the other end (figure 6.16a). But, since the method of ligation and annealing is not 100% efficient, some of the DNA oligonucleotides did not assemble into the intended structure and some variations were formed, like the ones, but not limited to the conformations shown in figure 6.16 b-i. The figure shows some of the DNA arrangements that were possible to be formed as a result of inefficient ligation process.

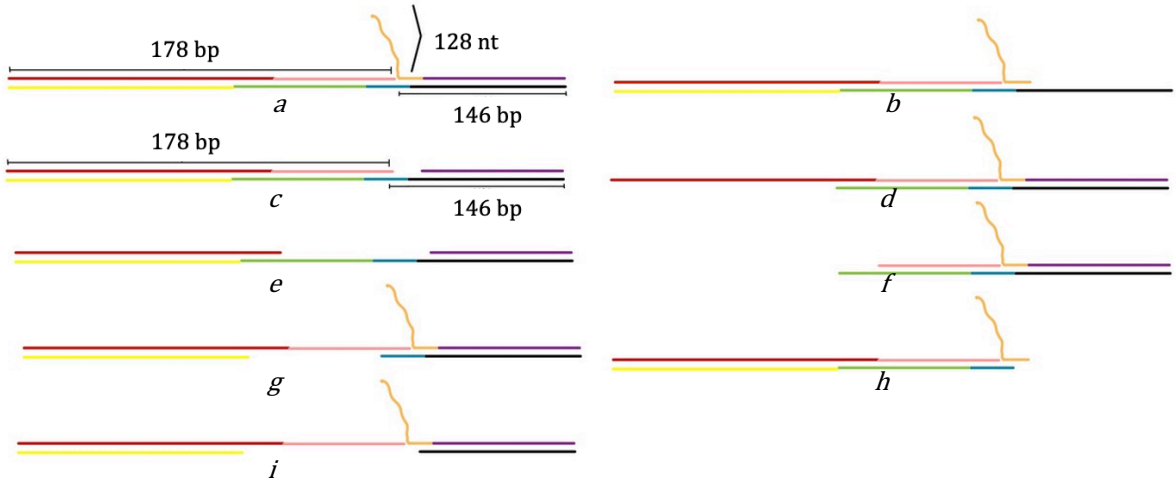


Figure 6.16: Some possible DNA structures formed by annealing when one or two of the oligonucleotides are not annealed. a is the flap DNA conformation formed as a result of perfect annealing and ligation, b, d, e and g have one oligonucleotide missing while e, f, h, and i have two oligonucleotides missing.

Some of these DNA assemblies could be used to study the interaction of FEN with DNA as the protein has demonstrated to interact with dsDNA either having a ssDNA flap or an overhang[1], with the difference that FEN recognises and binds the former better. In figure 6.16, structures a, b, d, f-i show a flap that the protein could bind to and figures b, d, f show an overhang that could interact with FEN. In the figures 6.17-18, various DNA structures formed could be seen. In image 6.17a, the molecules encircled were the DNA molecules with an overhang and not the flap DNA. Similarly, another overhang DNA could be seen in figure 6.17 b labelled x, while the rest of the molecules encircled were flap DNA.

Like the DNA conformations depicted in figure 6.16e, there were other conformations possible if more than one oligonucleotide were not ligated, hence the overhangs seen in figures 6.17-18. The efficiency of the ligation process varied when the duration of ligation reaction was increased from 30 minutes to overnight (16 hours), hence the percentage of overhang or flap DNA was more in the latter case, as could be seen in the large-scaled images of figure 6.18. On the contrary, the large-scaled images in figure 6.17 a-b showed that the presence of a lot of short DNA fragments around the dsDNA that were possibly unligated DNA fragments. It should be mentioned that the DNA molecules were not resolved and gel eluted to extract the molecules of required length, *i.e.* the flap/overhang molecules of length ~300 bp. The purification step using agarose gel electrophoresis or native PAGE was omitted because the sample obtained from gel elution displayed a lot of noise when imaged using AFM. Hence, the samples showed the presence of more contaminants (of possible short DNA fragments) than the overnight ligation sample. However, the frequency of the lengths of ssDNA overhangs or flap were very different in the two setups, as explained later.

In figure 6.17, the possible DNA structures seen in the AFM images are shown. Flap DNA molecules were seen in figures d, e, f encircled, i and j while the figures c and g show an overhang DNA. In figure h and k, the DNA structure was the one like the one sketched in figure 6.16e with one of the oligos in the centre missing along with the oligo that was supposed to form the flap. In figure 6.17f, one DNA molecule, marked by the green arrow, showed the presence of a very long ssDNA branching from a dsDNA, which was possibly formed as a result of the non-central oligos not annealing at all or another large section of the intended arrangement missing. Similarly, the flap molecules were seen (encircled in figure 6.19) in the overnight ligation DNA sample.

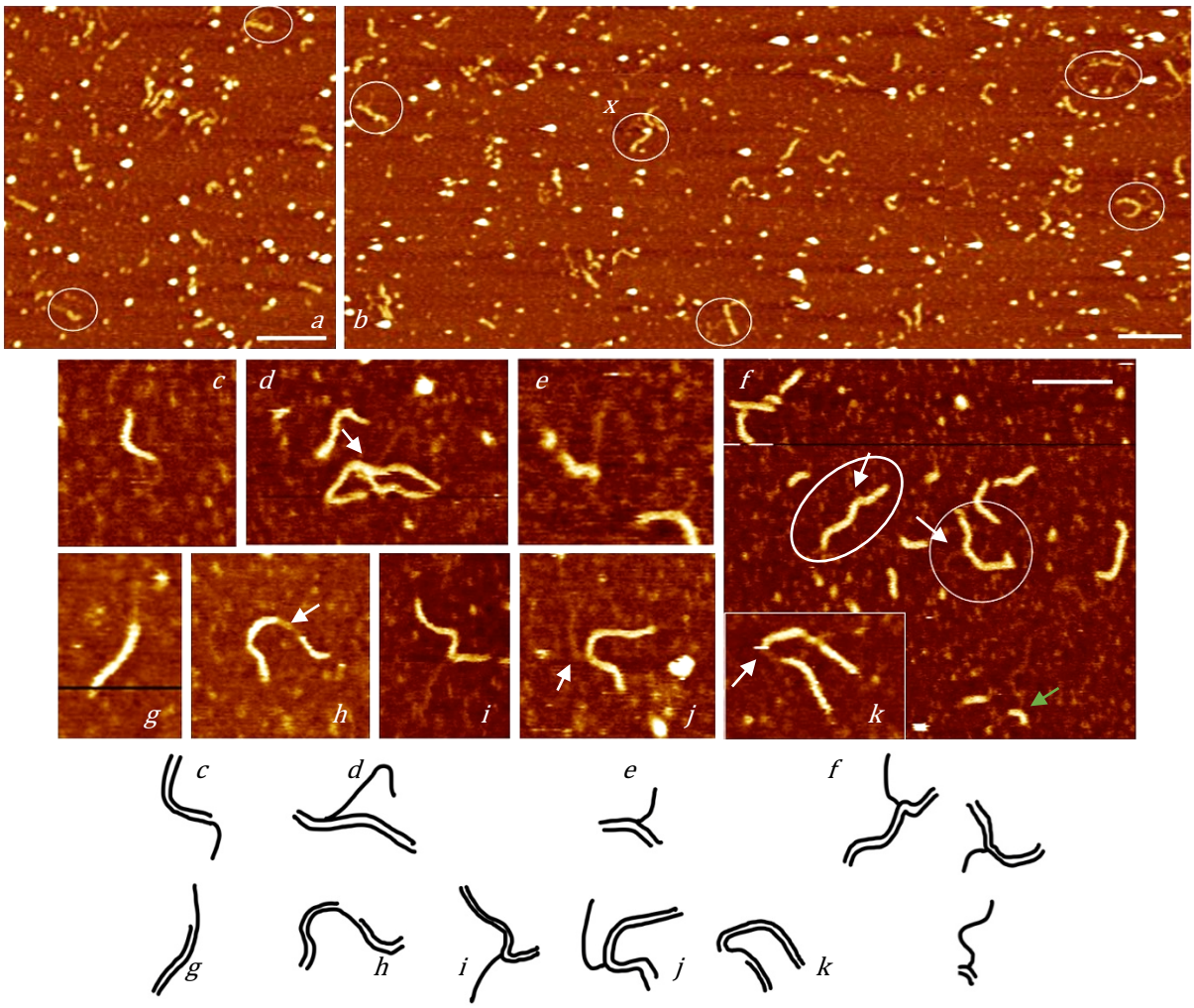


Figure 6.17: flap DNA (annealing time 30 min):
imaging in air: a-b. height 3 nm, scale bar 100 nm,
In a, the molecules encircled are overhang DNA Another overhang DNA can be seen in b, labelled x, while the rest of the molecules encircled are flap DNA.
imaging in buffer (10 mM HEPES): scale bar 50 nm: c-f,i,k. height 4.5 nm, g,h,j. height 4 nm
Various DNA shapes could be observed here. Flap DNA seen in d,e,f encircled, i and j while the figures c and g show an overhang DNA. In h and k, the DNA structure corresponding to figure 6.16e. In 6.17f, the DNA molecule, marked by green arrow, shows the presence of a very long ssDNA branching from a dsDNA. (Below) Sketches of DNA molecules in figures c-k.

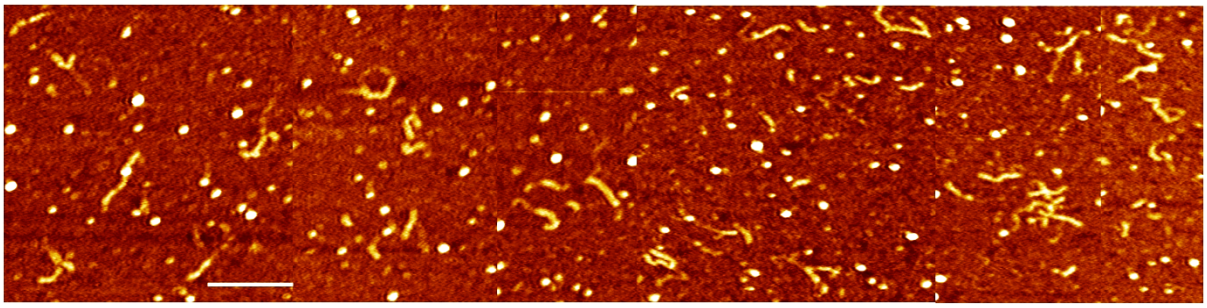


Figure 6.18: flap DNA (annealing time ~16 hr): imaging in air: height 2 nm, scale bar 100 nm. This is a collage of various images with the same scale. Various DNA shapes could be observed here. The percentage of overhang or flap DNA increases for overnight ligation than 30-minute ligation.

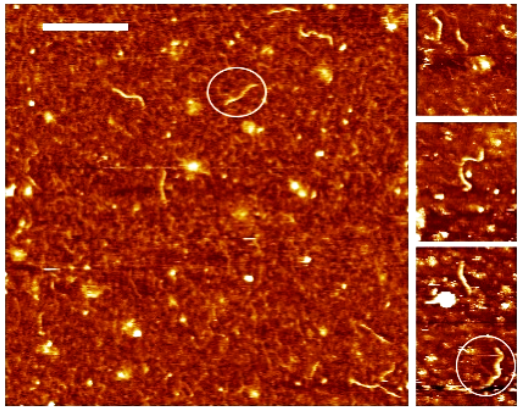


Figure 6.19: flap DNA (annealing time ~16 hr): imaging in buffer (10 mM HEPES): height 4 nm, scale bar 100 nm. Various DNA shapes could be observed here.

It is noteworthy that the flap or overhang were successfully seen in liquid imaging, as compared to the DNA assembled by annealing in chapter 5, which though prepared in the same way, differed in length (figure 6.20). In case of 100 bp DNA molecules, there were extremely rare flaps seen in liquid imaging while most of the DNA appeared as double-stranded. In 300 bp DNA images, however, there were a large number of ssDNA branching out of the dsDNA either as flap or overhang. The possible explanation was that in the short DNA, the ssDNA flap was short and not prominent when close to the thicker (hence higher on the AFM image) dsDNA from which it was branching out. Moreover, if the ssDNA was folded or looped randomly, it would have appeared as a 'blob' that could be mistaken as background noise. Shorter flaps also had lesser anchor points hence were less likely to immobilize well on the surface. Hence, it could be concluded that this method of flap DNA production proved to be the most effective out of the ones attempted so far.

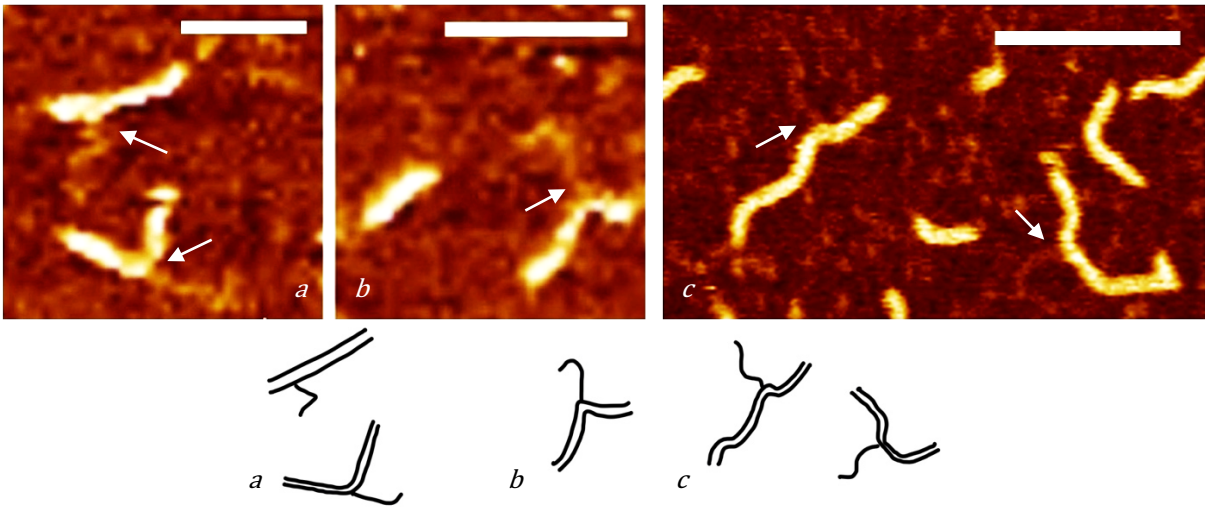


Figure 6.20: a and b: 100 bp flap DNA with 50 nt flap on PLO: images in water: height 4 nm, scale bar 25 nm and 20 nm respectively. Some rare flap DNA where the flap is visible **c: 300 bp flap DNA with 128 nt flap on PLO: images in buffer** height 4.5 nm, scale bar 50 nm. Many flaps DNA could be seen. (Below) Sketches of flap DNA molecules in figures a-b

To understand the effectiveness of the ligation reaction to produce flap/overhang DNA, the lengths of the flap or overhang DNA was traced using NeuronJ. A comparison of the length distribution corroborated the effectiveness of overnight ligation reaction in the production of flap/overhang DNA as compared to 30 min reaction (figure 6.21). The measured lengths in nm were converted into lengths that the DNA would have in terms of nucleotides, using the equation 6.2 (see section 6.3.1) to calculate the relationship between ssDNA length and the number of nucleotides. The equation can be rewritten as

$$\frac{nt}{10}\ X\ 7.15 = L\ (nm)$$

Where n is the number of nucleotides in the ssDNA and L is the length in nm of that number of nucleotides in DNA if the DNA was an elongated thread instead of a double helix. Considering the approximate error of 4 nm in the resolution by the AFM cantilever tip to locate each end of DNA and the error in tracing the lengths of the DNA strands, the length of the DNA would have an error of 8 nm. Using the above equation, that would amount to 11.12 nt or ~11 nt. Hence, the data for length measurements of the flap of DNA was binned into 11 nt size.

A comparison revealed that for DNA ligated overnight (figure 6.21 a), a large percentage of the ssDNA were 99-110 and 121-143 nt long, while the length of the flap was 128 nt. This meant that the flaps that were visible on the AFM images were possibly the intended DNA structures and not some DNA oligonucleotides randomly immobilized close to the dsDNA. It is worth mentioning that the lengths of the 8 oligonucleotides that were used to assemble into flap DNA were 79, 150, 120, 80, 33, 128, 99 and 79 nt (see appendix 7 for details). The peak observed for DNA of length 99-110 nt were possibly the fraction of oligonucleotides of lengths 79, 80, 33 and 99 nt that immobilized close to one another and hence gave the impression of length around 99-110 nt. Though 2 of the oligonucleotides (120 and 128 nt long) did fall into the length category of the required length of flap DNA, those oligonucleotides theoretically amounted to only $1/4^{th}$ of the total ssDNA oligos present and should not have amounted to many ssDNA observed if the DNA had incompletely ligated. *I.e.* If the DNA ligation had been inefficient, there should have been more ssDNA of lengths varying around the lengths of the individual oligonucleotides, and not frequencies peaking around the intended length of the flap of 128 nt.

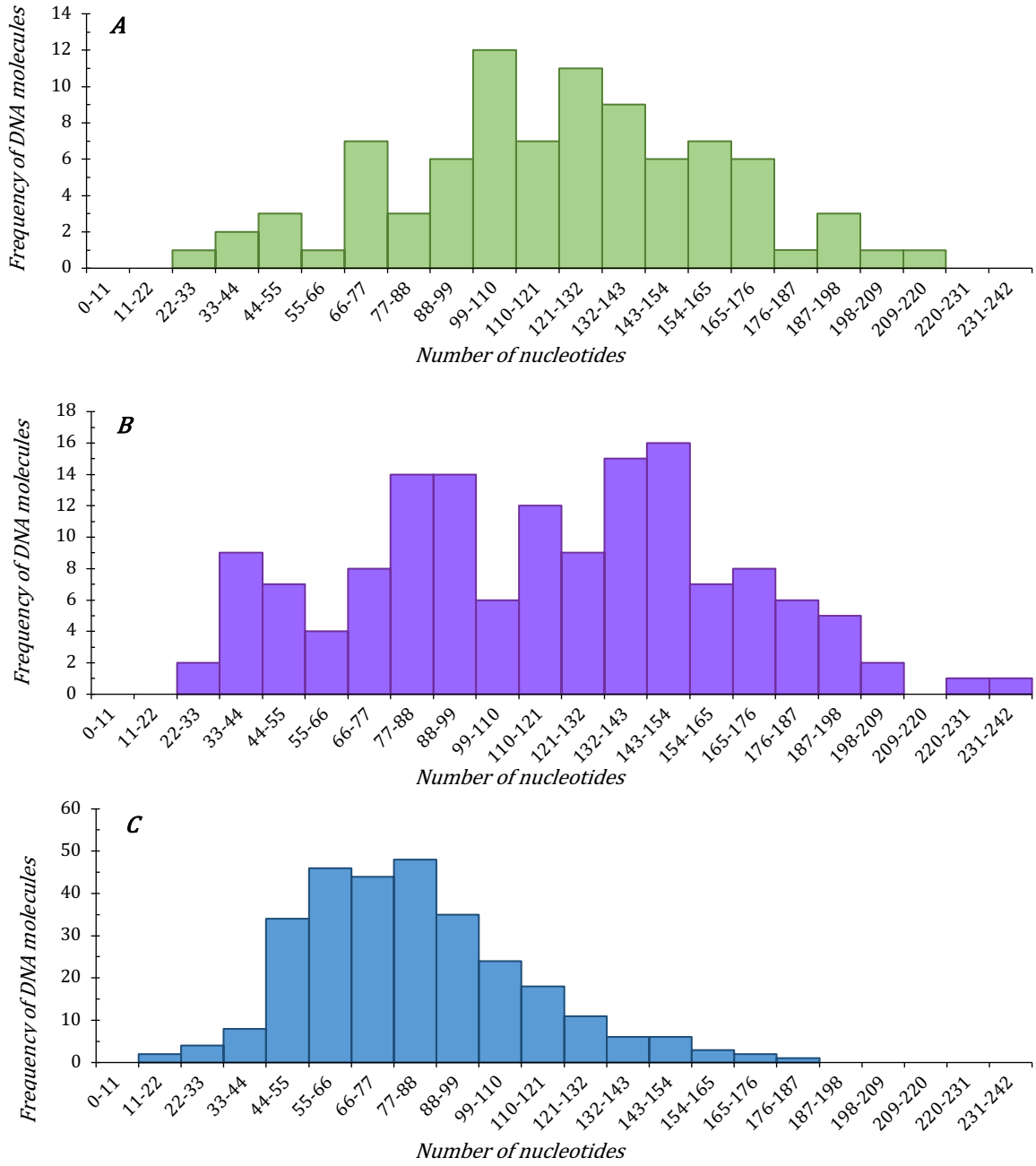


Figure 6.21: Histograms of the frequency of the lengths of flaps on DNA: x-axis: number of nucleotides, y-axis: frequency of DNA molecules with those many nucleotides in the ssDNA branch.
A. DNA ligation to form flap/overhang DNA overnight: the ssDNA lengths peak at 99-110 nt and 121-143 nt, the latter corresponding to the lengths of flap DNA.
B. DNA ligation to form flap/overhang DNA for 30 minutes: ssDNA lengths varied across the various lengths showing that the ligation process had not been very effective.
C. Ligated ssDNA: most of the ssDNA was around 50-90 nt and corresponded to the unligated oligonucleotides and indicated that no ligation occurs without annealing.

This was the case when ligation was performed for 30 minutes only (figure 6.21 b). Most of the ssDNA lengths measured fell either into the 77-99 range or 132-154 range, which was the size group of the un-ligated oligos. Remarkably, there were very few DNA of the range of 121-

132 nt, which was the target length of flap DNA required, hence indicating that the DNA that are observed here were possibly just the oligonucleotides of length 120 and 128 nt, and that the ligation reaction had been inefficient. Furthermore, there were more ssDNA in the 34-44 nt range for 30-minute ligation than overnight ligation, again indicating the presence of un-ligated DNA oligos in the 30-minute ligation sample. These plots indicated that some fraction of oligonucleotides ligated to form the intended flap DNA with the efficiency of ligation being higher for overnight ligation reaction than 30-minute reaction time.

It was possible that some of the ssDNA oligonucleotides ligated to one another to form a long ssDNA that could be randomly of a length close to the targeted 128 nt long flap, and immobilized close to the dsDNA giving the false impression that the ssDNA of suitable length close to the dsDNA was a flap DNA. To examine if this was a possibility, oligonucleotides were ligated overnight and not allowed to anneal to form dsDNA. This DNA sample was then imaged and measured to compare the frequency of the various lengths of ssDNA obtained (figure 6.21c). The frequency distribution revealed that most of the ssDNA lengths ranged between 44-99 nt, which corresponded to the lengths of oligonucleotides. Furthermore, the number of DNA of lengths corresponding to the flap DNA length (128 nt) was a very low percentage of the total DNA present. Hence, it could be safely assumed that the DNA that did fall into the group of 121-132 nt were largely the oligonucleotides of 120 and 128 nt length. This signified that ssDNA did not ligate, or if they did, the percentage was very low. This observation allowed the assumption that the DNA of lengths 120-130 nt formed while synthesizing flap/overhang DNA by overnight ligation (figure 6.21 a) were more or less the flap DNA, and not oligonucleotides that had ligated and immobilized close to the dsDNA to give the false impression of flap DNA

6.5. Conclusion

The above experiments demonstrated that assembly of long DNA oligonucleotides could be a reliable method of synthesis of flap DNA that would be suitably immobilized on PLO treated mica for high-speed AFM to study DNA-protein interactions. The other methods of synthesis of branched DNA experimented with in this chapter, namely

a. the synthesis of ~ 300 bp long overhang DNA fragments by assembly of PCR products, and
b. polymerization of single-stranded M13 DNA in the absence of exonuclease fragment of DNA polymerase (*i.e.* the klenow segment alone) to produce closed dsDNA with a flap,

both proved to be ineffective.

While synthesis of sufficient amounts of overhang DNA using asymmetric PCR and annealing of the PCR products would have been a possible substrate for FEN, the frequency of the desired structures formed was very low. At the same time, there was a large number of ssDNA present in the sample that had a comparable size to the desired DNA product and could not be perfectly separated be electrophoresis. Moreover, the DNA molecules formed loops and other unwanted arrangements of DNA and also moved when imaged in liquid, thus making it difficult to obtain good resolution images that could be used for understanding the action of FEN.

The synthesis of flap DNA by polymerization of M13 ssDNA molecules produced closed dsDNA that were too large to remain circular when immobilized onto PLO treated mica. The closed dsDNA itself was ~ 6400 bp long, while the polymerization reaction by Klenow fragment formed a single-stranded flap of unknown length on the dsDNA, which further added to the length of the molecule. Also, the images of dsDNA were similar to the images for ssDNA, and it could be assumed that the polymerization reaction had not been efficient and the DNA imaged were actually ssDNA. The large flap DNA formed convolutions and clumps making it difficult to resolve any DNA helix and thereby was not suitable for study of interactions with FEN.

The final method of oligonucleotide assembly to form ~300 bp flap DNA worked best as the length of the DNA was not too short that it mobilized (chapter 5), nor was it too long that the DNA folded and convoluted to inhibit the observation of DNA conformations and changes that

would possibly happen as a result of its interaction with FEN during dynamic imaging (chapter 7) (DNA molecules were seen extended in a worm-like conformation immobilized on the mica surface). The images also revealed the ssDNA flaps/overhang clearly and did not mobilize during consecutive scans. Hence, this DNA sample was used to study the interaction with FEN using AFM by dynamic imaging, the details of which are discussed in the following chapter.

Bibliography

1. Harrington JJ, Lieber MR. DNA structural elements required for FEN-1 binding. J Biol Chem. 1995;270(9):4503-4508.
2. Sayers JR, Eckstein F. Phosphorothioate-based site-directed mutagenesis for single-stranded vectors. In: McPherson MJ, ed. Directed Mutagenesis: A Practical Approach (Practical Approach Series) 1st Edition. first. Oxford University Press; 1991:49-68.
3. Rivetti C, Codeluppi S. Accurate length determination of DNA molecules visualized by atomic force microscopy: evidence for a partial B- to A-form transition on mica. Ultramicroscopy. 2001;87(1-2):55-66.
4. Hansma HG, Revenko I, Kim K, Laney DE. Atomic force microscopy of long and short double-stranded, single-stranded and triple-stranded nucleic acids. Nucleic Acids Res. 1996;24(4):713-720.
5. Japaridze A *et al*. Toward an Effective Control of DNA's Submolecular Conformation on a Surface. Macromolecules. 2016;49:643−652.

Chapter 7

Dynamic Imaging Flap DNA- Flap endonuclease Protein Interactions

7.1. Introduction

The aim of this chapter is to discuss the methods and results obtained for experimentation with dynamic imaging of protein-DNA interactions. But before diving into it, it was important to understand what the protein looked like. While the previous 2 chapters have focussed solely on imaging of DNA control sample and discussed the methods of synthesis, purification and immobilization of DNA for AFM, the protein sample has not been talked about. The first section of this chapter will focus on the results of imaging of protein, and their effects on the efficiency of dynamic imaging of protein-DNA interactions.

Following the methods of protein synthesis in chapter 4, various mutations and sources of FEN were produced and purified to be imaged in air and liquid imaging conditions on the AFM. It is worth emphasizing that while DNA imaging experiments focussed primarily on synthesizing an adequately sized DNA molecule and its methods of immobilization, the aim of protein imaging experiments was to explore a suitable buffer and the imaging conditions. It was necessary to ensure that the protein immobilized efficiently on PLO treated and PLO free mica, the buffer contents did not produce imaging noise, how the cations in the buffer affected the protein immobilization, what was the size of the protein to avoid being confused with noise and what was the temperature for the catalytic reaction during AFM imaging.

As already discussed in the previous chapters, DNA and protein had to be immobilized enough on the surface to be detected by the cantilever tip yet must be sufficiently flexible to allow interaction. It was worth checking if the protein immobilized better on PLO-free mica or PLO treated mica, since the charged amino acid residues on the surface of FEN or DNA Polymerase I would allow immobilization on both. Since the DNA immobilized on PLO treated mica and not mica alone, it would have been problematic if the protein immobilized better on the latter

than the former. Hence, experiments were performed to check the immobilization parameters for the protein samples.

A remarkable advantage of using AFM is that it allows imaging samples at a very low concentration/ amount. This however poses the problem that a complex buffer with many ingredients would contribute to image noise due to the sensitivity of the instrument. Therefore, it was necessary to focus on limiting the buffer composition (*eg*. presence of glycerol) of the protein solution and the reaction/imaging buffer during AFM imaging while ensuring that the buffer conditions were adequate for the DNA-protein interactions and protein stability. This also posed the limitation of time for which the protein sample could be stored or immobilized in the absence of buffer reagents necessary for the stability of the protein. Furthermore, the composition of protein solution should not interfere with the interactions of DNA or protein with the immobilization surface, and thereby the image quality.

A major detail of the protein under discussion is that it is a very small molecule as an AFM sample. DNA polymerase I measured from crystallography data is about 13 nm at the longest and 5 nm at the widest with the FEN domain alone being ~ 7 nm at the longest (figure 7.1)[1]. With the limitation in resolution of the cantilever tip, the protein would appear as a long blob in the case of DNA PolI and more or less as a spherical blob in the case of FEN. In the presence of contaminants or large contents in the buffer, the protein would be difficult to distinguish from the noise.

An important parameter in the AFM imaging of DNA-protein interaction was temperature for the stability and activity of the protein. *Thermus aquaticus* DNA Polymerase I was used to study the behaviour and shape of protein by static and dynamic imaging of protein binding to DNA without any activity. This enzyme source was chosen because it could be prepared and stored at room temperature (22ºC) hence did not require freezing. This allowed for long experimentations with sample immobilization without worrying about the stability of the protein being compromised at room temperature. But, since TaqPolI performs its activity at 37ºC, it posed the additional requirement of heating the sample during AFM imaging. Thus, the study of action of active FEN was performed using DNA Polymerase I other than from *Thermus aquaticus*, T_7 in this case.

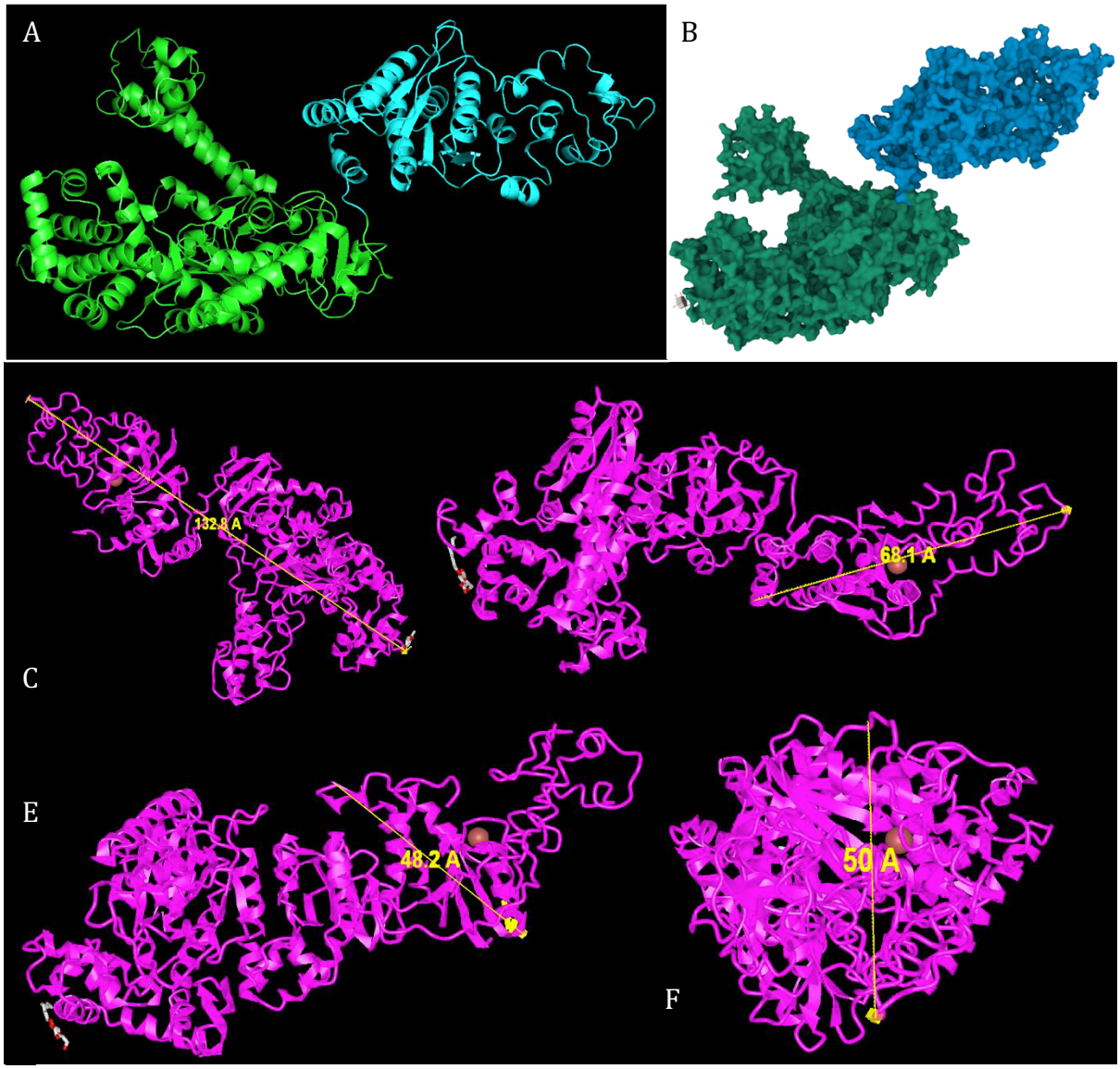


Figure 7.1: A. Cartoon model and B. Molecular surface model of Crystallography structure of *Thermus aquaticus* DNA Polymerase I reproduced from Protein Data Base: the green domain is the Klenow fragment and the blue domain is the Flap endonuclease. The Klenow fragment shows a distinct shape of 'hand, fingers and thumb'. **Shape and size of *Thermus aquaticus* DNA Polymerase I:** showing C. length of TaqPolI, D. length of FEN domain, E. breadth of FEN domain, F. width of FEN domain.

The most critical aspect of dynamic imaging DNA-protein interaction was time. The duration of endonuclease reaction should not be too short to be time resolved into multiple frames during dynamic imaging. For this reason, it was relevant to experiment with various buffer compositions that would control the rate of the reaction. At the same time, it was required to test numerous imaging conditions that would allow imaging to be as fast as possible while ensuring maximum spatial resolution. So, experiments were performed to analyse how the protein would behave during dynamic imaging without the presence of DNA on the surface

(section 7.2.2). Also, protein with a DNA binding site but without any active site was imaged to comprehend the duration of protein binding to the DNA, before the active protein could be dealt with. Having achieved a suitable time frame for the binding reaction, the active protein reacting with DNA was imaged in a suitable buffer.

In this chapter, the results for various scenarios of protein imaging are discussed: DNA polymerase I from *Thermus aquaticus* and T_7 with and without active FEN imaged in air and buffer conditions immobilized on mica or PLO treated mica. After static imaging, the protein negative control was dynamically imaged to understand how the protein immobilized on the PLO treated mica surface without the presence of DNA. Following the negative control experiments, the dynamic images for DNA interacting with active/inactive FEN domain containing DNA polymerase I were examined.

7.2. Imaging negative control of protein

7.2.1. Static imaging inactive FEN domain Thermus aquaticus DNA Polymerase I

Firstly, *Thermus aquaticus* DNA Polymerase I containing an inactive FEN (D119K or D142K mutation to disable the activity of FEN but retain the binding capacity) was imaged. The protein sample was diluted in 5 mM HEPES buffer and immobilized on mica. The sample was washed with 25 mM HEPES buffer in case of images a-e and was not washed with buffer or water for images f-i of figure 7.2. Following this, the sample was blown dry under N_2. The images revealed that the protein appeared as blobs with a width of ~20 nm (figure 7.2 i) and lengths of about 50 nm for the longest appearing molecule (*eg.* the molecule encircled in white in figure 7.2 g). There was a significant difference in the images when the sample was washed vs not washed with buffer, the proteins appeared to be larger and revealed more details when they had not been washed with buffer. In these cases, the sample was not washed with water, like it was for DNA sample, because in water, the absence of any buffer would affect the stability of the protein. There were molecules immobilized randomly on some sections of the mica surface (figure 7.2 d) while in some cases, there were clusters of protein visible (figure 7.2 a and f). Interestingly, the proteins immobilized in various conformations on the mica surface (figure 7.2 g and i). Mostly the molecules appeared as blobs, but on closer inspection, some of them revealed somewhat elongated structures folded together (encircled in blue). In some very rare cases, the protein appeared in an extended form where the domains were clearly visible:

elongated sphere-shaped FEN domain and the klenow domain with a hand, thumb and finger shape (encircled in white, where the upper end of the protein is the FEN and the lower end shows the hand shape of klenow).

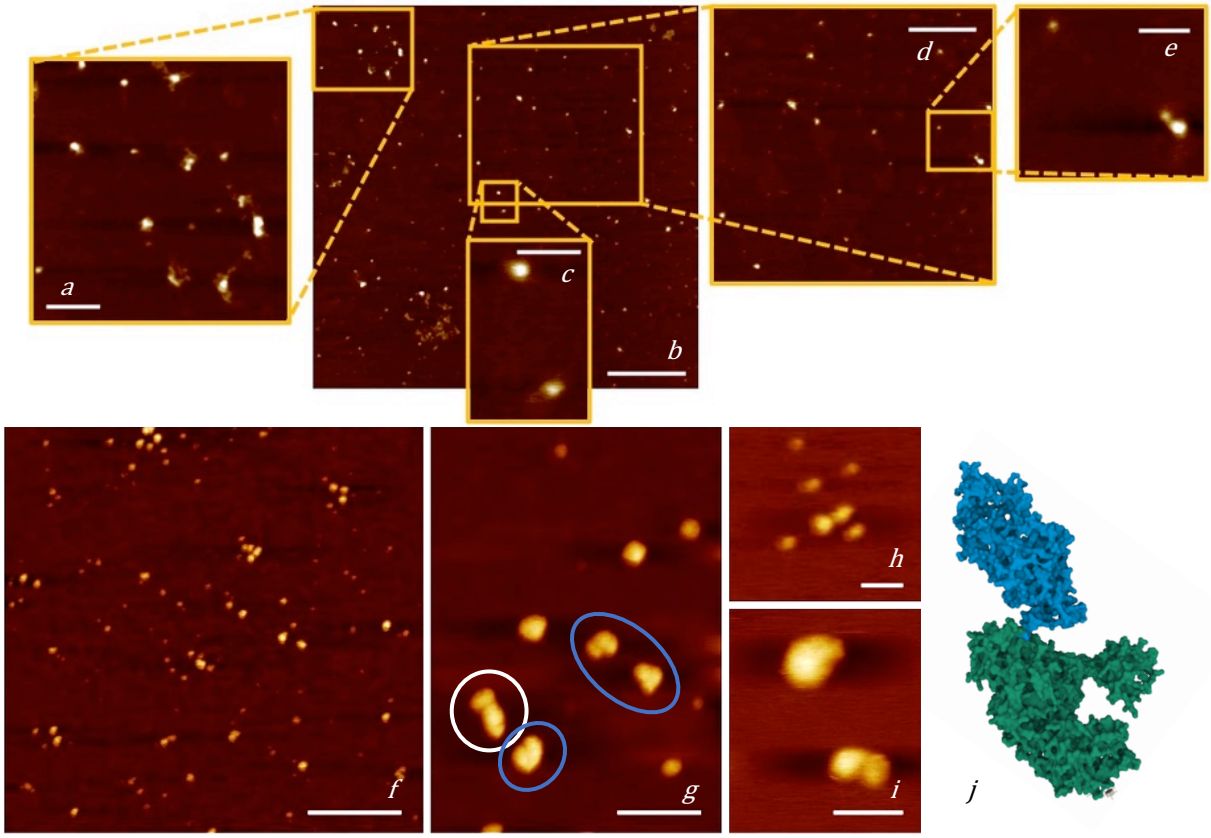


Figure 7.2: TaqPolI immobilized on mica, imaging in air. Height 3 nm. scale bar: a. 100 nm, b. 400 nm, c. 50 nm, d. 200 nm, e. 50 nm. f-g: height 9 nm, scale bar 400 and 100 nm respectively, h and i: height 7 nm, scale bar 50 nm. Most of the protein could be seen as blobs in air, but in some cases (g and i), the two domains can be resolved with the claw shape of the klenow fragment distinguishable. j. the molecular surface model of crystallography structure of TaqPolI depicting the orientation of the protein encircled in white in figure g.

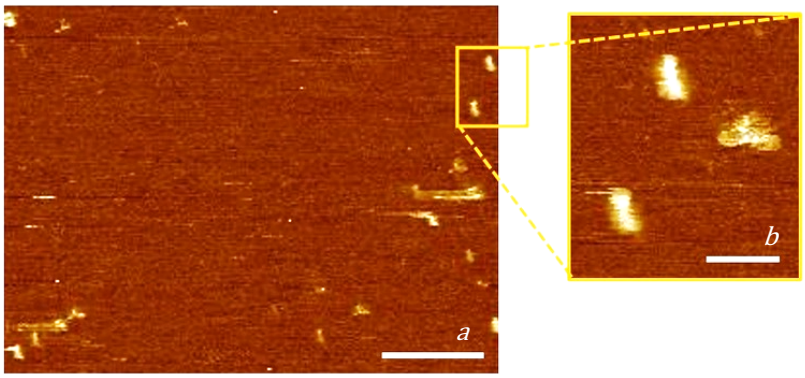


Figure 7.3: TaqPolI immobilized on mica, imaging in 25 mM HEPES.
Height 6 nm. scale: a. 200 nm, b. 50 nm. The protein appears as elongated blobs when imaged in buffer and the distinguishable domain or the claw shape of the klenow (seen in air) could not be resolved.

It was noteworthy that good conformations of the protein could be seen in air images but when the protein was imaged in 25 mM HEPES buffer, the protein concentration reduced (the sample in figure 7.3 was the same sample used for imaging in air in figure 7.2 a-e), and the resolution was compromised. Therefore, it was necessary to modify the sample immobilization process to ensure that the protein did not immobilize while imaging in buffer during the later dynamic imaging experiments. Also, it was imperative to experiment with the buffer composition for protein immobilized on PLO treated mica as that was the ideal substrate used for DNA.

TaqPolI solution was prepared in 25 mM HEPES, 2mM Ca^{2+} and 5 mM K^{+}, and immobilized on mica and PLO treated mica. The composition of the solution was based on the ideal requirement of the enzyme to react with the DNA substrate, and the buffer composition utilized for PCR using TaqPolI. The reason for experimentation with immobilization of protein solution on mica, even though the DNA sample was immobilized on PLO treated mica was to use the images on mica as a control to understand the amount and conformation of the protein that was immobilizing.

Figure 7.4 shows the images for protein in buffer solution immobilized on mica while figure 7.5 shows the same sample immobilized on PLO treated mica. The protein immobilized on mica in a similar fashion as figure 7.2 where the protein solution did not contain any cations. There were some clusters of protein seen while other protein molecules were evenly immobilized. It is worth mentioning that here the sample was washed with 25 mM HEPES buffer before drying with N_2 gas. When the same sample was imaged in buffer (figure 7.4 b and c), the composition of which was the same as the buffer in which the protein sample was prepared, some protein molecules mobilized and the number of molecules on the surface reduced. There were no details of the protein structure visible, and the protein appeared as blobs. However, there was no motion of protein during consecutive frames (figure 7.4 c) which was important to ensure during dynamic imaging of the enzyme reaction.

When the protein solution was immobilized on PLO treated mica (figure 7.5), the protein immobilized randomly and did not form clusters during imaging in air. On the other hand, while imaging in buffer solution containing Ca^{2+} and K^{+}, the immobilized protein formed circular patches on the surface, probably due to interaction of cations in the buffer with the PLO surface thus interfering with the immobilization. Here again, the protein appeared as blobs and no specific conformation of the protein could be seen.

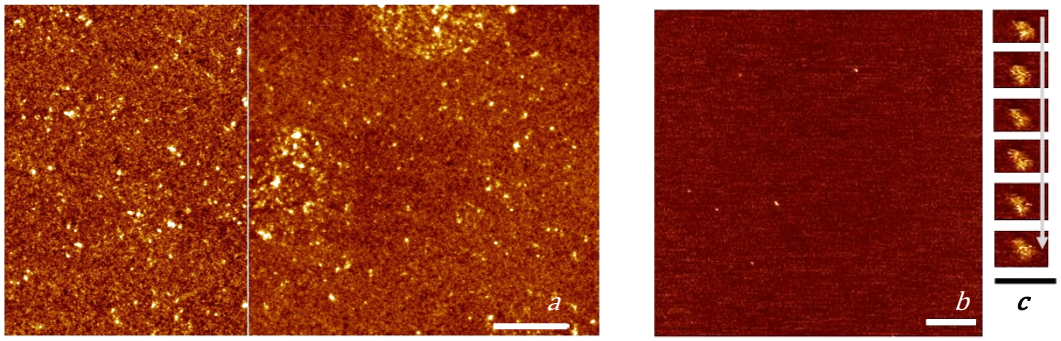


Figure 7.4: TaqPolI in buffer solution (25 mM HEPES, 2 mM Ca^{2+}, 5 mM K^{+}) immobilized on mica: a. imaging in air : Height 2 nm, scale: 400 nm**. b and c. Imaging in buffer (25 mM HEPES, 2 mM Ca^{2+}, 5 mM K^{+}):** height 5 nm, scale b. 100 nm, c. 20 nm. The protein molecules appear as blobs in air and buffer imaging conditions. The arrow in c indicates that the images are in series, i.e. consecutive frames. The series of images show that the protein molecules were well immobilized and did not move during the scans.

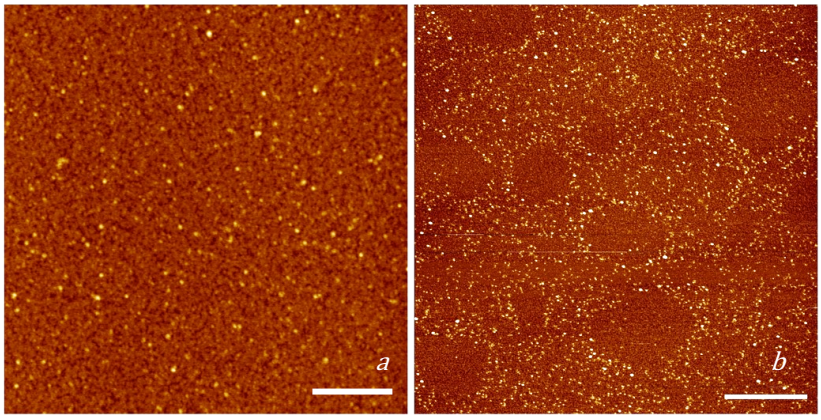


Figure 7.5: TaqPolI in buffer solution (25 mM HEPES, 2 mM Ca^{2+}, 5 mM K^{+}) immobilized on PLO treated mica: scale 400 nm. a. imaging in air, height 4 nm, b. imaging in buffer (25 mM HEPES, 2 mM Ca^{2+}, 5 mM K^{+}) height 6 nm. The protein molecules seen as blobs. In image b, the proteins could be seen encircling gaps, which were possibly formed during drying the sample.

The above-described buffer solution for sample preparation was the most suitable buffer composition and was applied to image DNA-protein interactions during dynamic imaging. There was very less mobilization of the protein due to presence of cations, no noise due to the buffer contents and the protein immobilized well on the PLO treated mica as there was no movement of molecules seen during consecutive frames.

7.2.2. Static imaging inactive FEN domain alone

Theoretically, it would be easier to image DNA interactions with FEN domain alone, without the presence of klenow fragment, that could contribute to interference with DNA binding to FEN by competition. With this aim, *Thermus aquaticus* DNA Polymerase I FEN domain was synthesized and purified (see section 4.2.5.1 and appendix 8 for details of protein production and purification) to be imaged using AFM. The protein solution in buffer was diluted with water and immobilized on mica and PLO treated mica to be imaged in air (figure 7.6).

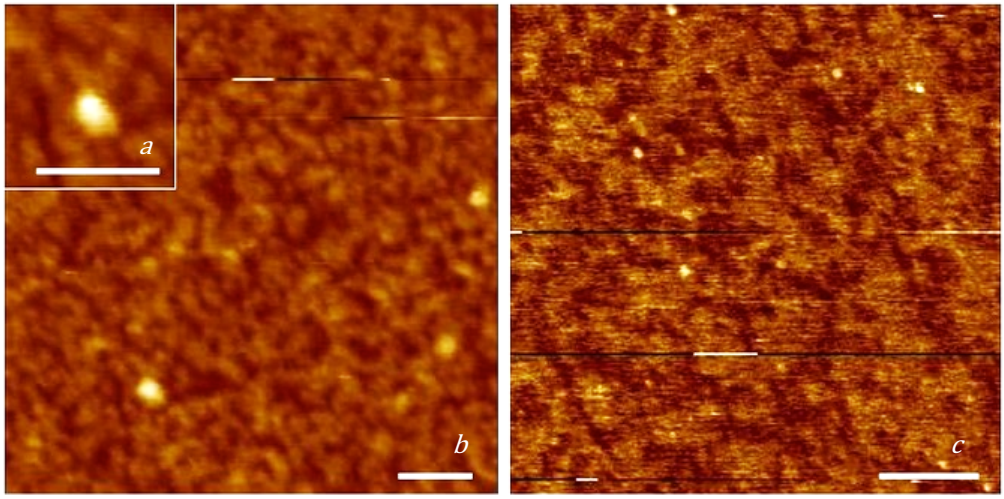


Figure 7.6: Imaging in air of flap endonuclease (*Thermus aquaticus* DNA Polymerase I FEN domain) immobilized on PLO treated mica: a-b. height 4 nm, scale bar 50 nm, **immobilized on mica**: c. height 600 pm, scale bar 200 nm. The protein molecules appear as blobs in air.

On PLO treated mica, the protein molecules appeared as blobs with size ~ 20 nm (figure 7.6 a) at the longest, which was different from the longer length of protein (~ 50 nm) observed for TaqPolI (figure 7.2 g). This was in accordance with the fact that the FEN is less than half of the entire DNA PolI. The protein size was very small and the shape too unremarkable to be distinguished from noise seen in the DNA images (see chapters 5 and 6). So, it would have hampered the observation of DNA-protein interactions. Therefore, the FEN protein sample was discarded for static images (but was used for dynamic imaging with active FEN, as described in section 7.4.2), and the larger DNA PolI was used instead, despite the klenow fragment causing possible competitive binding to the DNA substrate. It is worth mentioning, that the length obtained from the crystal structure of the protein is about 13 nm and 7 nm at the longest for DNA PolI and FEN respectively (section 7.1). Yet the sizes appeared to be larger on the AFM image, probably because the tip was also imaging the hydration layer around the protein and there could be tip convolutions that also affected the dimensions of the protein molecule.

7.2.3. Dynamic imaging TaqPolI with inactive FEN domain (protein negative control)

Once it was established that TaqPolI with inactive FEN domain (D119K or D142K) could be immobilized well on the PLO treated mica surface, dynamic imaging of the protein sample was experimented with to understand

1. how the protein would immobilize on the surface without any DNA present
2. the concentration of protein that would suitably immobilize on the surface with an adequate amount of protein immobilization to interact with DNA, and without possible overcrowding of proteins to make the DNA invisible and
3. the time duration required for the protein to immobilize.

Hence, the PLO treated mica was imaged in buffer solution of 25 mM HEPES, 2mM Ca^{2+} and 5 mM K^{+} and during imaging, the protein solution prepared in the above buffer was added (figure 7.7). From the images, it was determined that the protein usually took around 5-9 minutes to immobilize on the surface and after about 12-13 minutes, the surface became too crowded for the molecules to be visible. The protein immobilization was not reduced as compared to static imaging in liquid imaging (figure 7.5). Also, when the protein was added, it was diluted by the buffer already on the mica surface while imaging. Yet, the immobilization of protein was not affected. Having thus established that the protein negative control experiments worked well, DNA-protein experiments were ready to be performed.

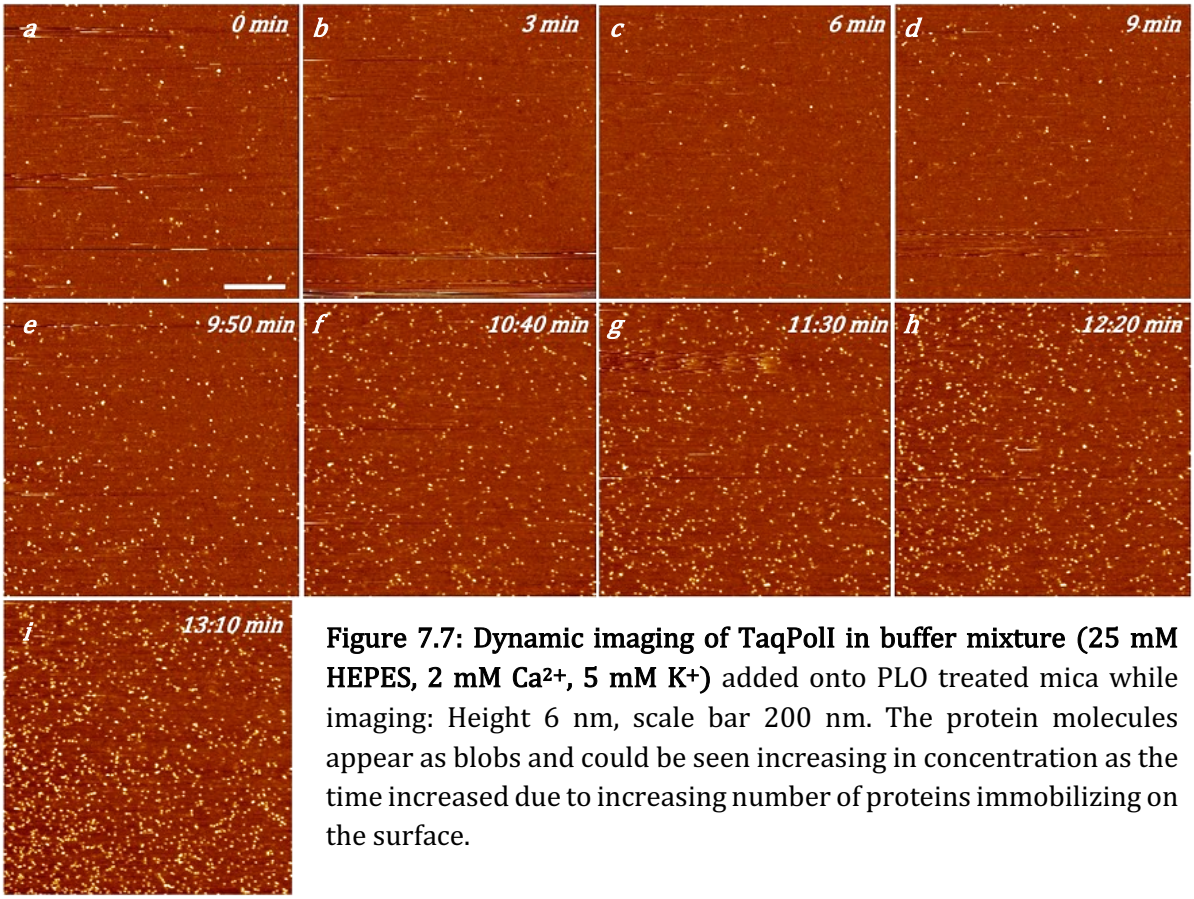


Figure 7.7: Dynamic imaging of TaqPolI in buffer mixture (25 mM HEPES, 2 mM Ca^{2+}, 5 mM K^{+}) added onto PLO treated mica while imaging: Height 6 nm, scale bar 200 nm. The protein molecules appear as blobs and could be seen increasing in concentration as the time increased due to increasing number of proteins immobilizing on the surface.

7.3. Static Imaging of Flap DNA- Taq PolI

Various combinations of DNA and DNA polymerase I were experimented with to optimize the imaging concentration and obtain control images. The sample preparation process followed was that flap DNA sample (~ 5 nM) was prepared in the buffer solution containing 25 mM HEPES, 2 mM Ca^{2+} and 5 mM K^{+}. TaqPolI with inactive FEN domain (D119K or D142K) was added to this reaction mixture and allowed to incubate for ~ 5 min. The solution was diluted 10 times and immobilized on mica and PLO treated mica.

Firstly, 100 bp DNA was allowed to interact with TaqPolI and immobilized. In the case of immobilization on mica (figure 7.8), the protein molecules could be seen on the surface, but not DNA. On the contrary, when immobilized on PLO treated mica (figure 7.9), both DNA and protein could be seen bound together, though the percentage of flap seen was extremely low. Also, as already mentioned in chapter 5, if there were any flaps at all, being single stranded, they would not have a large height in contrast to the much thicker thus higher dsDNA. This would make the ssDNA flaps difficult to distinguish. Furthermore, because the strands were short (flaps being 50 nt long), they would not have been easily detected if they folded or immobilized very close to the dsDNA section. Hence, it was safely assumed that a significant percentage of flap (or overhang DNA produced because of inefficient ligation reactions) was not visible. Given that not all the flap/overhang DNA would bind to protein in a solution, more so when the DNA was immobilized on the flat surface during dynamic imaging, out of the flap/overhang DNA that were visible, the percentage seen reacting with protein would be meagre. In accordance with this, there were very scarce flap DNA molecules seen with protein bound to them, like the one interaction in figure 7.9 c where the flap DNA was seen binding to the protein at the branch point.

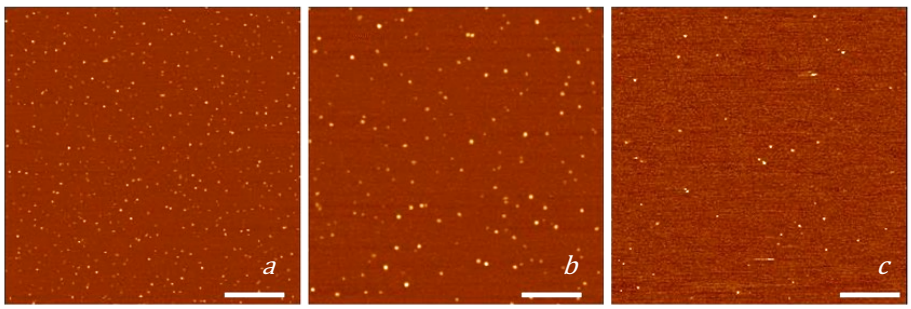


Figure 7.8: 100 bp flap DNA - TaqPolI (inactive FEN) prepared in buffer (25 mM HEPES, 2 mM Ca^{2+}, 5 mM K^{+}) immobilized on mica: imaging in air: a. height 4 nm, scale 400 nm. b. height 4 nm, scale 200 nm. **imaging in buffer (25 mM HEPES, 2 mM Ca^{2+}, 5 mM K^{+}):** c. height 8 nm, scale bar 400 nm. The protein molecules appear as blobs in air and buffer imaging conditions.

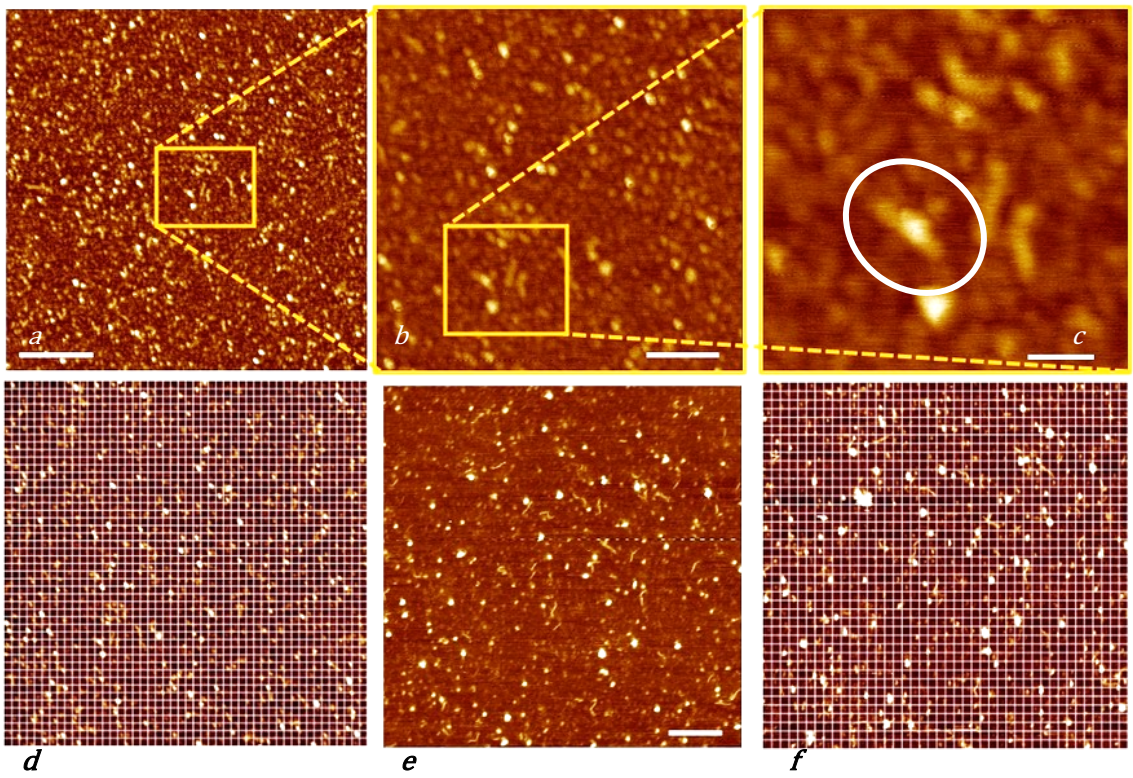


Figure 7.9: 100 bp flap DNA - TaqPolI (inactive FEN) prepared in buffer (25 mM HEPES, 2 mM Ca^{2+}, 5 mM K^{+}) immobilized on PLO treated mica: imaging in air: a. height 2 nm, scale 200 nm, b. height 4 nm, scale 100 nm, c. height 3 nm, scale 30 nm, d. gridlines drawn on an image taken in air. Imaging in buffer (25 mM HEPES, 2 mM Ca^{2+}, 5 mM K^{+}): e. height 6 nm, scale bar 100 nm, f. gridlines drawn on image e. Using the gridlines, the number of DNA and protein that occurred close together were counted to compare with the probability of occurrence of those molecules together. This allowed the estimation whether it was an interaction and not random immobilization on the PLO surface.

Similarly, when the same sample was imaged in buffer, protein could be seen immobilized very close to the DNA, but since the single strands were not seen in liquid (as already discussed in chapter 5), it could not be verified if the protein was indeed bound to the DNA or had randomly immobilized in its vicinity. To ensure if the latter was not the case, the probability of protein molecules landing on the DNA randomly was calculated.

As shown in figure 7.9 d and f, large-scaled images obtained by imaging in air and buffer were divided into grids. *Eg*. for figure 7.9 d, which had been imaged in air, the image was divided into 48x48 grid to obtain 2304 cells of cell size 20 nm each, based on the assumption that the size of the protein molecule is a slightly elongated sphere with diameter 20 nm. Since the image was 1 µm x1 µm, the size of each cell equated to 21 nm. The number of proteins counted on the image (assuming the blobs seen in the image were proteins and not contaminants) was 123. Since the size of each cell was assumed to be roughly equal to the size of a protein molecule,

the number of protein cells on the image were 123. The number of DNA counted on the image were 71. As the size of each DNA was roughly 30 nm long and 20 nm wide, it was considered that each DNA molecule was about (round-figure) 2 cells, thus giving 71 x 2=142 cells of DNA on the image.

$$p(protein) = \frac{no.of\ protein\ cells}{total\ number\ of\ cells} = \frac{123}{2304} = 0.053$$

$$p(DNA) = \frac{no.of\ DNA\ cells}{total\ number\ of\ cells} = \frac{142}{2304} = 0.0616$$

p(DNA and protein occurring together randomly) = p(protein) x *p(DNA)*

=0.053 x 0.0616 = 0.003267,

which was the probability of finding DNA and protein together in one cell. Hence, the probability of DNA and protein together in the entire image of 2304 cells was 2304 x 0.003267 = 7.526 which rounds to 8. Thus, even if the DNA and protein were not supposed to bind to each other, on average 8 DNA and protein molecules would have randomly immobilized close to one another on a 1 μm x 1 μm image frame to give the appearance of a DNA-protein interaction.

Similarly, figure 7.9 f shows 42x42 gridlines on the figure 7.9 e. The number of proteins on the image was 133, so 133 protein cells and DNA were 45, thus 90 DNA cells. Here, the sample was imaged in buffer, yet the length of the DNA was comparable to the length in air and was equated to 2 grid cells.

$$p(protein) = \frac{no.of\ protein\ cells}{total\ number\ of\ cells} = \frac{133}{1764} = 0.07$$

$$p(DNA) = \frac{no.of\ DNA\ cells}{total\ number\ of\ cells} = \frac{90}{1764} = 0.05$$

p(DNA and protein occurring together randomly) = p(protein) x *p(DNA)*

=0.07 x 0.05 = 0.0035,

which was the probability of finding DNA and protein in one cell. Thus, the probability of DNA and protein together in the entire image of 1764 cells was 1764 x 0.0035 = 6.174 which rounds to 6. The number of interactions actually seen were 8, which was comparably close to the probability calculated for protein and DNA randomly immobilizing together. Hence, in both the above cases, it was reasonable to believe that more DNA-protein interactions should have been present to ensure that the binding was not random. Therefore, experiments were

performed to synthesize longer DNA (chapter 6) such that more interactions could be easily distinguished.

Since TaqPolI requires a high temperature for activity, the DNA-protein solution prepared in a buffer was incubated at 60-70ºC before immobilization to ensure their binding. Yet, the images produced (figure 7.10) did not show any DNA-protein interactions that could be observed to understand the process better. There appeared to be interlinked DNA molecules in a way that it was difficult to distinguish the DNA-protein binding.

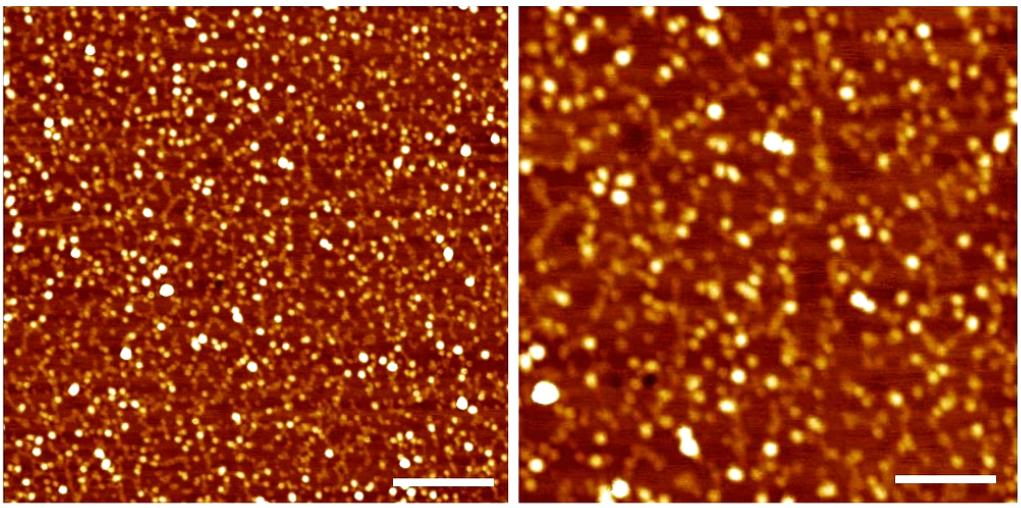


Figure 7.10: 100 bp flap DNA –TaqPolI (inactive FEN) (heat treated 60-70°C) prepared in buffer (25 mM HEPES, 2 mM Ca^{2+}, 5 mM K^{+}) immobilized on PLO treated mica: imaging in air: height 2 nm, scale 200 nm and 100 nm respectively. The heat treatment allowed the TaqPolI adequate temperature for DNA binding. The molecules appeared as a meshwork and the interactions could not be resolved.

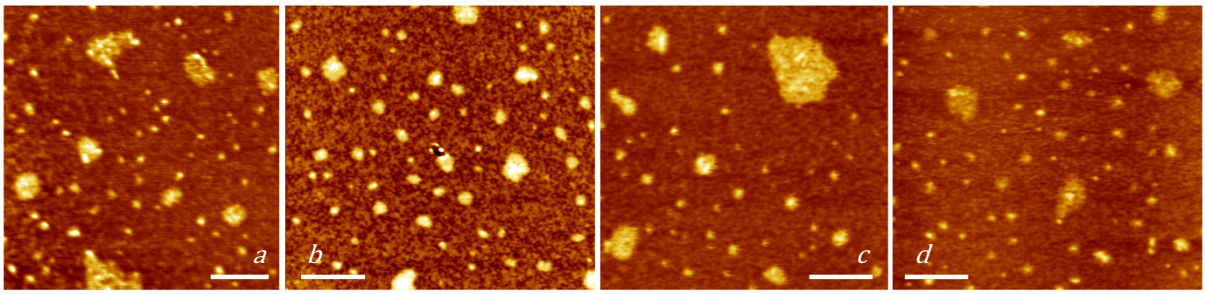
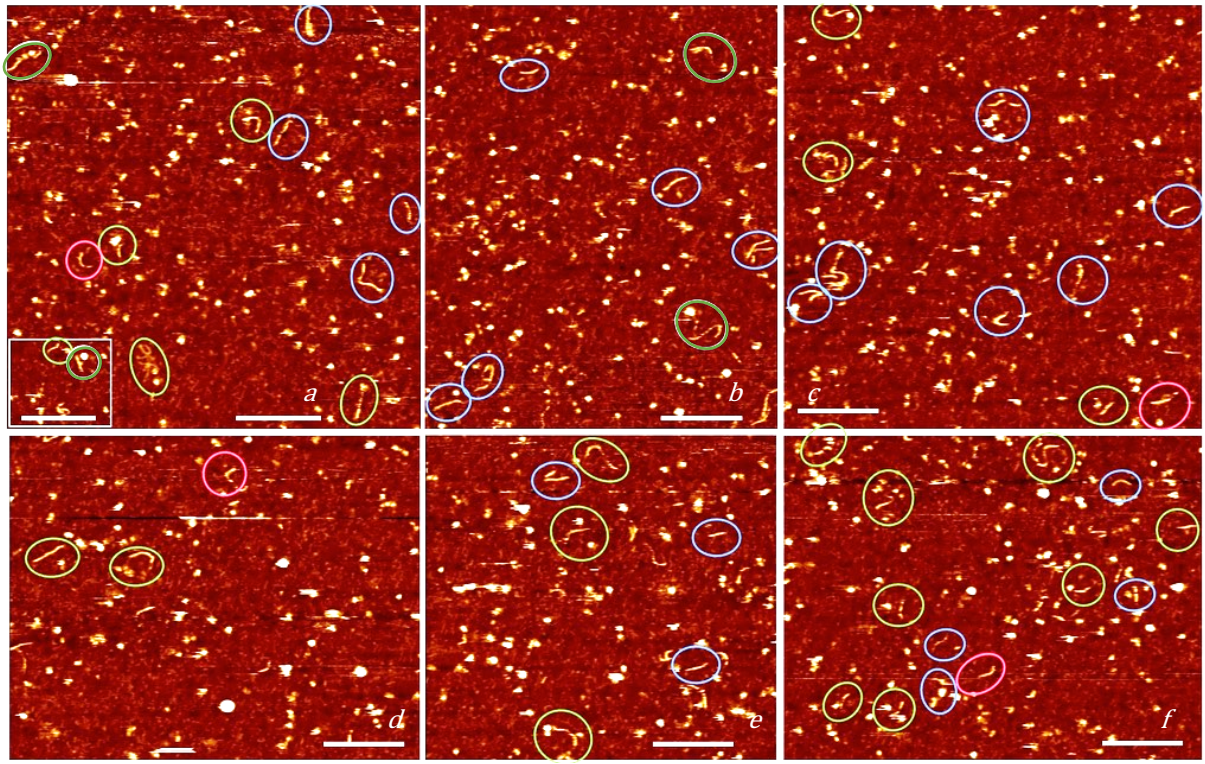


Figure 7.11: 300 bp flap DNA – TaqPolI (inactive FEN) mixed and incubated at: a-b: 65°C/5-10 min, c: 22°C/5 min and d:25-35°C/5-10 min. Diluted and immobilized on PLO treated mica, imaging in air, scale bars 400 nm, height a. 2 nm, b-c. 4 nm. The molecules are seen as aggregates and cannot be resolved.

The similar method of sample preparation was also tried with longer DNA. In figure 7.11, 300 bp DNA with flap/overhang (synthesized by annealing, chapter 6) was incubated with TaqPolI in HEPES buffer at 65ºC, 22ºC and 25-35ºC for 5-10 min, diluted in 10 mM HEPES and immobilized on PLO treated mica. The purpose of heating the sample was to ensure that there was suitable temperature for TaqPolI activity and if that could increase the efficiency of DNA-protein binding. However, the images showed that heating the reaction mixture caused clumping of molecules to the extent that it was difficult to comprehend the individual DNA or protein apart. Therefore, this technique was not utilized further for preparation of DNA-protein reaction mixture, but the initial method utilized for 100 bp DNA (for figure 7.9) was preferred since it produced better images.

When 300 bp DNA was mixed with protein in a buffer solution containing 25 mM HEPES, 2 mM Ca^{2+} and 5 mM K^{+}, allowed to react for 5-10 min, diluted in the same buffer and incubated on PLO treated mica, the images revealed numerous DNA molecules with protein bound to them or immobilized in the vicinity (figure 7.12). There were various types of DNA molecules seen, depending on the efficiency of the ligation reaction when the DNA was prepared (for details, refer section 6.4 and figure 6.16 and 18), showing both the overhang and flap DNA molecules. Interestingly, since longer DNA fragments ensured that more single stranded flaps/overhangs could be easily distinguishable, more interactions were visible as compared to the images obtained for 100 bp DNA (figure 7.9). To support the above statement, DNA molecules were counted on the images (figure 7.12) and it was observed that many DNA molecules revealed a flap or overhang with some percentage of those interacting with the protein or having protein immobilized close to them, giving the impression of a possible interaction (encircled in green). Many flap/overhang DNA (encircled in pink) did not have any protein close by or bound to them. There was a significant amount of DNA that did not have a flap/overhang attached to them (encircled in blue) thereby indicating that the ligation reaction had not been 100% efficient or that there were some ssDNA branches that were somehow not visible on the surface. Nevertheless, from the count it was revealed that about 50 % of the DNA molecules were possibly interacting with the protein because of the latter's presence close by, while about 10 % of the DNA were flap/overhang without any protein bound. The rest of the 40 % did not have a flap/overhang branching at all. From these numbers, it could be safely concluded that this buffer mixture, length of DNA sample and the protein concentration ensured that there were adequate interactions, no movement of DNA in the subsequent frames during imaging, no unwanted noise, or limitation to the visibility of ssDNA branching as

flap/overhang from dsDNA. Hence, the above buffer, DNA and protein sample and immobilization conditions were used for the next step in the experimentation, *i.e.* dynamic imaging of the interactions.



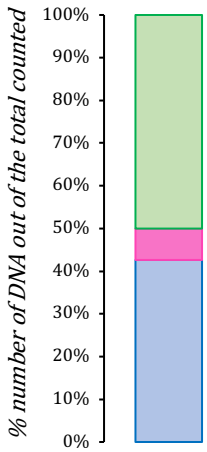


Figure 7.12: (above) 300 bp flap DNA–TaqPolI (inactive FEN) prepared in buffer (25 mM HEPES, 2 mM Ca^{2+}, 5 mM K^{+}) immobilized on PLO treated mica: height 10 nm, scale bars 100 nm.

Encircled in green: possible flap-protein interaction, pink: flap or overhang DNA without protein interacting, blue: DNA without visible overhang/flap DNA.

(right) % conformations of DNA and its binding to protein based on the count of DNA-protein images, like the ones above: green: possible interaction between flap/overhang DNA and protein, pink: flap or overhang DNA without protein interacting, blue: DNA without visible overhang/flap DNA.

7.4. Dynamic Imaging DNA-Protein Interactions

In the experiments so far, the methods of DNA and protein immobilization have been established. The DNA sample had been imaged in buffer conditions to obtain high resolution images (section 6.4) and the dynamic images for the protein sample as negative control were successfully obtained (section 7.2.3). In these sections, it was established that the DNA could be imaged with suitable resolution without any mobilization in the consecutive images, and that more and more protein molecules could be seen immobilizing onto the PLO treated mica during dynamic imaging.

As already described in section 2.2.4, DNA Polymerase I requires the presence of Mg^{2+} as a cofactor in the buffer at a suitable pH 8. Mg^{2+} allow efficient binding and activity of FEN domain of DNA PolI to DNA for the replication process to occur [2–6], while Ca^{2+} augment the binding of FEN alone to the DNA, without allowing any endonuclease activity [6–8]. For this reason, DNA PolI containing active FEN domain was imaged in a buffer containing Mg^{2+} while the protein with inactive FEN was imaged in a buffer containing Ca^{2+}. K^{+}, on the other hand, was added to the buffer to enhance the activity of the enzyme [9]. Additionally, DTT was added to the buffer to ensure the stability of the protein by protecting the disulphide bonds of the cysteine residues in the protein. Also, it enhanced the activity of TaqPol [10].

For dynamic imaging, the strategy was to immobilize DNA on PLO treated mica in a buffer solution containing Ca^{2+}/ Mg^{2+} and K^{+} and DTT, but since the presence of Mg^{2+} oversaturated the positive charge on the surface thus interfering with the DNA-PLO interaction required for DNA immobilization, the DNA was immobilized in a buffer containing Ca^{2+}. Also, DTT was omitted from the DNA solution as it contributed to noise. This process assumed that the buffer in which DNA was immobilized (followed by washing with H_2O) on PLO did not affect its interaction with the protein or modify the composition of the buffer used as the imaging buffer bubble. Therefore, the DNA solution was prepared thusly, immobilized on PLO treated mica and imaged in Ca^{2+}/ Mg^{2+} containing buffer (former for inactive FEN and latter for active FEN) in the presence of K^{+} and DTT (for active FEN). DNA was imaged continuously to observe any changes in conformation without the presence of the protein. The frame rate was usually limited to about 30 s per frame.

This section discusses the various conformation changes that were observed when the protein was interacting with the DNA molecules' single-stranded flap or overhang.

7.4.1. TaqPolI with inactive FEN interacting with flap/overhang DNA

The purpose of experimentation with TaqPolI with inactive FEN (D119K or D142K) was to ensure that the DNA was interacting *i.e.* binding to the flap or overhang DNA. It was important to guarantee that the ssDNA branches were not cleaved by the FEN too quickly for the AFM to observe. Hence, the estimation of the speed of imaging was done by using inactive FEN domain as a negative control experiment.

It is important to mention that many possible interactions that were observed due to protein molecules immobilizing close to the double strand or the single strand segment of DNA could have been random immobilization of the protein. The various conformation changes observed during the protein binding have been discussed below.

In figure 7.13, the protein was added 3 frames before the first image shown. Each frame was 36 s long (the previous 3 frames not shown). The protein was not visible until frame 2 shown (protein marked by the arrow), which is about 180 s after the protein was added. Here, the protein appeared and disappeared in frame 3, reappearing at a different location in frame 4, thereby indicating that the molecule was randomly moving about the buffer drop above the sample and would probably anchor somewhere on the PLO. The DNA molecule, contrary to the protein, did not move in the frames so far, thus signifying that DNA was well immobilized and would not randomly change conformation. The DNA showed a double-stranded fragment with a single-stranded flap which was fainter (of a lower height) than the former. In frame 12, a protein molecule could be seen close to the DNA's flap, possibly interacting, following which it moved in frame 13 and bound close to the probable branch point of the DNA flap. Following this binding, the DNA molecule apparently flipped (frame 15) and didn't move from then on. It was possible that owing to the interaction with the protein, the segment of DNA that was tightly bound to the PLO surface (and showed at a lower height) moved and thereby was observed as a higher molecule. This set of images thus depicted protein binding causing the flap DNA to move.

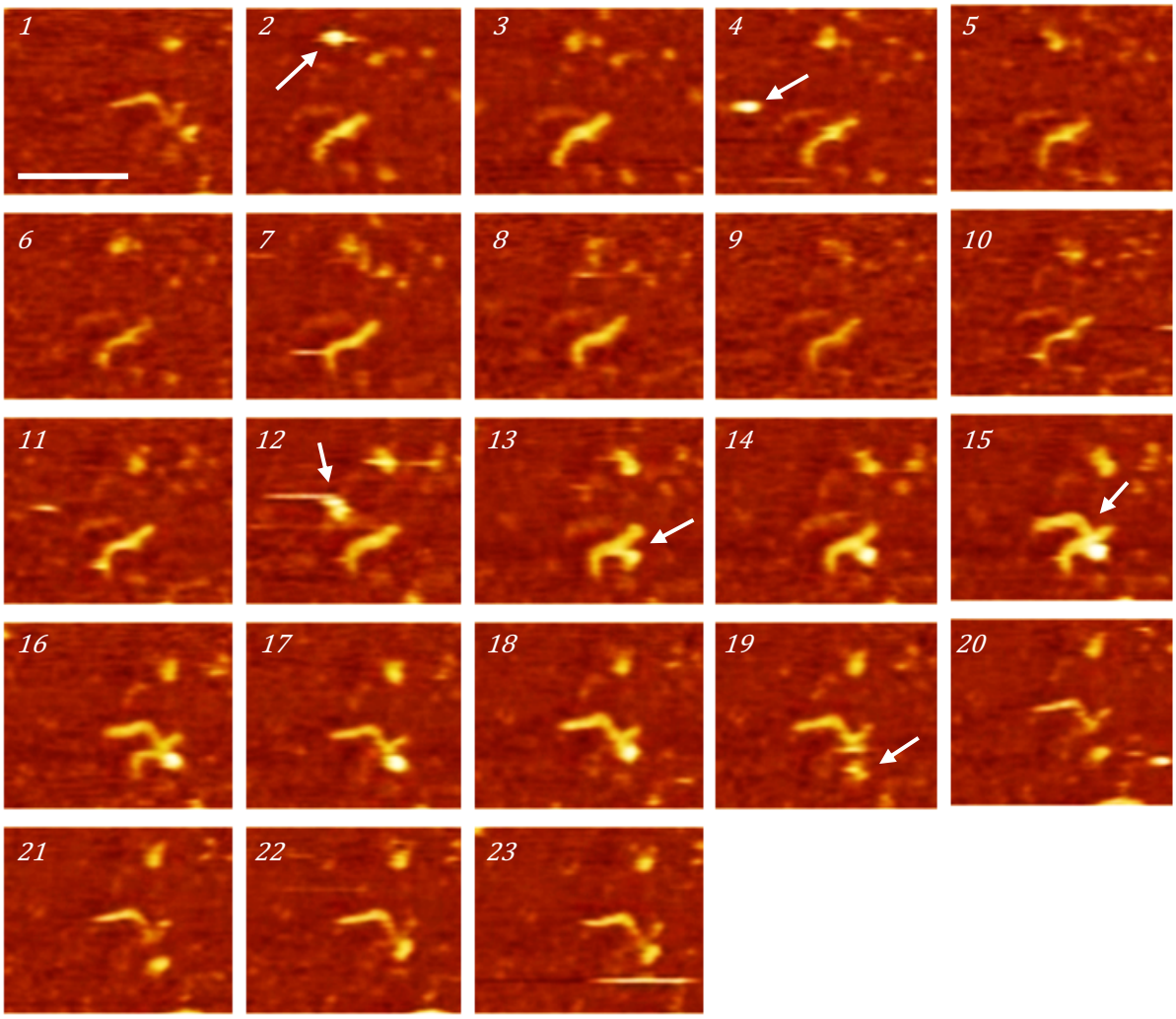


Figure 7.13: 300 bp flap DNA–TaqPolI (inactive FEN domain) imaged in buffer (25 mM HEPES, 2 mM Ca^{2+}, 5 mM K^{+}) immobilized on PLO treated mica, dynamic imaging: height 8 nm, scale bar 50 nm. Flap DNA seen here possibly changed the conformation due to binding of protein in frame 12. Refer Movie A in Supplementary Material.

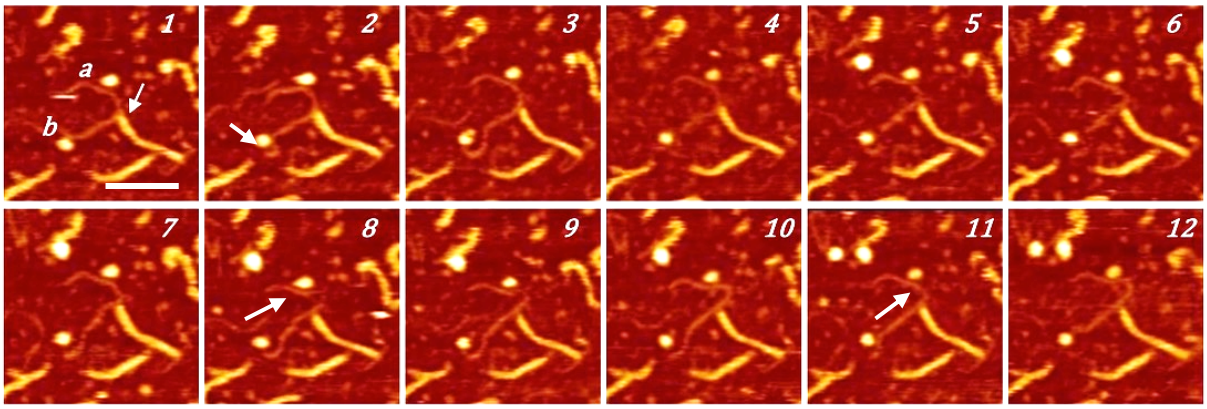


Figure 7.14: 300 bp flap DNA–TaqPolI (inactive FEN domain) imaged in buffer (25 mM HEPES, 2 mM Ca^{2+}, 5 mM K^{+}) immobilized on PLO treated mica: height 7 nm, scale bar 100 nm. Protein binding to ssDNA end 'b' caused the DNA to change the shape in frames 1-3. The end 'a' of ssDNA showed movement over the frames 11-12 due to possible binding of the protein. Refer Movie B in Supplementary Material.

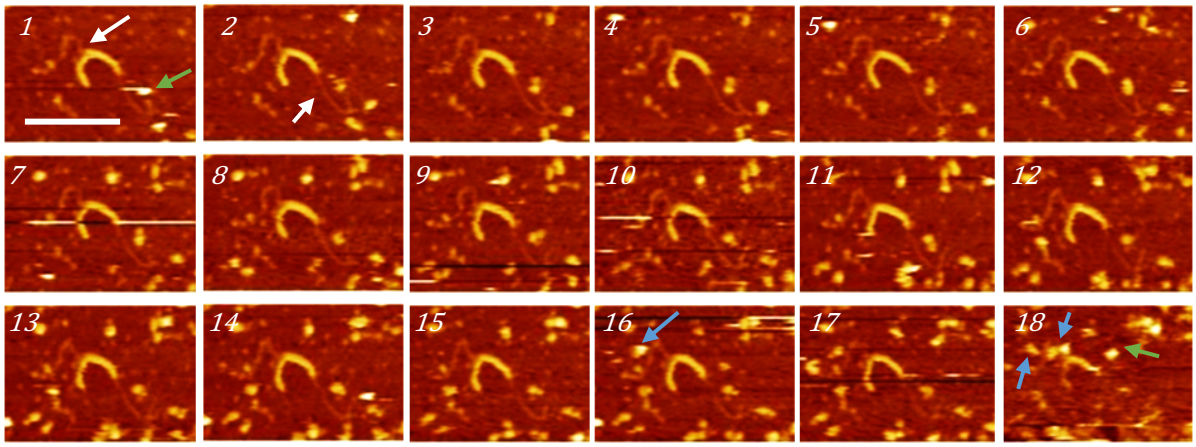


Figure 7.15: 300 bp flap DNA–TaqPolI (inactive FEN domain) imaged in buffer (25 mM HEPES, 2 mM Ca^{2+}, 5 mM K^{+}) immobilized on PLO treated mica, dynamic imaging: height 8 nm, scale bar 50 nm. The flap DNA seen here shows ssDNA moving in frame 18 after the binding of protein in frame 16. Refer Movie C in Supplementary Material.

Another set of images, figure 7.14, showed DNA molecule where a ssDNA attached to the dsDNA at the junction marked by the arrow in frame 1. The two ends of the ssDNA, labelled 'a' and 'b', remained fixed before the protein immobilized close to them. DNA end 'a' remained steady till frame 8, 7 minutes after the protein first came close to it in frame 1 (each frame is ~ 1 min), then bent close to the protein, marked by the arrow. The ssDNA then straightened slightly in frame 11 close to the protein, but came back to the bent position in the following frame 12. There was movement and change in shape seen at the b-end in frames 1-3, which could have been the result of protein binding to it or on the PLO surface slightly displacing the DNA end during immobilization.

Figure 7.15 shows another such set of images where the protein might have randomly immobilized on the PLO deceptively close to the DNA without actual binding to the flap or overhang. Here, the dsDNA had a ssDNA branch coming from the center (white arrow in frame 1) and appeared to also have a ssDNA overhang from its one end (white arrow in frame 2). The duration of each frame was 30 s and the first protein appeared about 6 minutes after the protein was added into the imaging buffer. In frame 1, a protein (marked by green arrow) was seen binding to the DNA or randomly immobilizing at the end of the overhang. It moved slightly in frame 2, but then immobilized. Many other protein molecules were also observed moving around the DNA molecule, but the DNA did not change its conformation in any of the frames, neither the flap nor the overhang. Nevertheless, there were some protein molecules aligning at the flap in frame 16 and then more in frame 18 (marked by blue arrows). Interestingly, the protein bound to the overhang (green arrow in frame 1) remained at the same place until the

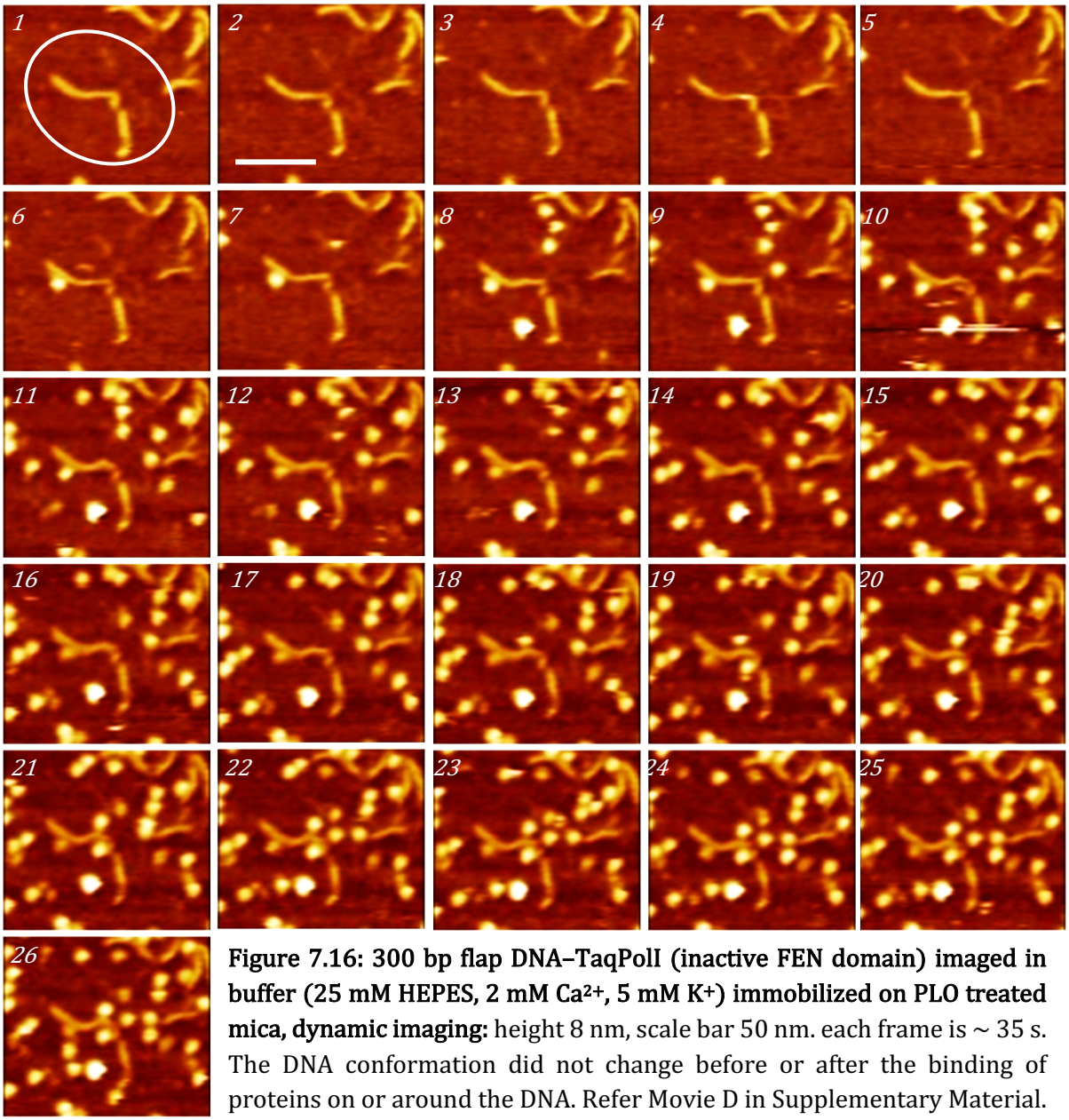


Figure 7.16: 300 bp flap DNA–TaqPolI (inactive FEN domain) imaged in buffer (25 mM HEPES, 2 mM Ca^{2+}, 5 mM K^{+}) immobilized on PLO treated mica, dynamic imaging: height 8 nm, scale bar 50 nm. each frame is ~ 35 s. The DNA conformation did not change before or after the binding of proteins on or around the DNA. Refer Movie D in Supplementary Material.

end in frame 18 where it moved (green arrow in frame 18). Further images could not be obtained as the cantilever became sticky.

Another set of images, figure 7.16, showed a flap DNA which was well immobilized with no change in the DNA conformation, before the protein was added or after the protein molecules had immobilized close to it. Here, the protein appeared frame 6 onwards, about 1 minute after the protein was added. There was no change in the movement of ds or ssDNA after the visibility of proteins and indicated the absence of interaction. Similarly, in figure 7.17, a flap DNA could be seen with the flap marked by the arrow in frame 3. The protein solution was added in a low concentration in 1st frame, each frame spanning ~ 1min. The proteins first appeared frame 16

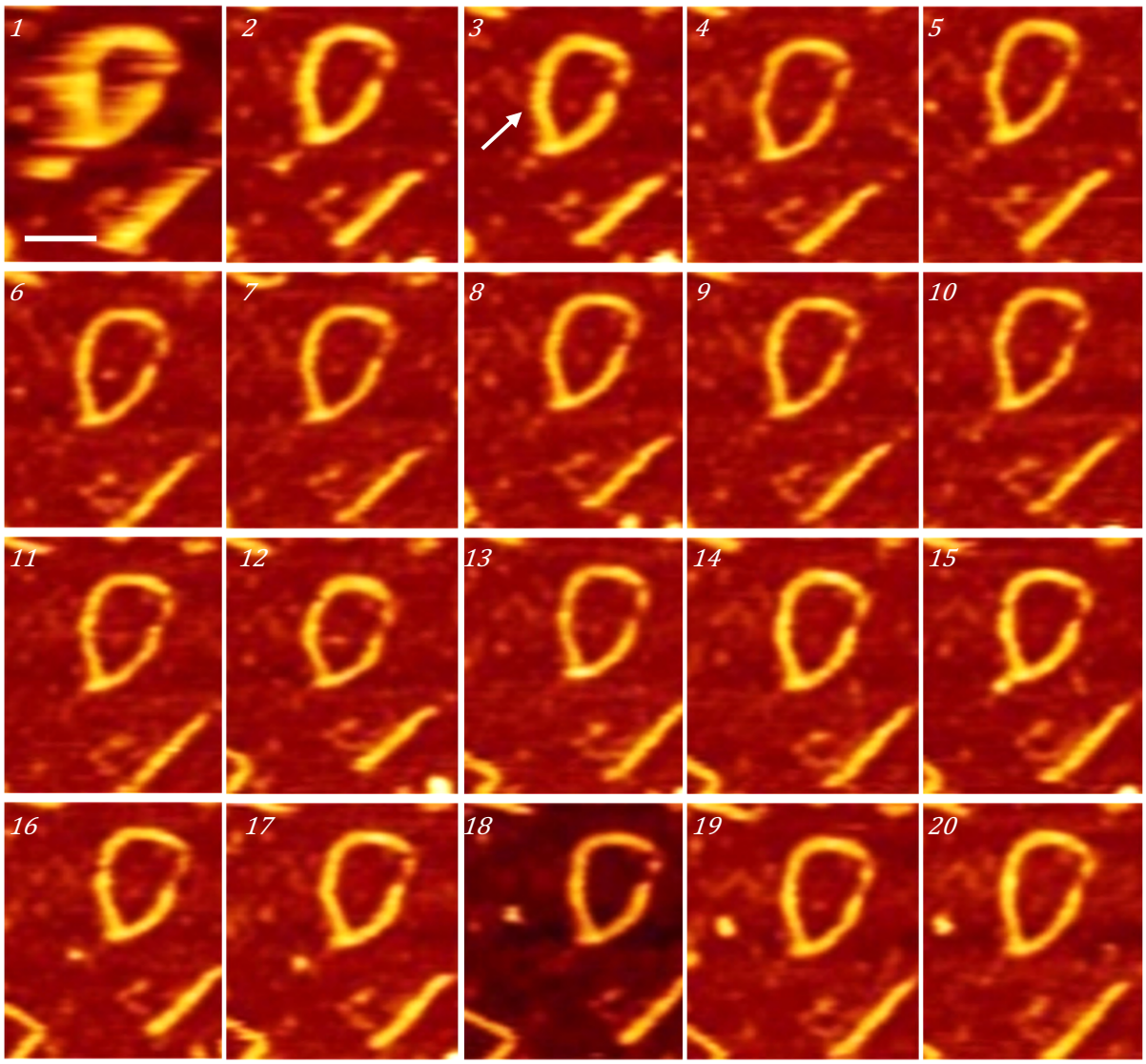


Figure 7.17: 300 bp flap DNA–TaqPolI (inactive FEN domain) imaged in buffer (25 mM HEPES, 2 mM Ca^{2+}, 5 mM K^{+}) immobilized on PLO treated mica, dynamic imaging: height 7 nm, scale bars 100 nm. The DNA conformation did not change before or after the binding of proteins on or around the DNA. Refer Movie E in Supplementary Material.

onwards (16 minutes after the protein was added), but neither did they appear close to the DNA nor did the latter show any movement, thereby indicating that there was no possible interaction between the DNA and the protein.

In figure 7.18 as well, there were 2 DNA molecules, labeled ‘a’ and ‘b’, which appeared to have a ssDNA between them, either a flap from the middle of ‘a’ or an overhang at the end of DNA ‘b’. The protein (encircled) appeared for a few frames (frames 1-9) 16 minutes after it was first added in a low concentration. Then, it was not visible in the later frames, *i.e.* after frame 9. Also, there was no movement of the DNA molecules, thus indicating an absence of DNA-protein interactions. In both the above cases (figure 7.17 and 18) the concentration of the

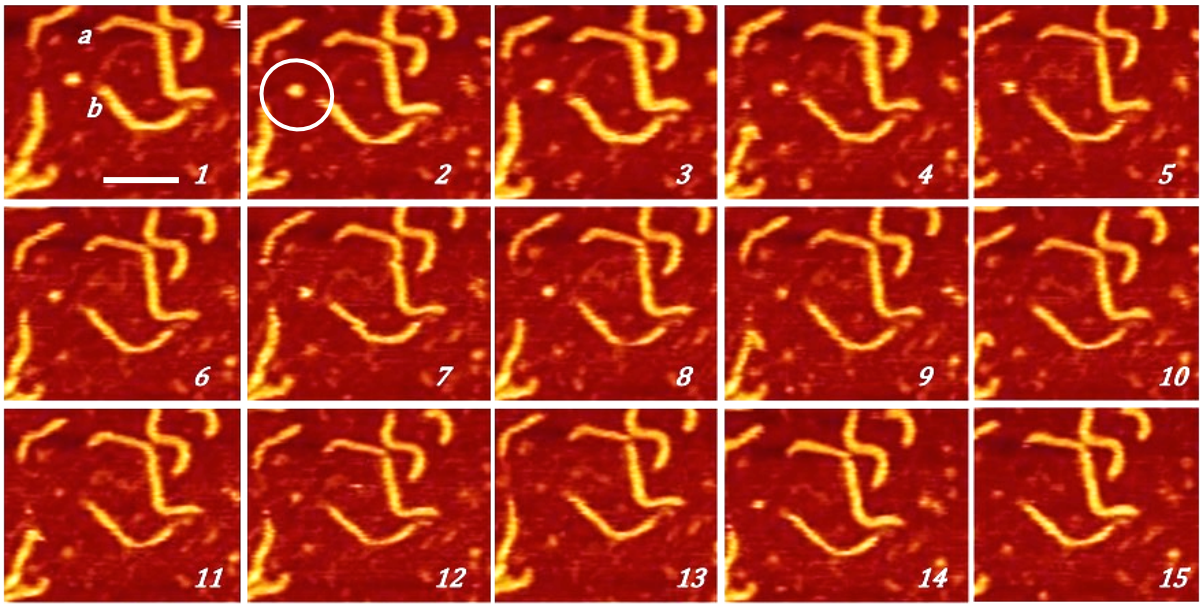


Figure 7.18: 300 bp flap DNA–TaqPolI (inactive FEN domain) imaged in buffer (25 mM HEPES, 2 mM Ca^{2+}, 5 mM K^{+}) immobilized on PLO treated mica, dynamic imaging: height 7 nm, scale bar 100 nm. The DNA conformation did not change before or after the binding of proteins on or around the DNA. Refer Movie F in Supplementary Material.

protein was probably not adequate. Hence, in the following experiments, the protein was added in a higher concentration, or if need be, was added twice after a few frames were imaged.

Additionally, in figure 7.19 (see appendix 10 for all frames of images, only frame 11-25 shown here), the protein did not interact with the flap DNA very well. The DNA encircled in frame 11 consisted of a flap structure. Even though the protein solution was added about 30 minutes before the protein first appeared, the number of protein molecules immobilized here was not large enough to saturate the surface (even after about 70 frames of 30 s each, imaged before the frame 1 shown here). The initial frames did not show any protein and there was no movement of the DNA molecule observed. The protein molecules did not immobilize close to the flap DNA until much later (encircled in frame 18) and even then, no movement of the DNA molecule, or any other molecule was seen. One DNA molecule, encircled in frame 20 was immobilized well in the previous frames but was knocked off between frames 20-21. Except this movement of DNA (which might have been random), there was possibly no DNA-protein binding occurring in these images.

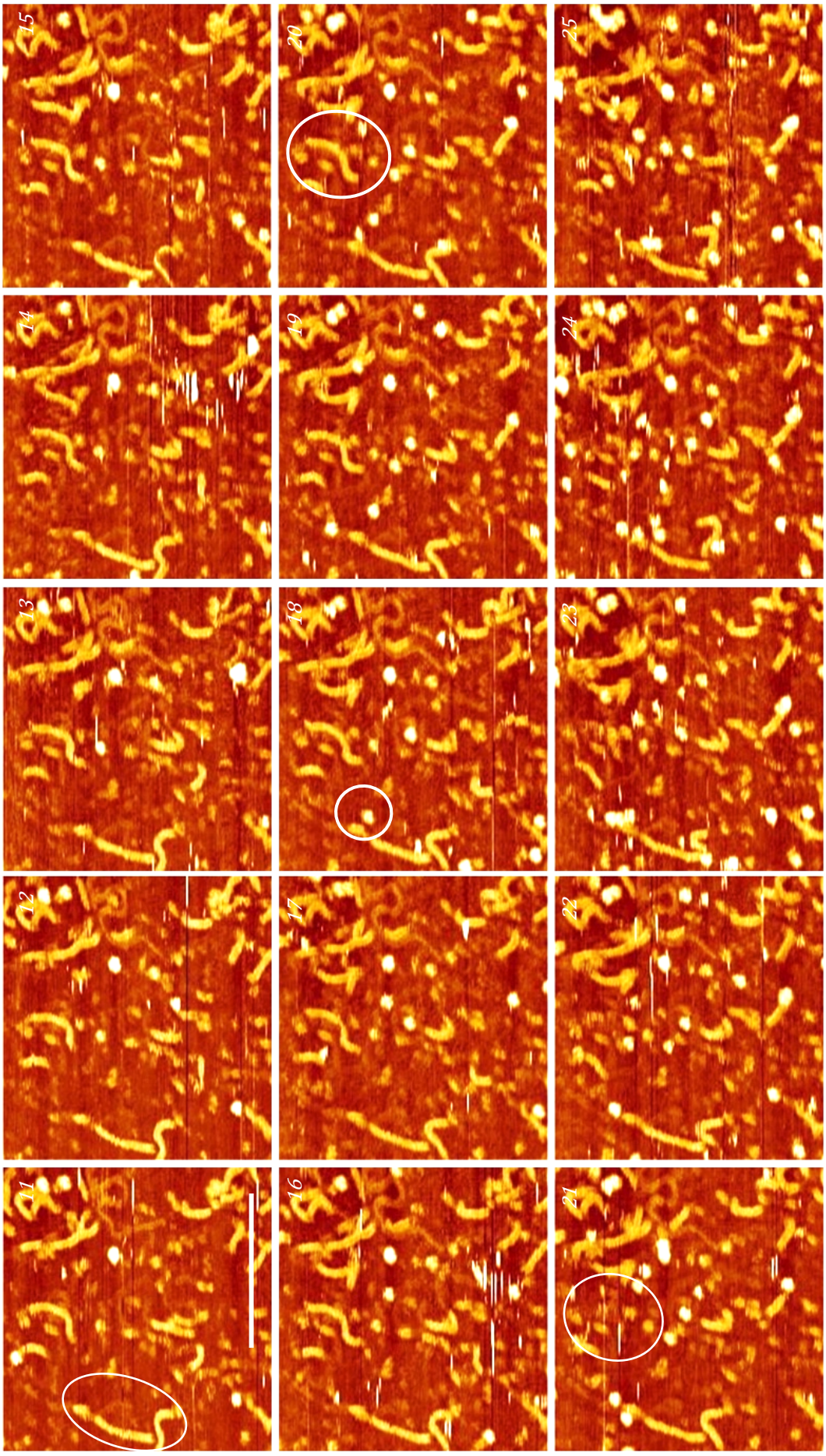


Figure 7.19: 300 bp flap DNA–TaqPolI (inactive FEN) imaged in buffer (25 mM HEPES, 2 mM Ca^{2+}, 5 mM K^{+}) immobilized on PLO treated mica, dynamic imaging: (see appendix 10 for all frames of images, only frame 11-25 shown here)height 8 nm, scale bar 100 nm. The DNA conformation did not change before or after the binding of proteins on or around the DNA. Refer Movie G in Supplementary Material.

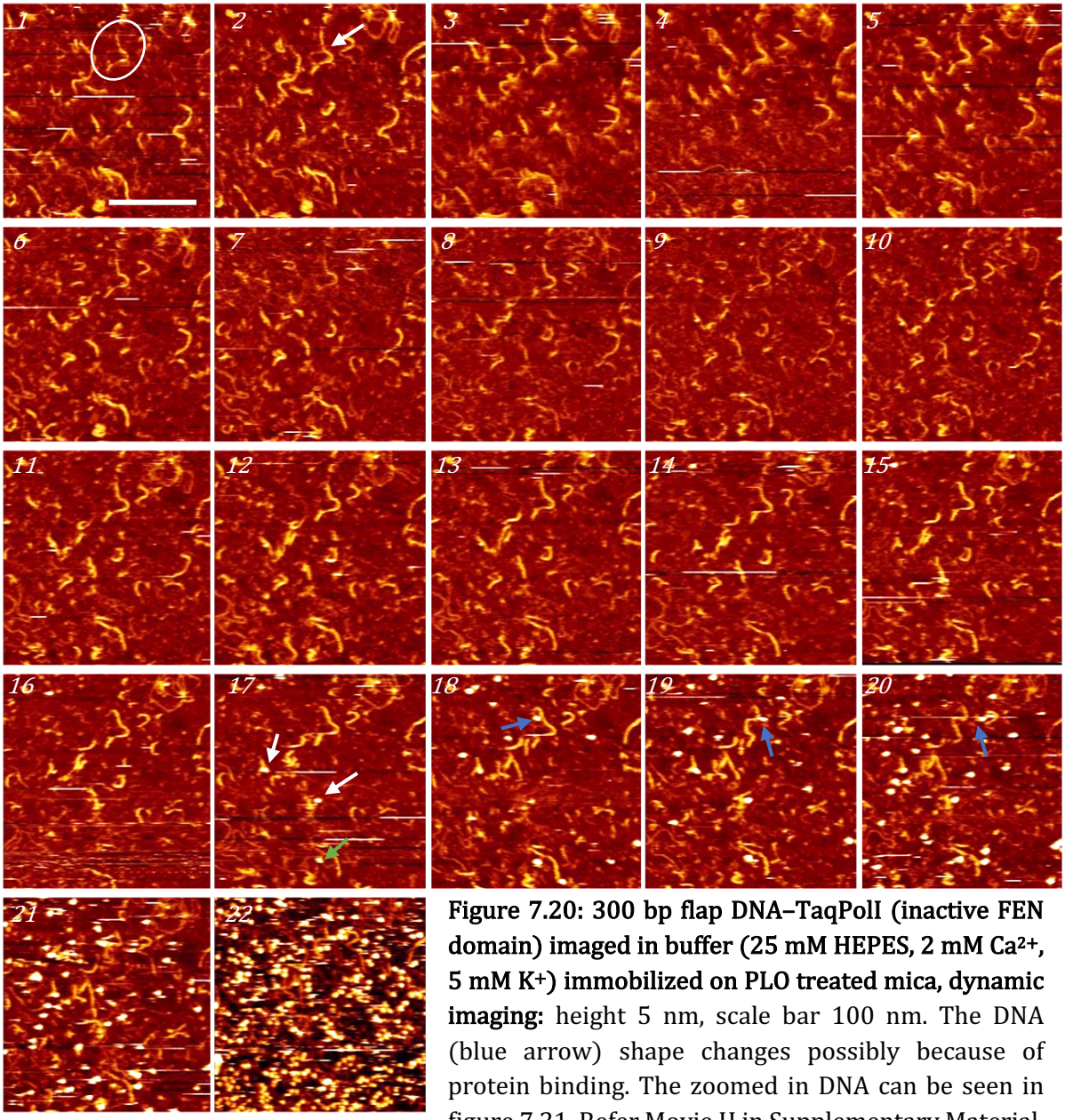


Figure 7.20: 300 bp flap DNA–TaqPolI (inactive FEN domain) imaged in buffer (25 mM HEPES, 2 mM Ca^{2+}, 5 mM K^{+}) immobilized on PLO treated mica, dynamic imaging: height 5 nm, scale bar 100 nm. The DNA (blue arrow) shape changes possibly because of protein binding. The zoomed in DNA can be seen in figure 7.21. Refer Movie H in Supplementary Material.

Contrary to the cases so far, there were some images where the DNA conformation did change because of possible binding by the protein. One such interaction could be seen in figure 7.20-21. The encircled DNA molecule in frame 1 showed a flap structure arising at the point marked by the arrow in frame 2. The shape of the DNA remained consistent throughout the frames in the absence of a protein molecule in the vicinity of the DNA. The protein was added in frame 1 and no protein appeared till frame 15, about 15 minutes after the addition of protein. Therefore, another aliquot of protein solution was added during frame 16 and immediately the molecules appeared in frame 17, marked by arrows. The protein marked by green arrow interacted with the DNA having a ssDNA flap and in the consecutive frames, remained bound to the DNA without causing any visible change in its conformation. This was possibly due to

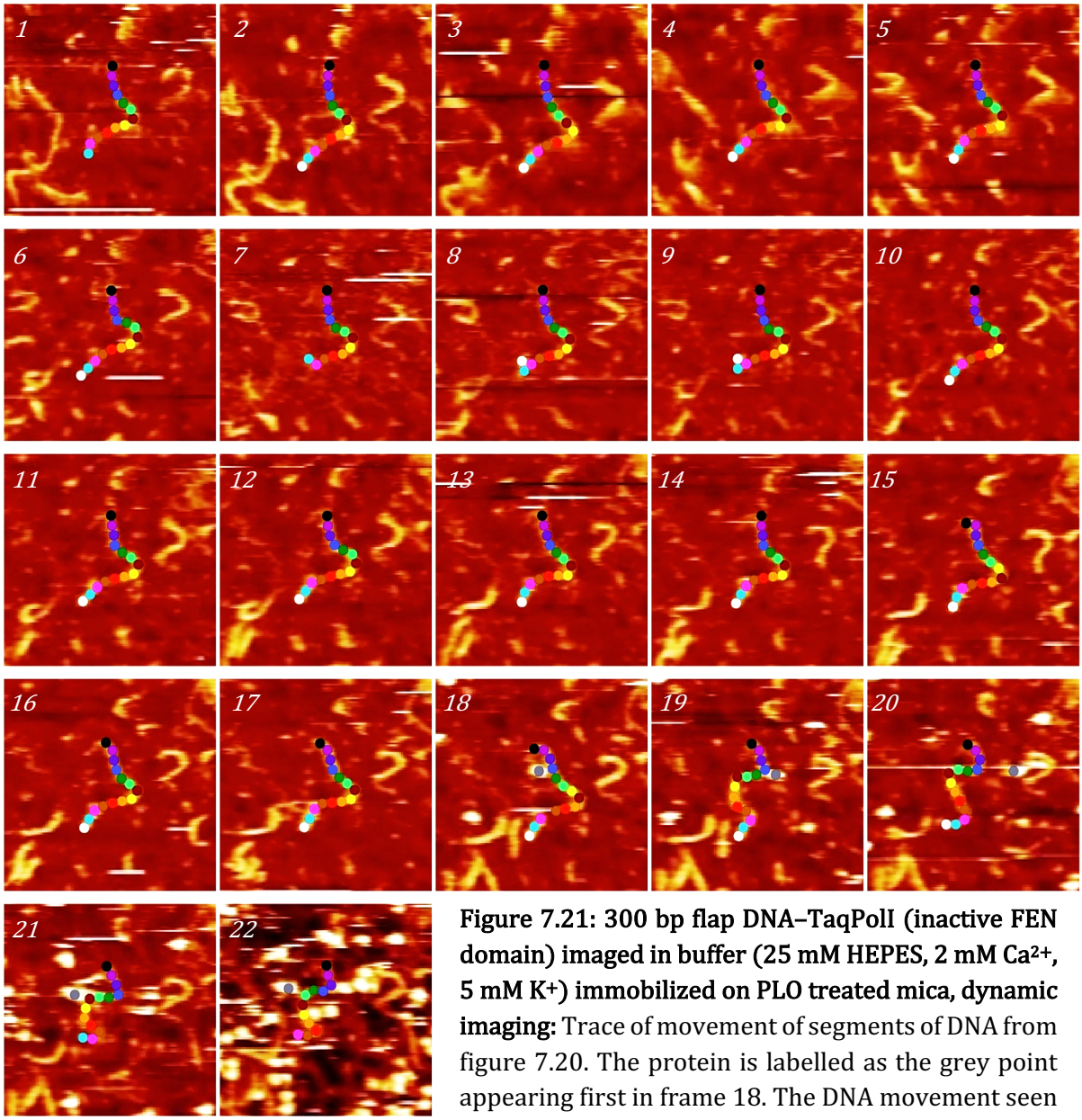


Figure 7.21: 300 bp flap DNA–TaqPolI (inactive FEN domain) imaged in buffer (25 mM HEPES, 2 mM Ca^{2+}, 5 mM K^{+}) immobilized on PLO treated mica, dynamic imaging: Trace of movement of segments of DNA from figure 7.20. The protein is labelled as the grey point appearing first in frame 18. The DNA movement seen clearly after interaction with the protein.

the low clarity of ssDNA in these images restricting the observation of any movement of ssDNA, if any occurred. On the other hand, another protein molecule, marked by blue arrow in frame 18, interacted with a well-immobilized DNA molecule. The protein apparently bound very close to the junction of ss-dsDNA causing it to move (frame 19) and then stayed more or less in that position. The other DNA molecules showed no change. This indicated that the protein bound to the DNA molecule and in the process moved it. The protein molecules crowded the surface after frame 22 restricting the observation of any interactions.

To understand and quantify the motion of DNA with and without the presence of protein, the DNA molecule under discussion was segmented into fragments of different colours and each

segment tracked using the software 'Tracker: Video Analysis and Modelling Tool'. This allowed the measurement of displacement of each of the DNA molecules in each frame of dynamic imaging. The displacement-frame plot thus obtained allowed an understanding of the way the DNA moved due to protein binding or interaction. However, it should be mentioned that due to the 'breathing' of AFM images, the molecules that are well immobilized also depicted some displacement values. But the presence of synchronicity of the segments indicated that the image was shifting over the frames, and not the molecules. This argument was used to comprehend the change in conformation of DNA in the presence of protein.

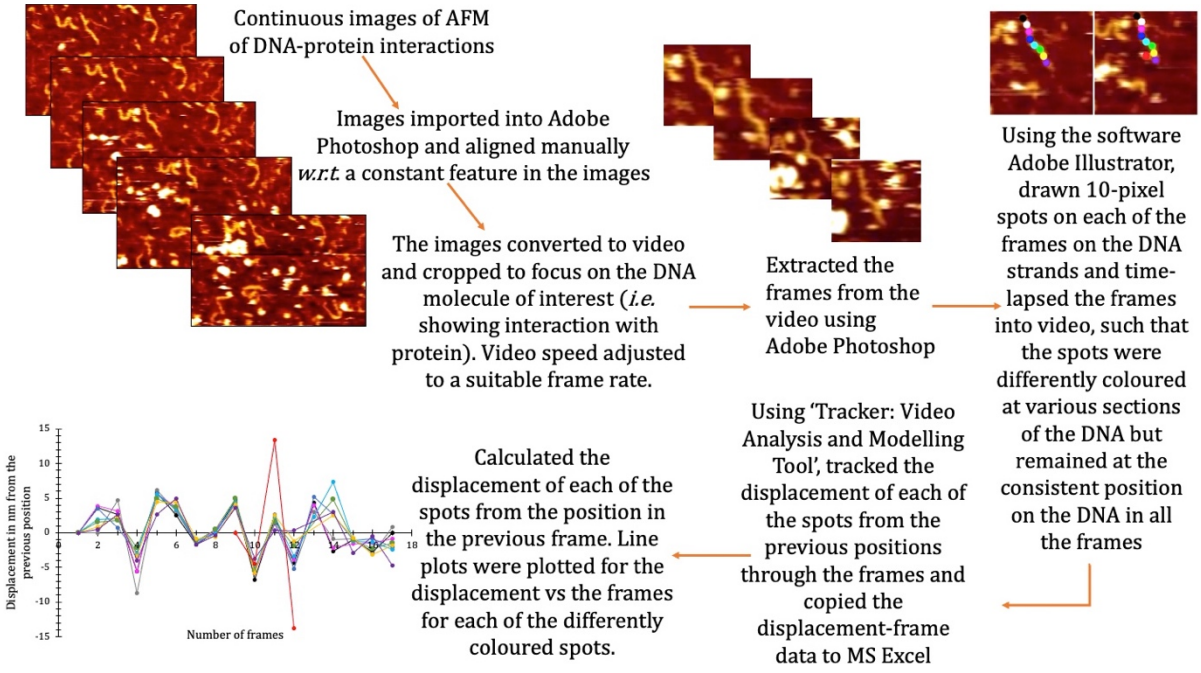


Figure 7.22: Outline of the image processing into videos for tracking the movement of sections of DNA under the influence of the protein.

DNA Motion Tracking: A method of tracking the motion of DNA under the influence of protein was developed (figure 7.22). Adobe Illustrator was used to mark segments of 5 pixel (or 10 pixels in the images where DNA was longer so reducing the number of segments and make the plots less crowded) on the DNA molecule in each of the frames. Each of the DNA segment were differently coloured to be tracked independently without confusion. These frames were then assembled into a video using Adobe Photoshop (by creating Time-lapse). The software 'Tracker: Video Analysis and Modelling Tool' was used to track each of the DNA segments and their respective displacements for each frame were plotted using Microsoft Excel. It should be emphasized that the displacements of each of the segments was measured as a magnitude of displacement vector from the previous position, and **not** from the initial

starting position. The plots allowed us to understand which section of the molecule moved and when*, w.r.t.* the position and time of the protein's appearance in the vicinity of the DNA molecule.

Figure 7.21 shows the DNA labelled with markers of 5-pixel length and figure 7.23 shows the displacement-frame plot corresponding to tracing the DNA markers. The images are the same set of scans as figure 7.20. As already discussed above, the DNA appeared to be well immobilized in figures 7.20-21 until frame 18 where the protein first appeared close to the DNA. From then on, the shape of the DNA changed indicating that the protein had bound to DNA causing the latter to come off the surface, move and immobilize again in a different conformation.

This could be corroborated by the plot in figure 7.23. The displacement lines of all the fragments/markers appear to be mostly synchronised until frame 18, with some markers showing slight deviations, namely the light pink and sky-blue markers around frames 6-11, that indicated the end of the DNA. These fluctuations were possibly due to the end of the DNA being more mobile. However, a significant fluctuation in the plots could be seen in frames 18-22 where the marker displacements were not only large but significantly unsynchronised, indicating the movement and conformational change of the DNA molecule. This indicated that the protein bound to the DNA causing some mobilization of latter in the process.

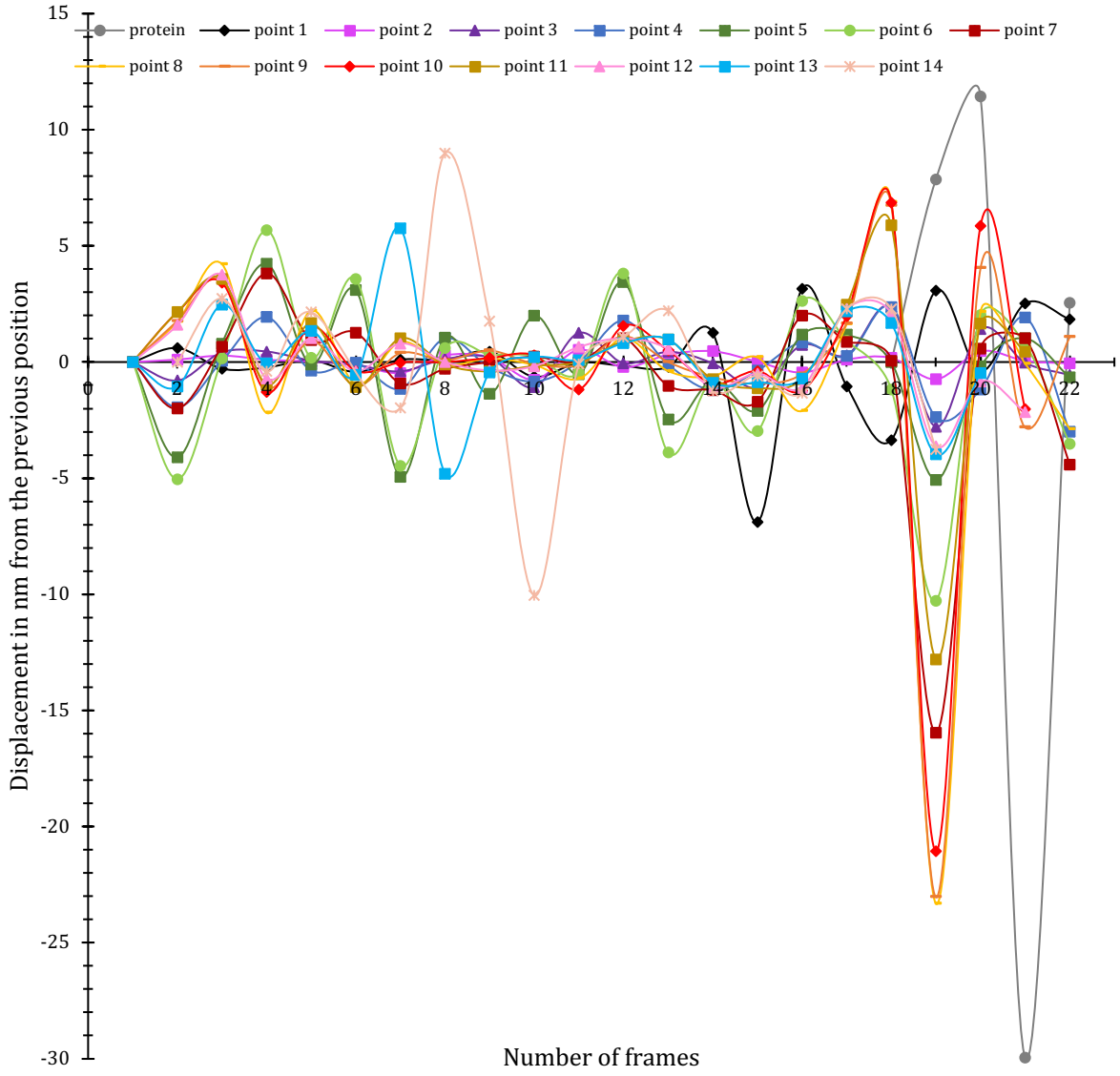


Figure 7.23: Displacement of segments of DNA in the presence of inactive FEN protein. The x-axis displays the number of frames and y-axis indicates the displacement in nm from their position in the previous frames of DNA fragments of colours corresponding to figure 7.21. The grey plot line indicates the motion of protein, which appears frame 18 in figure 7.20 and 7.21. The displacement of DNA segments increases after the protein interaction in frames 18 to 20.

Similarly, in figure 7.24, the DNA molecules circled in frame 1 each had a ssDNA overhang, arising from the points depicted by arrows in frame 2. The protein solution was added in frame 3, and about 2.5 minutes later, protein molecules appeared in frame 6. In frames 8 and 9, they were seen bound to the ss-dsDNA junctions. Likewise, in figure 7.25, the protein could be seen binding to ssDNA overhang branch point in frame 8. In both figures 7.24-25, as well as in figure 7.26, the protein molecules crowded about 5 minutes after the protein was added (as figures 7.24-26 are different zoomed in sections of the same set of images).

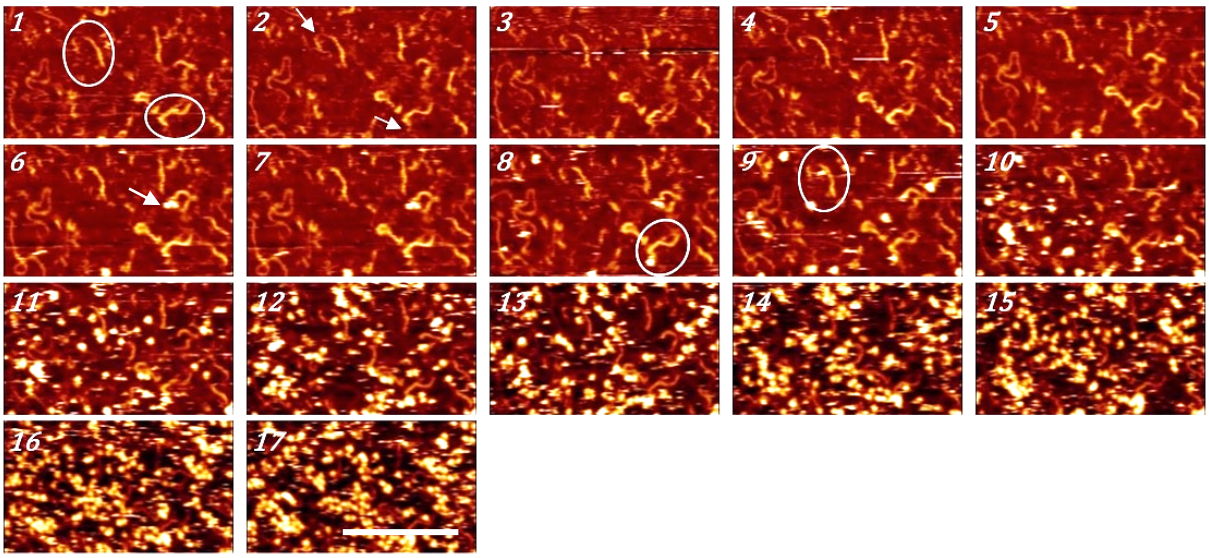


Figure 7.24: 300 bp flap DNA–TaqPolI (inactive FEN domain) imaged in buffer (25 mM HEPES, 2 mM Ca^{2+}, 5 mM K^{+}) immobilized on PLO treated mica, dynamic imaging: height 5 nm, scale bar 100 nm. The encircled DNA shows the molecules to which the protein appeared to bind. Refer Movie I in Supplementary Material.

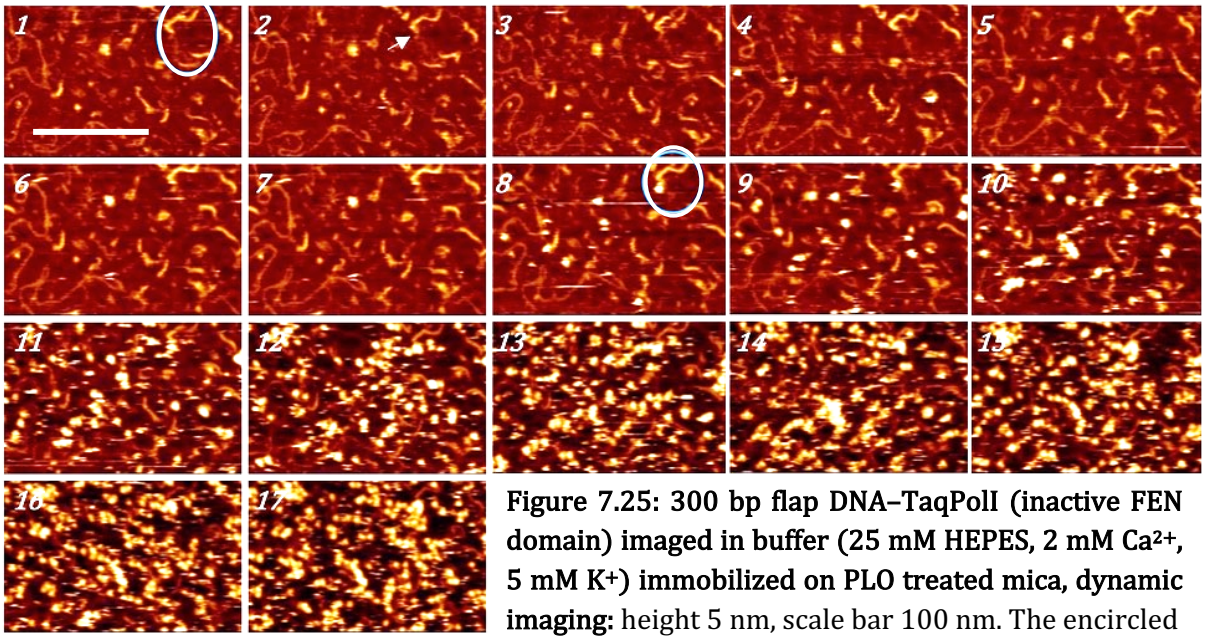


Figure 7.25: 300 bp flap DNA–TaqPolI (inactive FEN domain) imaged in buffer (25 mM HEPES, 2 mM Ca^{2+}, 5 mM K^{+}) immobilized on PLO treated mica, dynamic imaging: height 5 nm, scale bar 100 nm. The encircled DNA shows the molecules to which the protein appeared to bind. Refer Movie J in Supplementary Material.

Then, in figure 7.26, the DNA molecule encircled in frame 2 showed a flap DNA. The molecule's shape remained constant, even after the protein was added in frame 3, till the proteins appeared about 3 minutes later near it in frame 9, marked by blue arrow. From then on, the DNA molecule appeared to bend at the point of contact with the protein and in frames 10-12, the protein also moved around the axis of the DNA molecule (the motion of DNA can be seen zoomed in figure 7.27). After the 9^{th}-14^{th} frames of interaction, the DNA shape remained consistent, but could not be seen very clearly due to the crowding of proteins.

The above argument could be corroborated by figure 7.27 depicting the segmentation of DNA, under discussion in fig. 7.26, into 10-pixel segments and tracking the motion of each to obtain displacement-frame plots (fig. 7.28). The protein bound to DNA in frame 9 caused a slight bend in the DNA, seen visually in fig. 7.27, but did not produce any numerically significant deviation. The plot however indicated slight movement of some segments of DNA over frames 1-3, supported by minor lack of synchronisation of the plots. The presence of the protein from frame 9 onwards caused the DNA to move, and disturbed the synced line plots of displacement over frames 11-17. The protein and some section of DNA could not be tracked after frame 12 as there was crowding of other protein molecules. The DNA appeared to be more bent around the position of protein (greens, pale blue and yellow markers) in the later frames as compared to the straight form in the initial frames without the protein present. These observations thus signified that the protein binding to DNA caused a change in DNA mobility, possible bending and changing the conformation of DNA, especially near the position of protein binding.

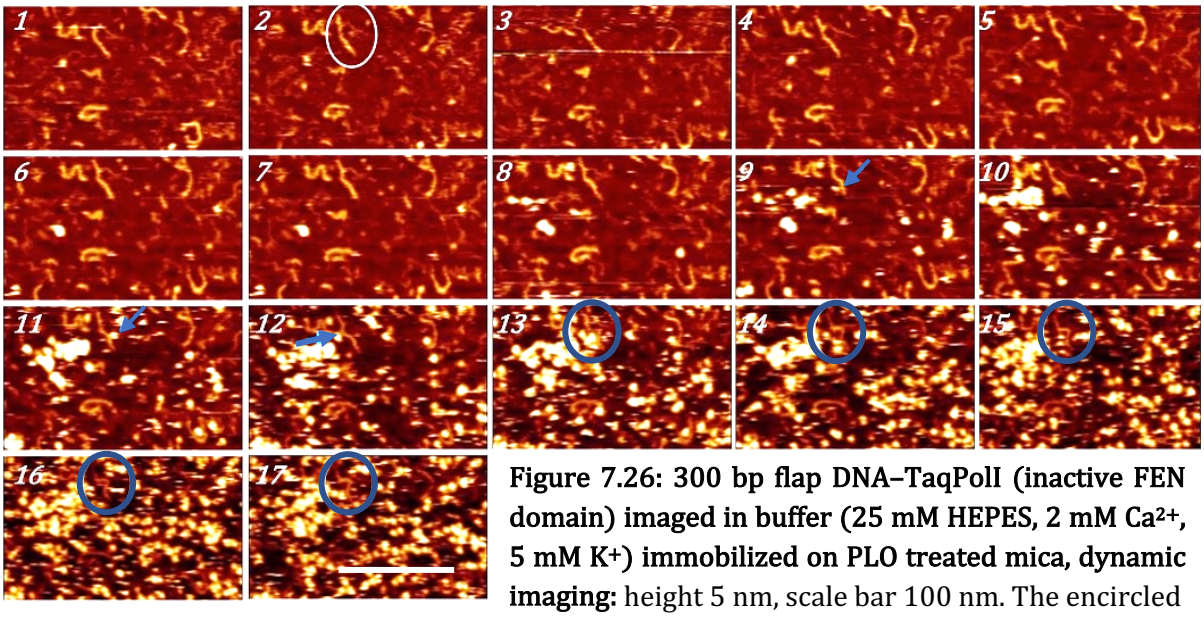


Figure 7.26: 300 bp flap DNA–TaqPolI (inactive FEN domain) imaged in buffer (25 mM HEPES, 2 mM Ca^{2+}, 5 mM K^{+}) immobilized on PLO treated mica, dynamic imaging: height 5 nm, scale bar 100 nm. The encircled DNA shows the molecules to which the protein appeared to bind and bend the DNA towards itself. Refer Movie K in Supplementary Material.

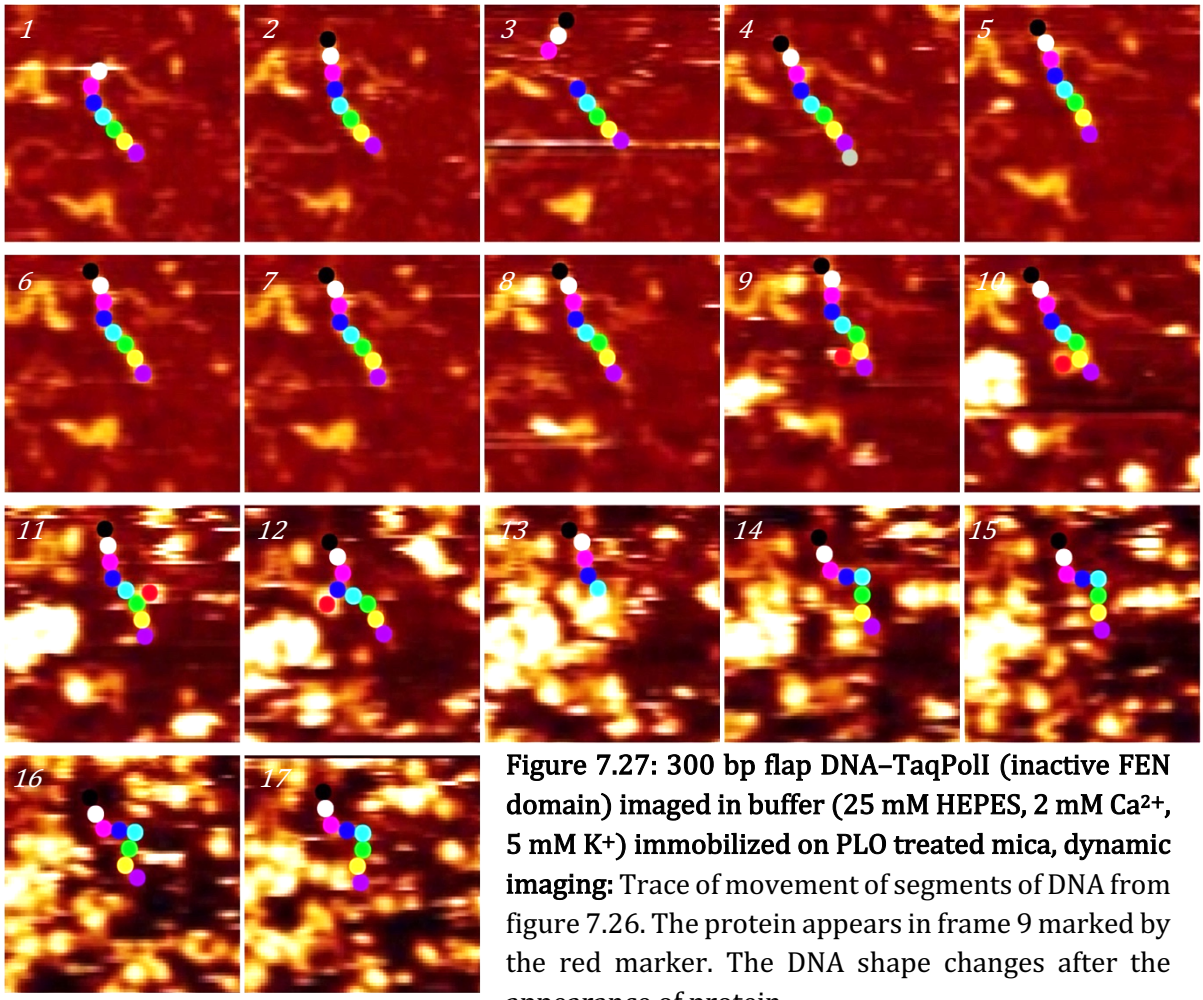


Figure 7.27: 300 bp flap DNA–TaqPolI (inactive FEN domain) imaged in buffer (25 mM HEPES, 2 mM Ca^{2+}, 5 mM K^{+}) immobilized on PLO treated mica, dynamic imaging: Trace of movement of segments of DNA from figure 7.26. The protein appears in frame 9 marked by the red marker. The DNA shape changes after the appearance of protein.

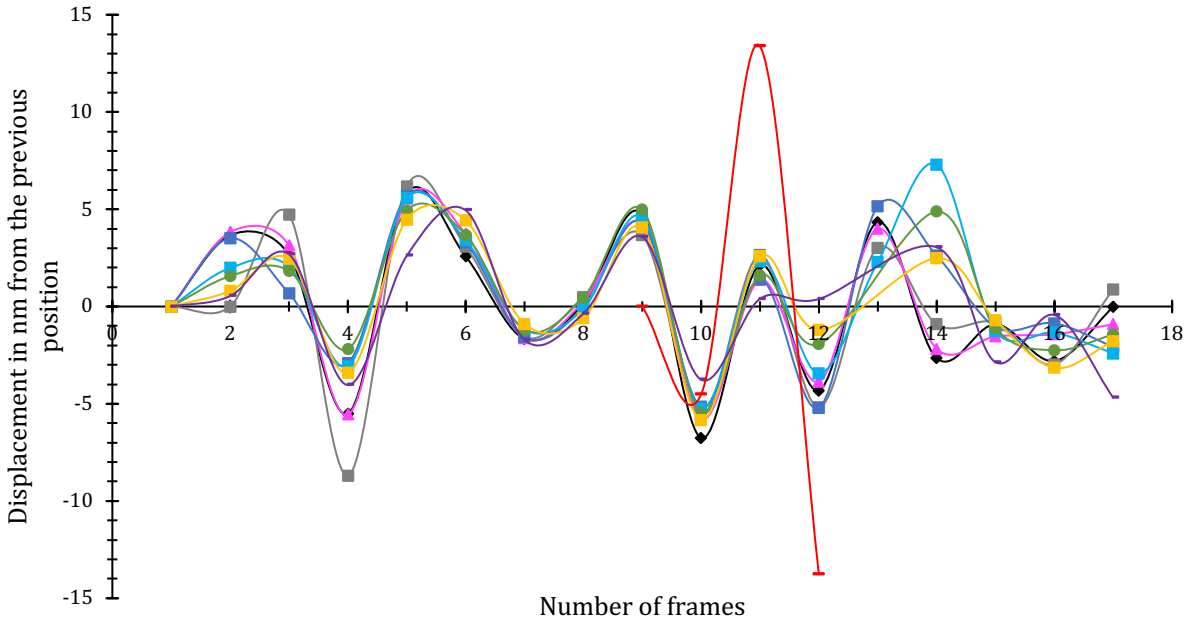


Figure 7.28: Displacement of segments of DNA in the presence of inactive FEN protein. The x-axis displays the number of frames and y-axis indicates the displacement in nm of DNA fragments of colours corresponding to figure 7.27. The red plot line indicates the motion of protein, which appears in the frame 9 in figure 7.26 and 7.27. The DNA moves between frames 13-14 and can be seen as a lack of synchronisation of the plot lines of the DNA displacements.

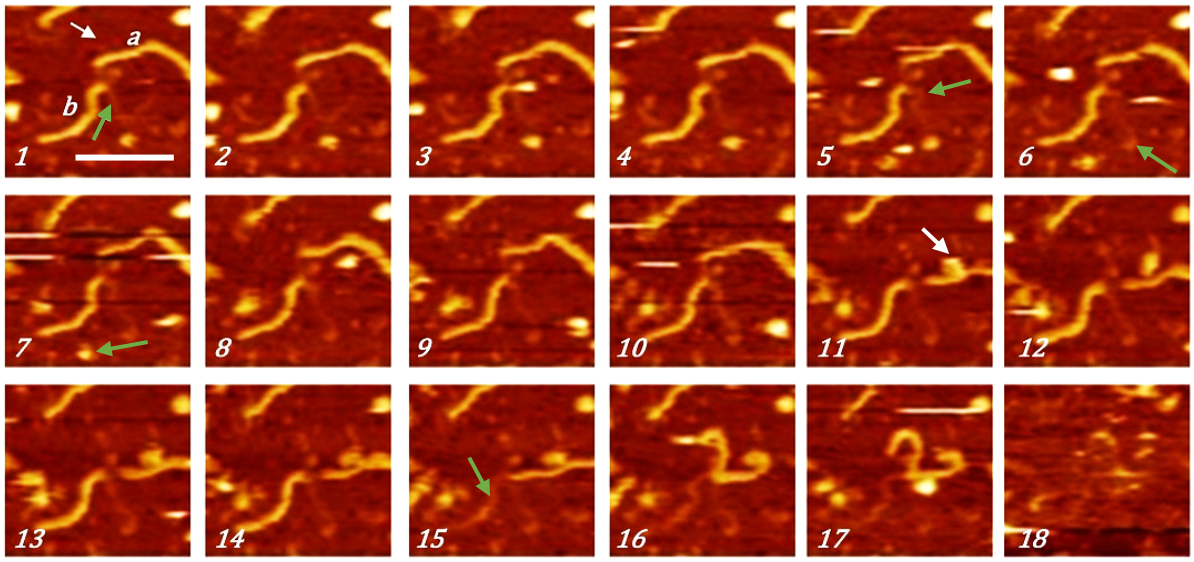


Figure 7.29: 300 bp flap DNA–TaqPolI (inactive FEN domain) imaged in buffer (25 mM HEPES, 2 mM Ca^{2+}, 5 mM K^{+}) immobilized on PLO treated mica, dynamic imaging: height 8 nm, scale bar 50 nm. The protein binding to DNA 'a' caused the DNA to move frame 11 onwards. DNA 'b' has ssDNA overhang that bent towards the protein in frame 6. Refer Movie L in Supplementary Material.

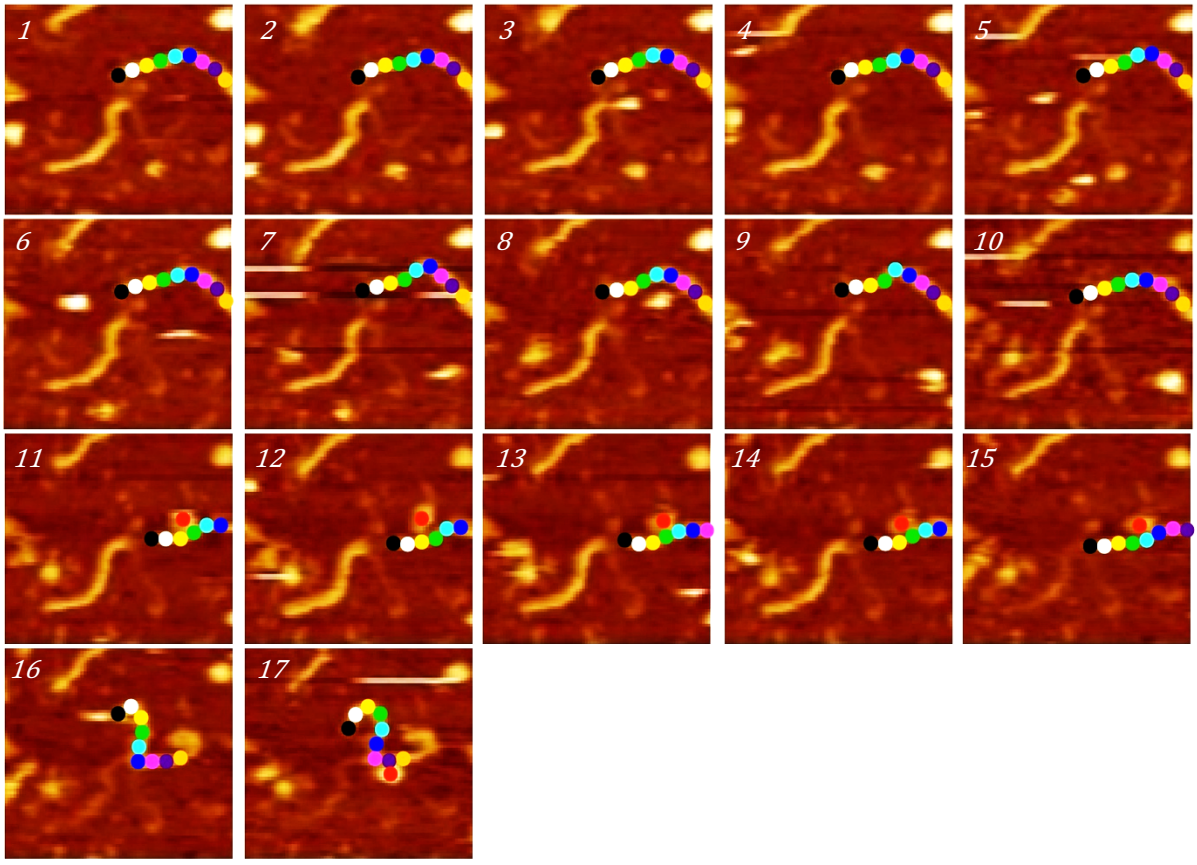


Figure 7.30: 300 bp flap DNA–TaqPolI (inactive FEN domain) imaged in buffer (25 mM HEPES, 2 mM Ca^{2+}, 5 mM K^{+}) immobilized on PLO treated mica, dynamic imaging: Trace of movement of segments of DNA from figure 7.29. The protein appears in frame 11 and marked by red marker. The shape of DNA can be seen to change after frame 11.

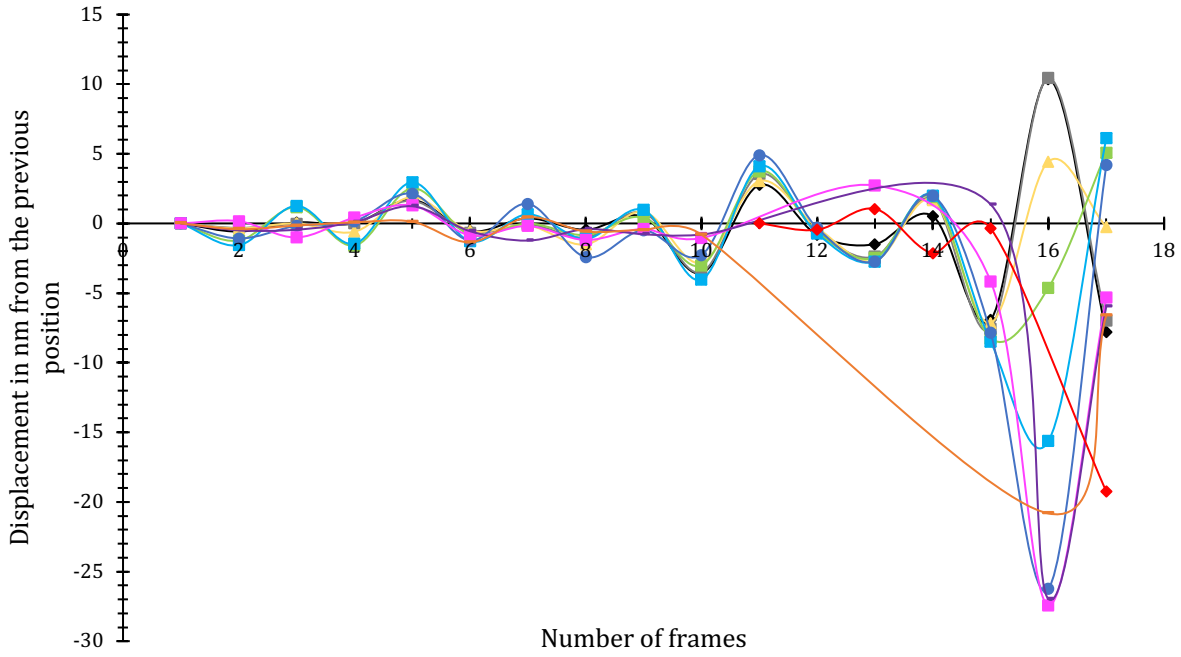


Figure 7.31: Displacement of segments of DNA in the presence of inactive FEN protein. The x-axis displays the number of frames and y-axis indicates the displacement in nm of DNA fragments of colours corresponding to figure 7.29. The red plot line indicates the motion of protein, which appears in frame 11 in figure 7.29 and 7.30. The displacement of the DNA fragments increases immediately in frame 11 and between frames 15-17 due to the presence of protein. The large gap in the positions of segment labelled as orange, in between the frames 10-16 is due to the DNA section being off the frame and not tracked for those frames.

In figure 7.29, two dsDNA molecules could be seen, 'a' and 'b'. DNA 'a' appeared to have a flap, marked by the white arrow, while 'b' had a ssDNA overhang at the end, marked by the green arrow. A protein molecule appeared to be interacting with the DNA from frame 1 onwards with DNA 'b' and with DNA 'a' from image 11 onwards. In case of DNA 'b', the molecule remained in a constant position before and after the protein was added, until the protein appeared close to it, as seen in frame 1. This first appearance of the protein occurred 6 minutes after the protein was added. In frame 5, after about 2 minutes of protein presence around it (each frame was ~ 30 s), the ssDNA overhang kinked (marked by the green arrow in frame 5) and the end of the overhang moved close to the protein (green arrow in frame 6). Following this interaction, the DNA remained immobile until frame 15, where the dsDNA segment probably moved and hence appeared faint (marked by green arrow) and later completely came off the surface in frame 18. The protein interacted with the DNA until frame 7, then moved away and was no longer visible in the consecutive frames.

DNA 'a', on the other hand, stayed immobilized until image 15. It didn't show any change in conformation until the presence of protein nearby in frame 11 (marked with a white arrow).

From then onwards, the DNA shape changed from a slightly bent form to a more or less straight shape. Also, the DNA moved such that some section of the molecule moved out of the frame of images. To quantify the movement of DNA due to protein interaction, the molecule was segmented by markers and each segment was tracked. The protein was marked in red (figure 7.30). Following the appearance of the protein close to the DNA in frame 11, the DNA segments which were initially displaced in synchronicity (figure 7.31), depicted immediate increase in displacement between frames 11-12. Also, the segments moved out of sync indicating motion of DNA, with large movement over frames 15-17. Here, the positions of the markers in the frames in which some DNA fragments were not visible, were interpolated. Thus, these series of images (figure 7.29-30) clearly depicted movement of DNA (both single stranded and double stranded), by the presence of protein.

Similarly, in figure 7.32, FEN binding with flap DNA caused the DNA to move. These frames were the last 30 frames of a 64-frame sequence. The initial frames were without the presence of protein even though the protein solution was added ~15 minutes before frame 1 here. There was no movement of the DNA molecule observed in any of those frames. The DNA encircled in frame 1 consisted of a dsDNA with no ssDNA visibly attached. A protein molecule immobilized close to it in frame 6, but there was no resultant change in the DNA conformation. However, in frame 16, another protein molecule immobilized close to the DNA and caused the DNA to move in frame 18, 1 minute (each frame is 30 s) after coming in contact. It is worth emphasizing that the DNA had retained a constant position for about 25 minutes during imaging but changed its conformation 1 minute after contact with this protein molecule. All the other DNA molecules remained immobilized in the previous and following scans.

The DNA was divided into segments and tracked to quantify the motion. The protein was labelled in red (figure 7.33) and appeared first in frame 16 causing the motion of DNA frame 18 onwards, as already mentioned. From the displacement-frame plot (figure 7.34) it was seen that the motion of the DNA markers was in sync until frame 17-18 where the displacement plots spread. From then on, the markers remained largely in sync, but not as well as the initial frames. Interestingly, the markers at the end of DNA, *i.e.* pink, black, and white/grey markers were the ones which showed large displacements even in the frames where the other markers were largely harmonized. This indicated that the ends of DNA were not perfectly immobilized. From these images, it could be concluded that the protein interacted with DNA to bind to it causing it to loosen from the surface briefly and immobilize again in a different position.

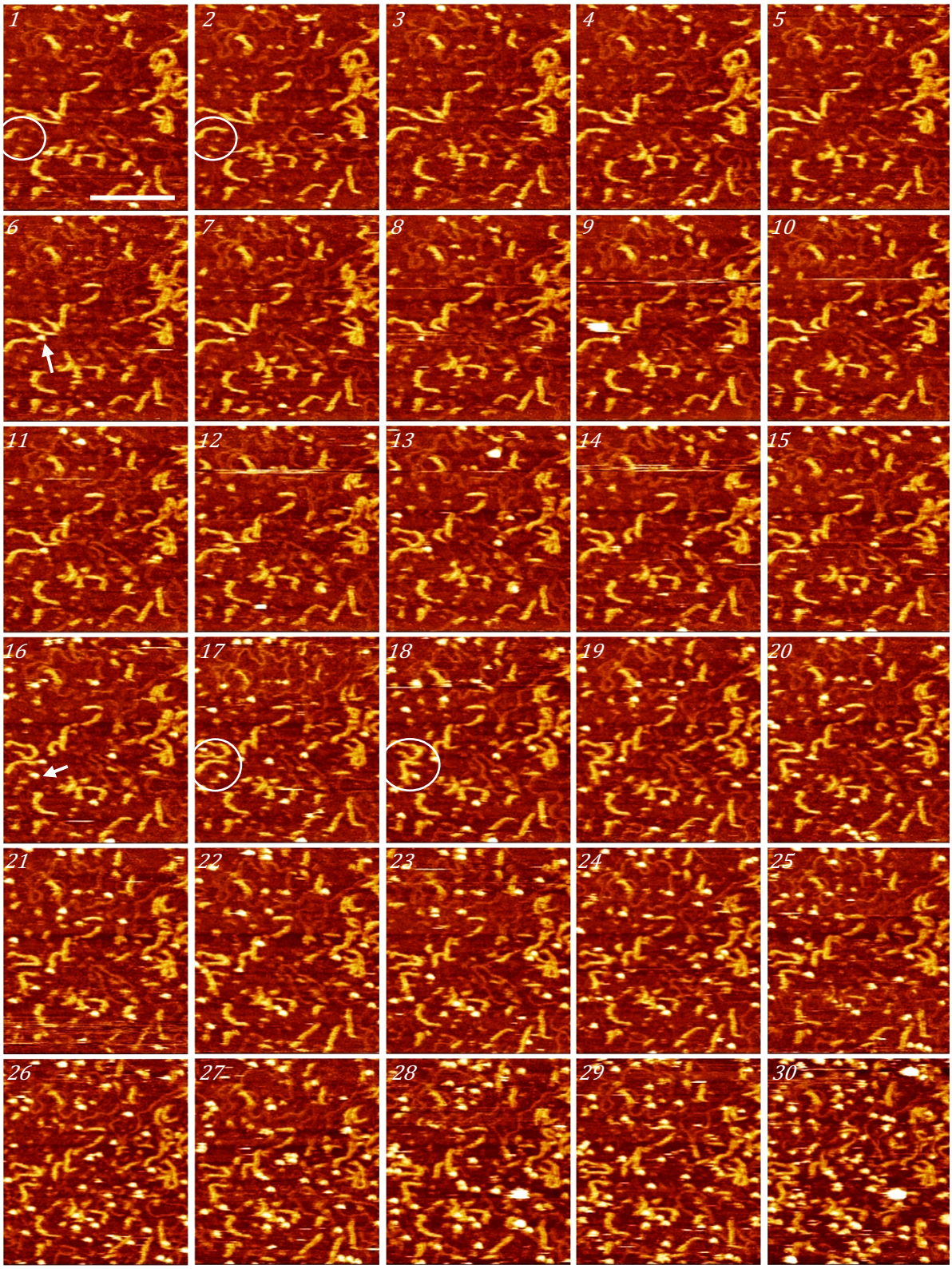


Figure 7.32: 300 bp flap DNA–TaqPolI (inactive FEN domain) imaged in buffer (25 mM HEPES, 2 mM Ca^{2+}, 5 mM K^{+}) immobilized on PLO treated mica, dynamic imaging: height 8 nm, scale bar 100 nm. The DNA encircled moved in frame 18 possibly because of protein binding in frame 16. Refer Movie M in Supplementary Material.

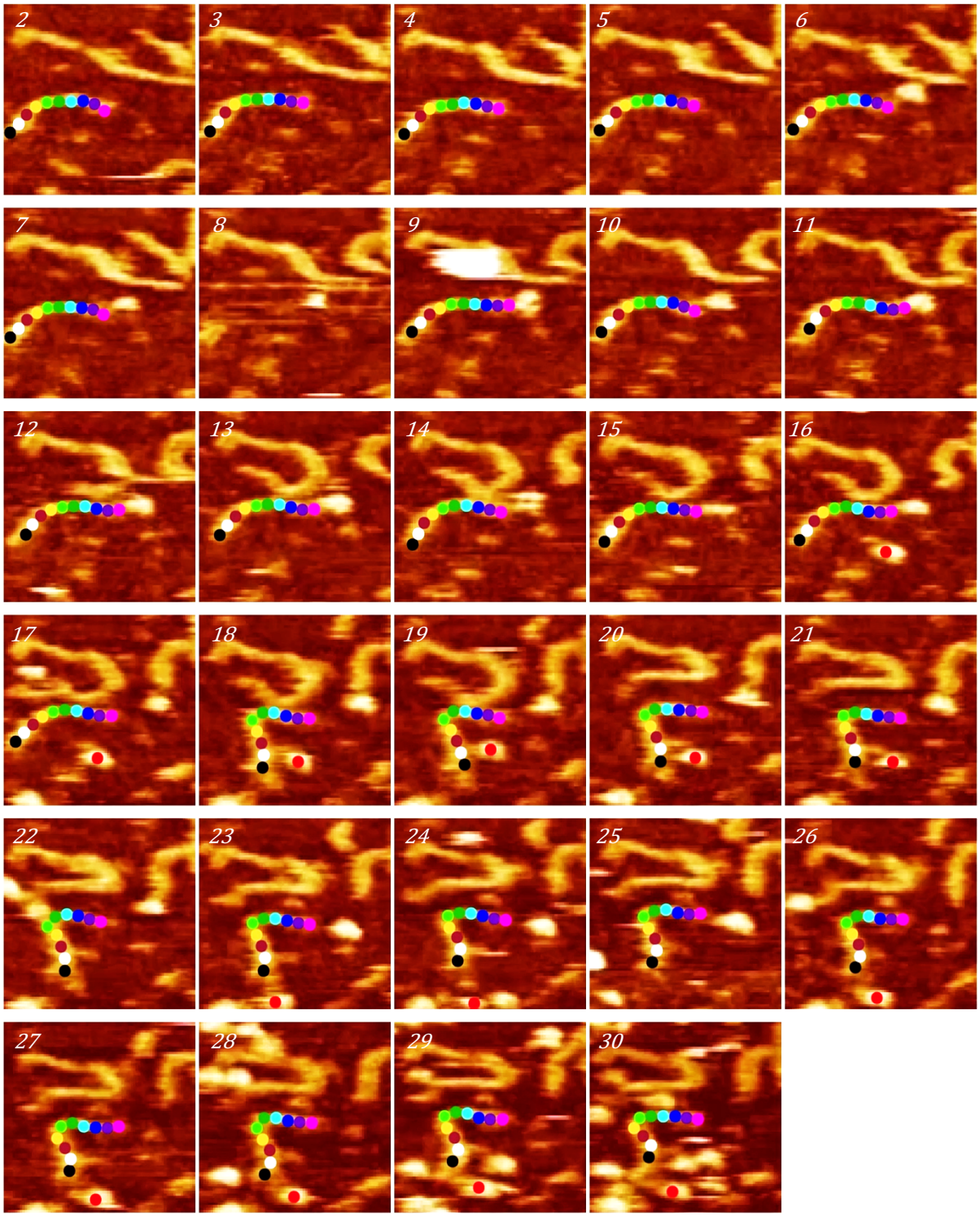


Figure 7.33: 300 bp flap DNA–TaqPolI (inactive FEN domain) imaged in buffer (25 mM HEPES, 2 mM Ca^{2+}, 5 mM K^{+}) immobilized on PLO treated mica, dynamic imaging: Trace of movement of segments of DNA from figure 7.32. The protein appears in frame 15 and is marked by the red marker. The DNA shape changed from frame 17 onwards.

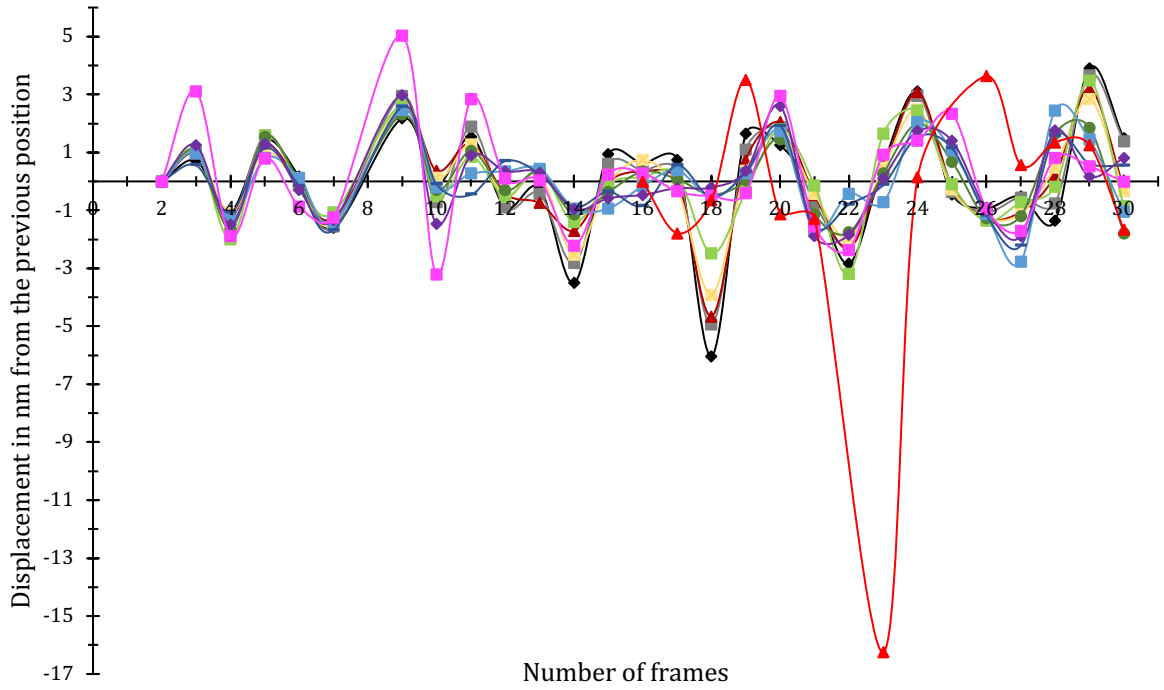


Figure 7.34: Displacement of segments of DNA in the presence of inactive FEN protein. The x-axis displays the number of frames and y-axis indicates the displacement in nm of DNA fragments of colours corresponding to figure 7.32. The red plot line indicates the motion of protein, which appears in frame 15 in figure 7.32 and 7.33. The DNA plot lines slightly move out of synchronisation in frames 17-21 indicating the effect of the presence of protein.

From the image discussed above, it could be understood that the protein molecule bound to DNA causing a change in the conformation of the latter because of mobilization of DNA from the surface due to protein binding, followed by immobilization again at the surface. There were many instances where the protein either did not immobilize around the DNA or immobilized on the PLO surface close to the DNA giving the false impression that the protein had bound the DNA substrate. In those cases, the DNA conformation change was not observed.

Proceeding from the observations made of DNA interaction with inactive FEN protein, the interactions of DNA were examined with active FEN protein, and this is discussed in the following section.

7.4.2. T_7 active FEN interacting with flap/overhang DNA

In the experiments so far, it was observed that the protein successfully bound to DNA and caused a change in its conformation despite it being well immobilized on the PLO treated mica. In this section, the interaction of DNA with active FEN protein was experimented with, in the presence of Mg^{2+} containing buffer, instead of Ca^{2+}. As already mentioned, Mg^{2+} enhance the activity of protein and act as a cofactor for the DNA polymerase I. In this case, the enzyme used was T_7 FEN, instead of the TaqPolI we had been using so far. This was to avoid the parameter of temperature for the reaction to occur as the T_7 enzyme could easily work at the 22ºC temperature of the Fastscan without additional heating required, as would have been the case for TaqPol (activity at 37ºC). It was interesting to note the changes on the DNA produced by the enzyme. The following paragraphs discuss the interactions observed when FEN was added to immobilized DNA (on PLO treated mica) in HEPES buffer (pH~8) solution containing Mg^{2+}, K^+ and DTT. It should be noted that the DNA was immobilized in a buffer solution containing Ca^{2+} and K^+ without any Mg^{2+} or DTT, to reduce the charge saturation by Mg^{2+} that would hinder with DNA immobilization to PLO treated mica. The absence of DTT allowed the reduction of noise by DTT sticking to the surface. But the imaging buffer contained Mg^{2+} and DTT as they were imperative for the activity and stability of the protein respectively.

Figure 7.35-36 consists of a set of images where each frame was 30 s long. The complete set of 77 frames can be seen in the appendix 11, but here only some frames are shown. The protein was added immediately before frame 1. A high concentration (10 µL of 550 µg/mL prepared in HEPES buffer (pH~8) solution containing Mg^{2+}, K^+ and DTT) of protein was added, yet the amount of protein immobilized around the DNA molecules was not sufficient, hence another 10 µL of protein solution was added during frame 19. Following this, the amount of protein immobilized increased significantly. Frame 1 shows a DNA molecule with a ssDNA that apparently attached to it (at the point marked by arrow in frame 4), but in the following frames, it became clear that it was not the case. It appeared that the ssDNA moved around the DNA molecules giving the impression of flap or overhang. The proteins immobilized around many DNA molecules (*eg.* those marked by arrows in frames 1 and 8) but no change in the shape of those dsDNA molecules occurred in all the frames except for the DNA molecule encircled in frame 1. It maintained its shape till frame 28, then moved. A protein molecule had immobilized

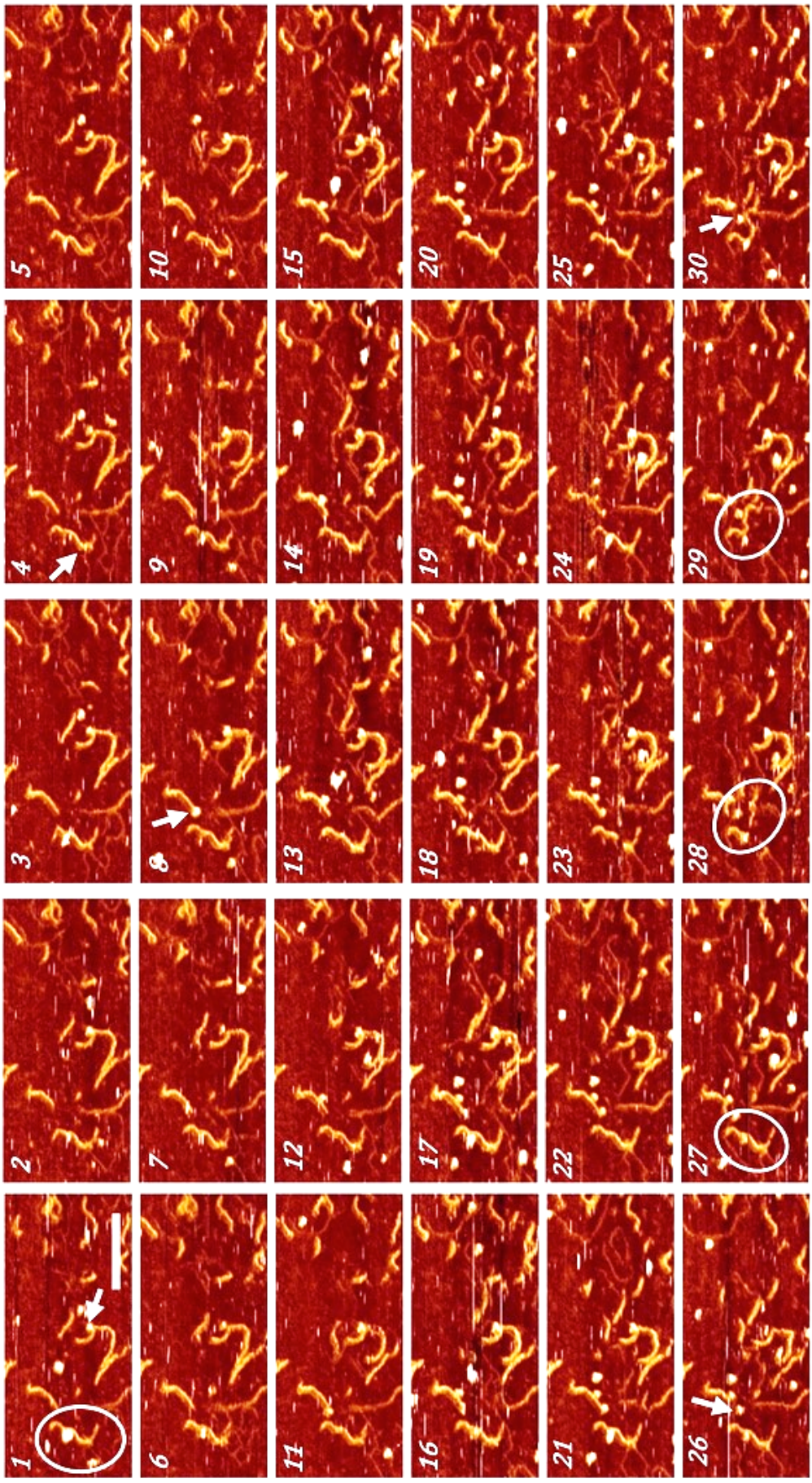
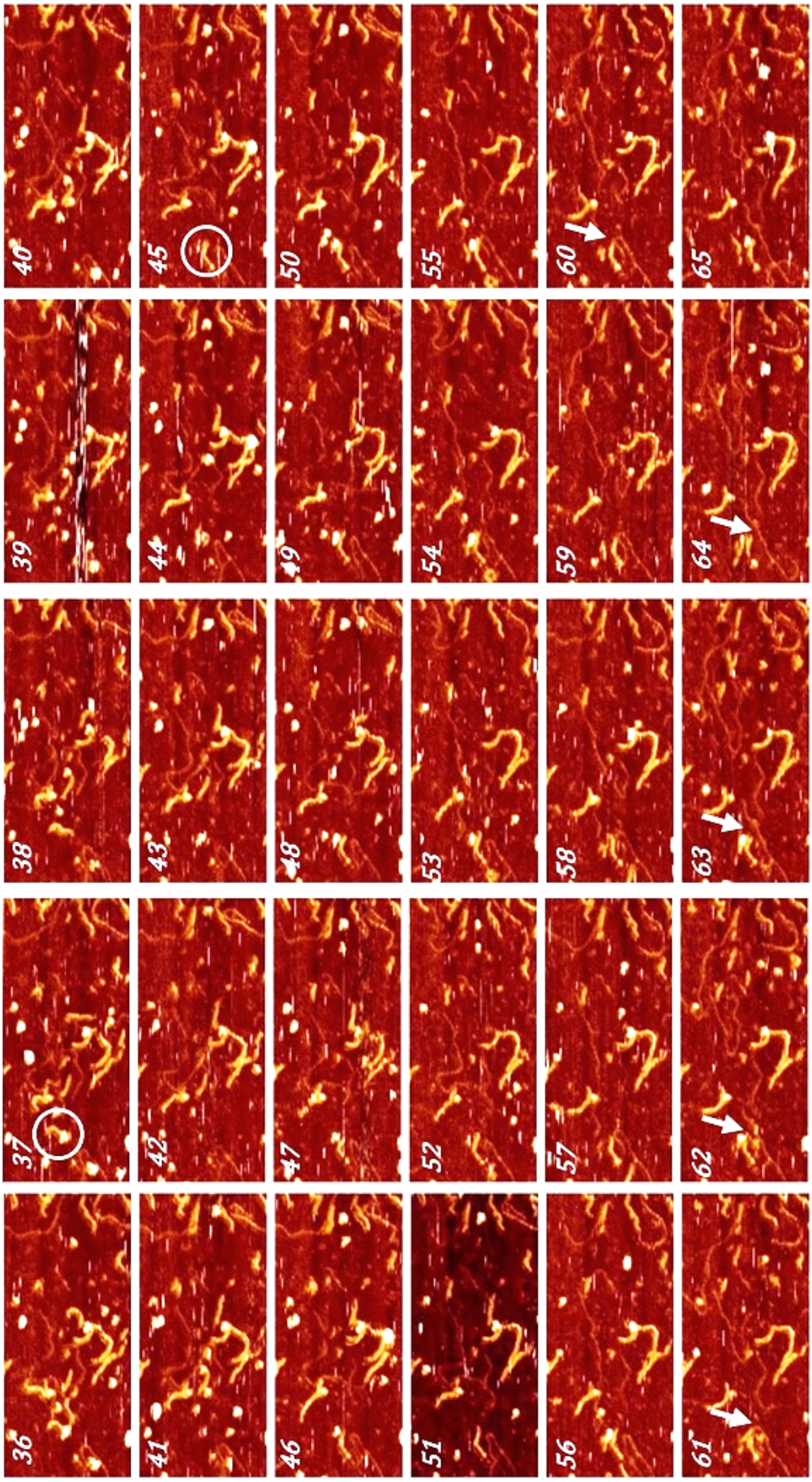


Figure 7.35: 300 bp flap DNA– T_7 (active FEN domain) imaged in buffer (25 mM HEPES, 2 mM Mg^{2+}, 5 mM K^+) immobilized on PLO treated mica, dynamic imaging: height 7 nm, scale bar 100 nm.(ctd...) Refer Movie N in Supplementary Material and Appendix 11 for all the frames.



(...ctd) Figure 7.36: 300 bp flap DNA– T_7 (active FEN domain) imaged in buffer (25 mM HEPES, 2 mM Mg^{2+}, 5 mM K^+) immobilized on PLO treated mica, dynamic imaging: height 7 nm, scale bar 100 nm. Refer Movie N in Supplementary Material and Appendix 11 for all the frames.

close to it (see arrow in frame 26) that caused the previously well-immobilized DNA to move about 1 min after apparently interacting with it (in frame 28). Then onwards, the protein molecule could be seen bound to DNA (indicated by the arrow in frame 30) and in frame 37, about 5 minutes after the protein binding to DNA, the dsDNA broke at this point of attachment and moved, and was not seen in the later frames. In frame 45, however, a new dsDNA with the same apparent size as the one mentioned above, became immobilized. It could have been another DNA molecule or the one mobilized in frame 37 as it also had a protein bound to it, but it was not possible to determine if it was the same one. All the other dsDNA molecules remained the same throughout, while there was (apparently random) movement of one ssDNA strand over frames 60-64 and following which it remained immobilized. It could be thereby understood that the protein had interacted with the dsDNA and caused it to move and break.

Figures 7.37-38 are another long set of images showing DNA cleaved by FEN (frames 85 to 108 of which can be seen in the appendix 12 and have not been shown here). The DNA molecule encircled in frame 1 had two ssDNA strands attached to it: the one at the top was a flap labelled ‘a’ while the one at the bottom was the overhang DNA labelled ‘b’. The negative control *i.e.* the buffer without the protein was added in frame 8 and the positive control i.e. the buffer with protein was added in frame 20. Frames 1-14 were 40 s long and the ones later were all 30 s long. There was no movement of the DNA seen due to the negative control solution. The protein molecule appeared in frame 42, ~12 minutes after its solution was added, while the DNA molecules remained well immobilized till frame 85. Additionally, there was a long ssDNA, that was visible clearly in frame 44. The protein appearing in frame 42 apparently hovered around it frame 42 onwards until it bound in frame 56. The shape of the ssDNA remained constant before this binding, but the interaction appeared to mobilize the DNA and it tended to change shape from the extended thread in frame 56 to slightly folding in 57. The ssDNA had consistent extended conformation previously but was disposed to fold in the following frames into various shapes. Another protein appeared to bind to this ssDNA in frame 60 and few more in frames 69 onwards. It is noteworthy that the protein molecules inclined to cluster around the ssDNA molecules, whether it was the above-mentioned ssDNA strand or the flap encircled in frame 53.

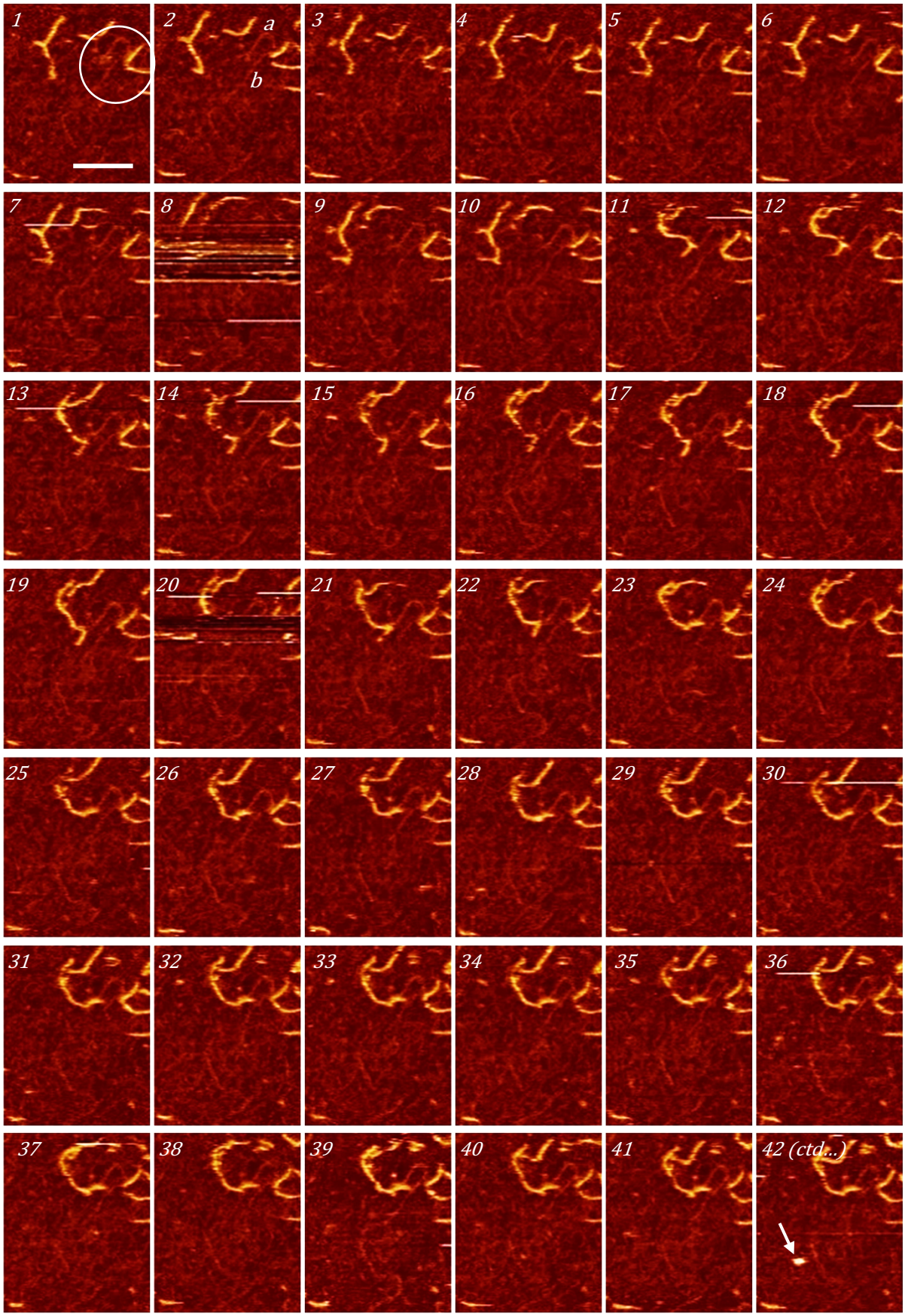


Figure 7.37: 300 bp flap DNA– T_7 (active FEN domain) imaged in buffer (25 mM HEPES, 2 mM Mg^{2+}, 5 mM K^{+}) immobilized on PLO treated mica, dynamic imaging: height 7 nm, scale 50 nm. Refer Movie O in Supplementary Material.

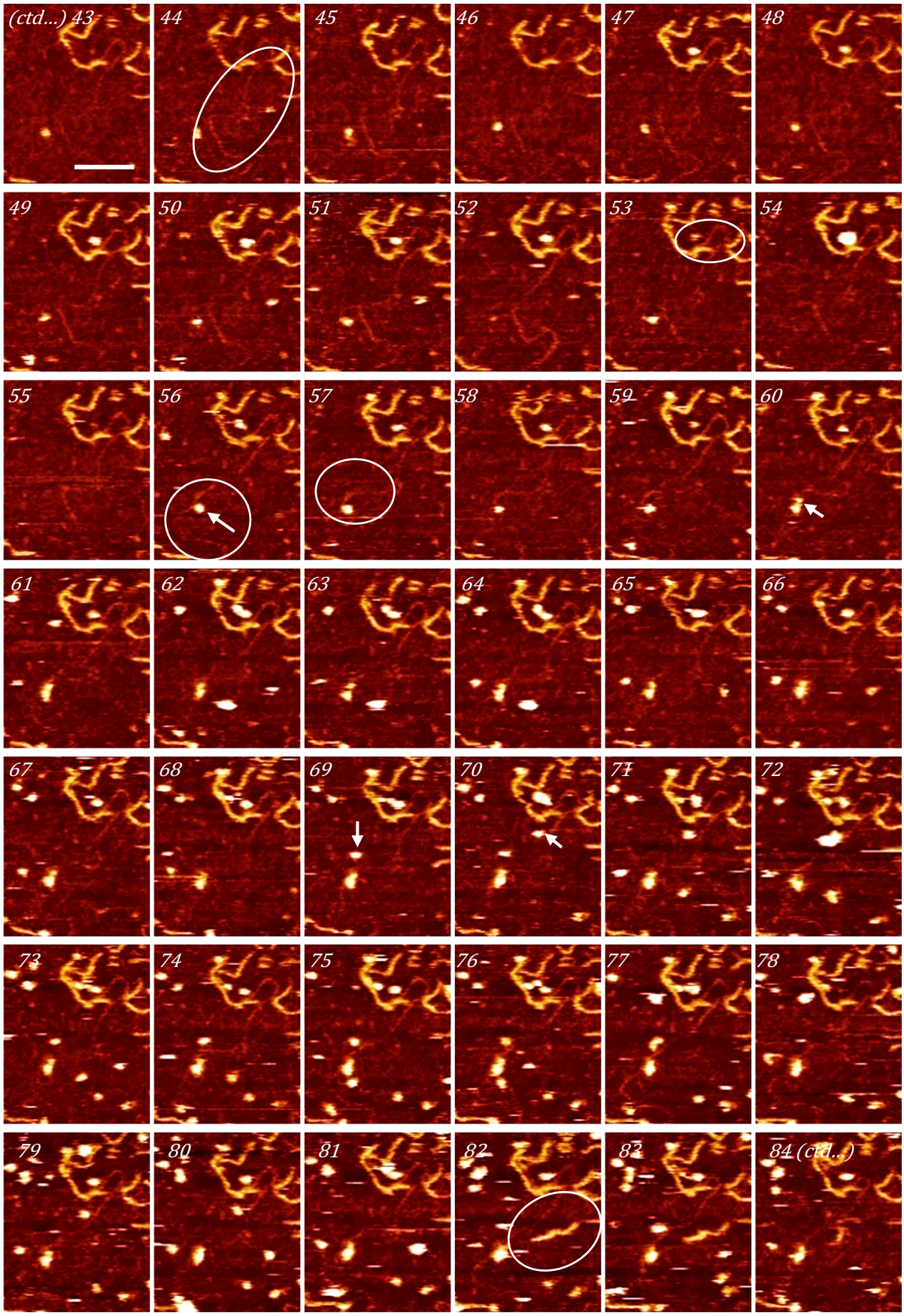


Figure 7.38: 300 bp flap DNA– T_7 (active FEN domain) imaged in buffer (25 mM HEPES, 2 mM Mg^{2+}, 5 mM K^{+}) immobilized on PLO treated mica, dynamic imaging: height 7 nm, scale bar 50 nm. Refer Movie O in Supplementary Material.

In addition to the above discussed cases, figure 7.39 shows a very interesting DNA-protein interaction process. The DNA molecule seen here was a ssDNA with a dsDNA attached close to one end. The protein appeared in frame 19, 17 minutes after the protein solution was added (each frame is 30 s and the protein was added a few minutes before frame 1 shown here). The DNA remained in a consistent shape till 23 where the ssDNA broke and the entire strands appeared to come off and slide the dsDNA in frame 24. To visualize the DNA behavior better, the molecule was divided into segments and the movement of the markers was tracked to obtain displacement-frame plots. As observed in figure 7.40 as well, a protein molecule (marked in red) could be seen bound to the end 'b' of the ssDNA in frame 24. This protein was partially visible at the bottom of the images in previous few frames. The ssDNA remained in that shape till 38 where it appeared to fold, and the protein attached to the end 'b' of the single strand bound to the other end 'a' of the ssDNA that had just come close to it. Frame 49/50 onwards, the protein then appeared to be bound to both the ends of the single strand instead of the lower end 'b' where it was bound before. The protein began to thread the upper end of ssDNA in the frames 50-61, after which the DNA was lost off the frames and the further reaction could not be seen. Thus, the protein binding to overhang DNA branch point had caused the ssDNA end to bend and thread along or through the protein, the exact nature of which could not be resolved.

The displacement-frame plot (figure 7.41) depicted DNA movement very different from the plots previously obtained for inactive FEN (section 7.4.1). The DNA fragments' displacement was largely synced in frames 21-23 and assumed to be thus in the previous images as the DNA remained very immobile. However, in frame 24, when the DNA started to fold, the displacement plots for the end of DNA (black, white, grey, and purple markers) were out of sync, with the extreme end of DNA (black) showing maximum shift from the synced curve of the other plots, followed by white, then grey and then purple. This indicated that the ends of the DNA had started to move bending the rest of the molecule with it. A similar plot behaviour could be observed in frames 27, 31, 36-37, 41, 43, 46, 47, 51 and 60, marking the frames of movement of the DNA end labelled 'a'. The end 'b' of DNA, marked by blues and greens began to move gently from frame 35 onwards, indicated by the large displacements, followed by frame 41, 43, 46, 51 and 60, where the displacement peaks occurred, signifying gentle gliding of the end 'b' of DNA as the end 'a' of DNA moves away in the later frames of imaging. The extreme ends of the DNA marked by orange, maroon, brown and prussian were only visible until frames 25 after which they could not be tracked. The displacement of some segments that were off the image in some frames were interpolated.

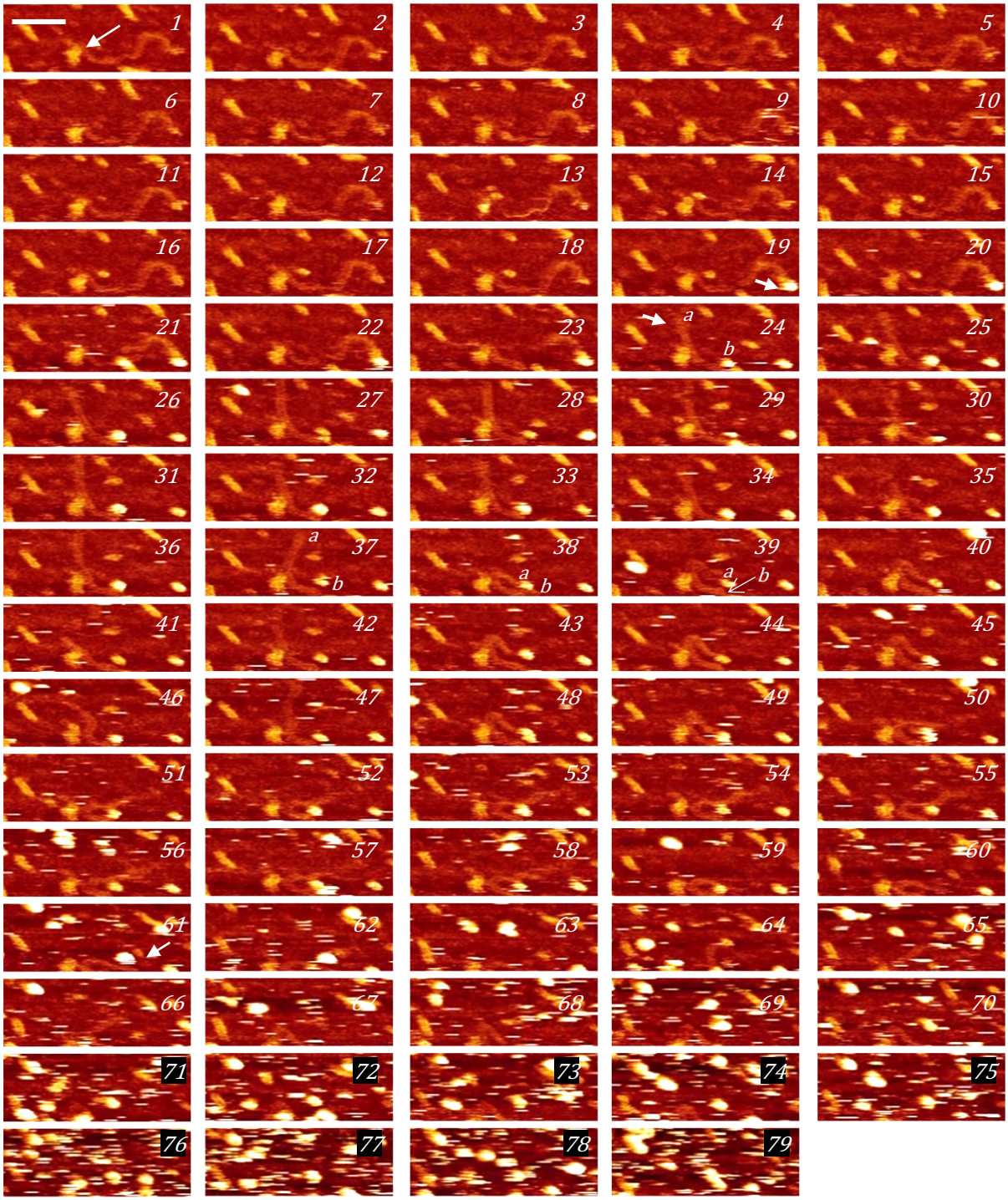


Figure 7.39: 300 bp flap DNA– T_7 (active FEN domain) imaged in buffer (25 mM HEPES, 2 mM Mg^{2+}, 5 mM K^{+}) immobilized on PLO treated mica, dynamic imaging: height 7 nm, scale bar 50 nm. The 'threading' of ssDNA could be seen along or possibly through the protein. Refer Movie P in Supplementary Material.

From the discussion above, it could be understood that the binding of protein to the branch point caused the ssDNA to bend, bind the protein, and slide along or through the protein. This corroborated the theory of Threading model of FEN binding and activity to DNA.

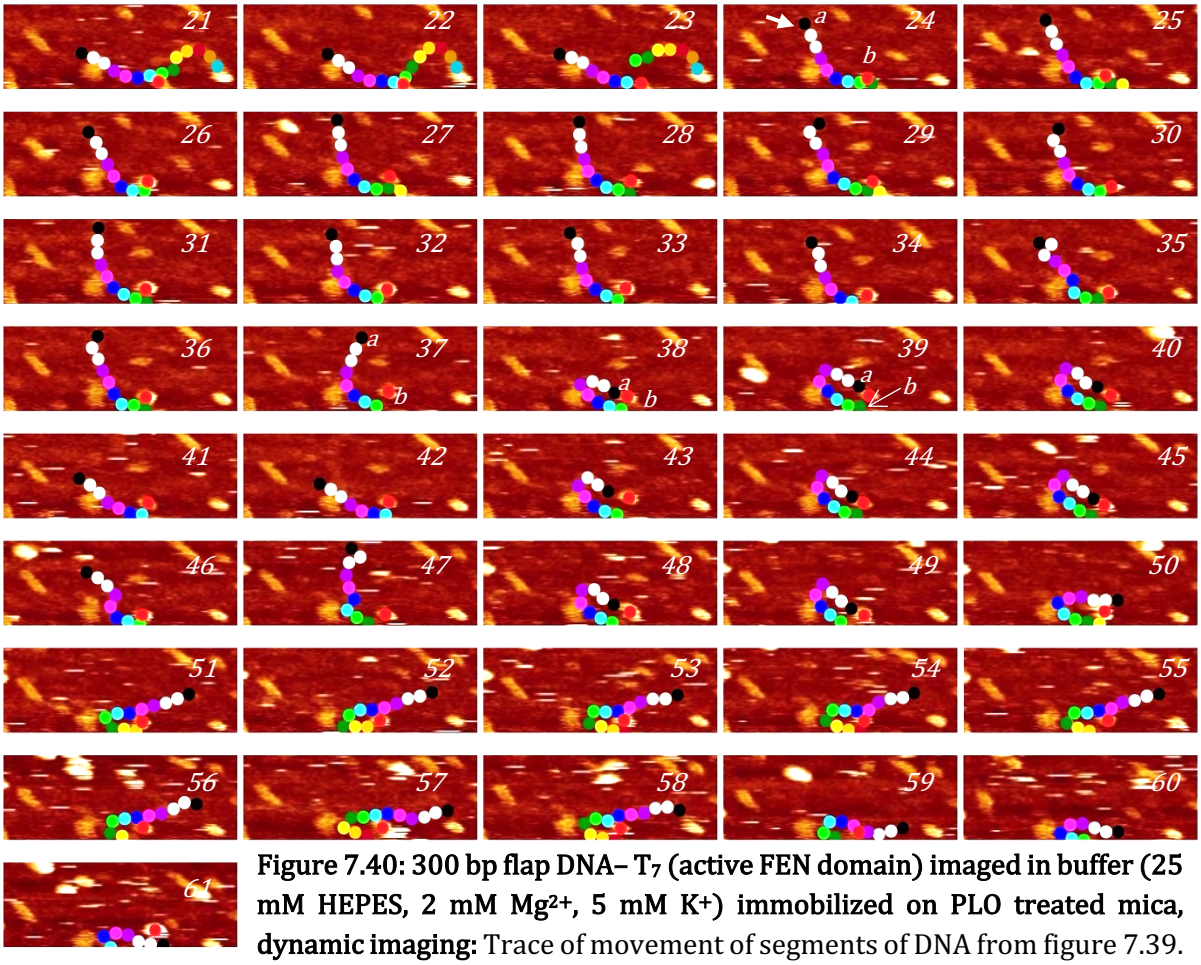


Figure 7.40: 300 bp flap DNA– T_7 (active FEN domain) imaged in buffer (25 mM HEPES, 2 mM Mg^{2+}, 5 mM K^{+}) immobilized on PLO treated mica, dynamic imaging: Trace of movement of segments of DNA from figure 7.39. The protein is marked by the red marker.

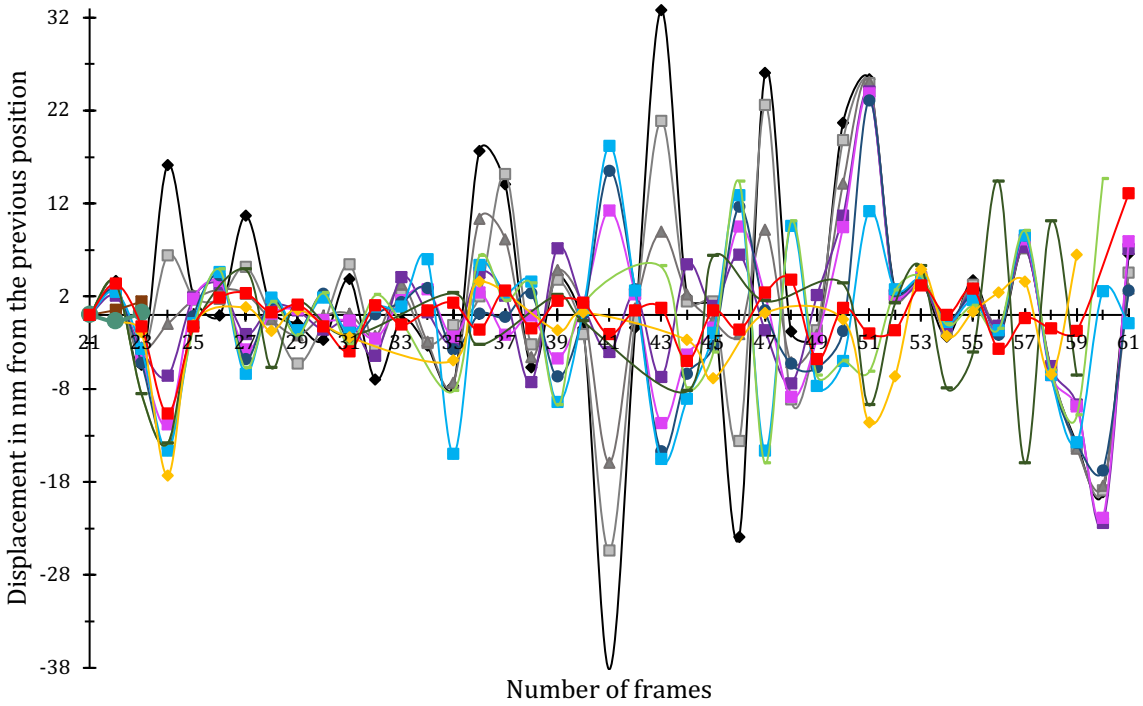


Figure 7.41: Displacement of segments of DNA in the presence of active FEN. The x-axis displays the number of frames and y-axis indicates the displacement in nm of DNA fragments of colours corresponding to figure 7.40. The red plot line indicates the motion of protein.

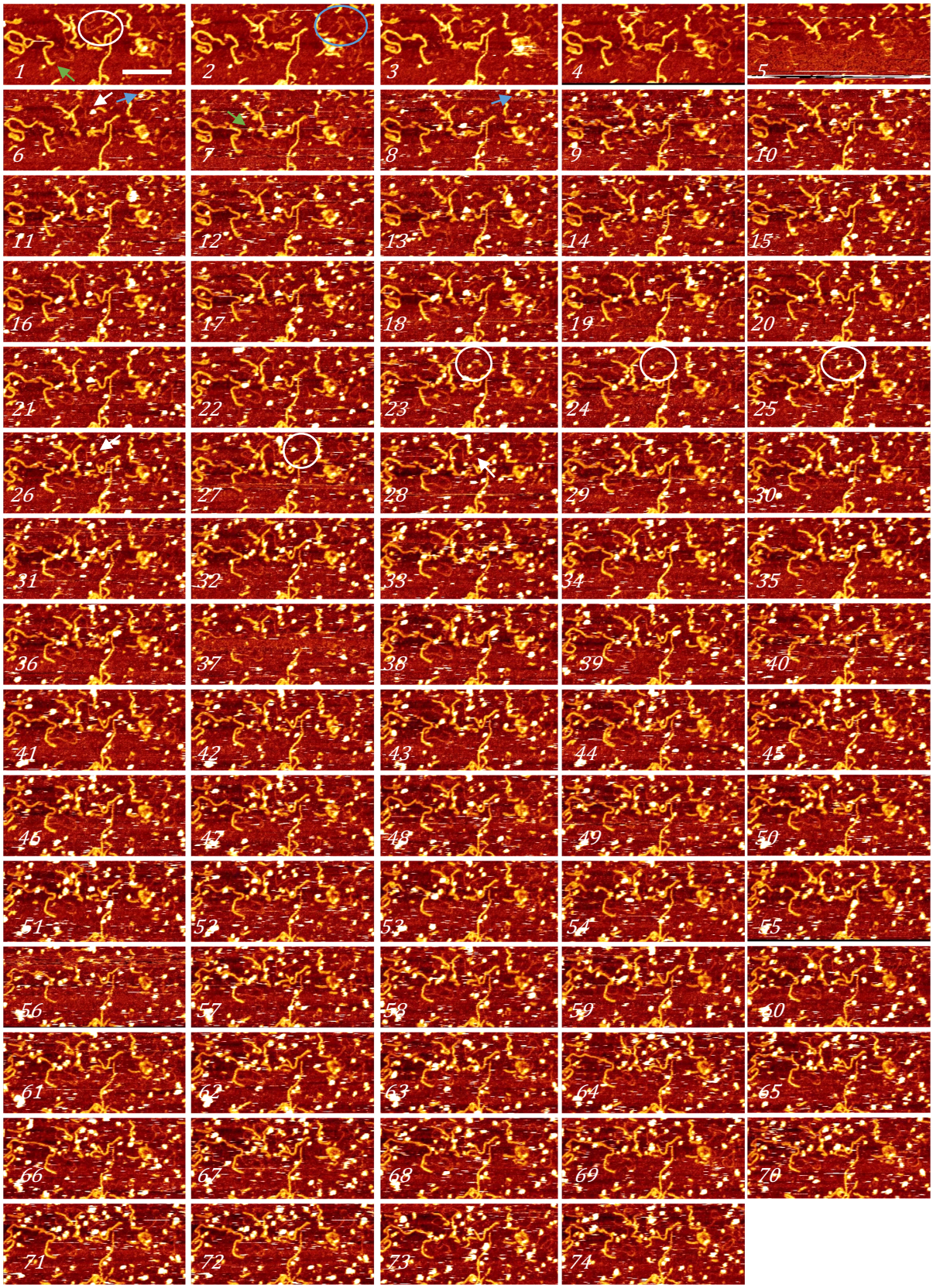


Figure 7.42: 300 bp flap DNA– T_7 (active FEN domain) imaged in buffer (25 mM HEPES, 2 mM Mg^{2+}, 5 mM K^{+}) immobilized on PLO treated mica, dynamic imaging: height 7 nm, scale bar 100 nm. The protein interacted with overhang DNA and cut off the dsDNA, causing it to move off, as seen frame 27 onwards. Refer Movie Q in Supplementary Material.

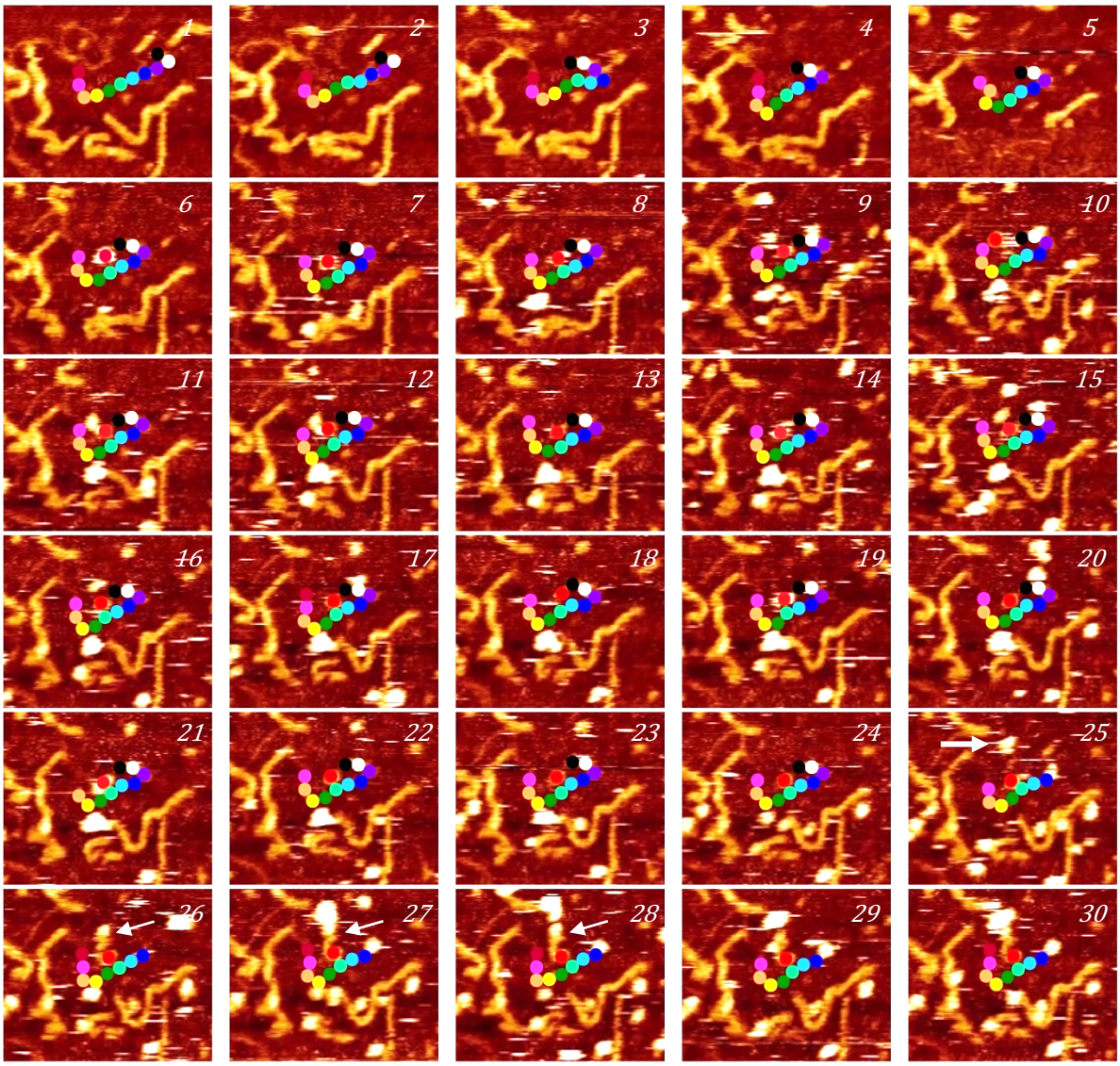


Figure 7.43: 300 bp flap DNA– T_7 (active FEN domain) imaged in buffer (25 mM HEPES, 2 mM Mg^{2+}, 5 mM K^{+}) immobilized on PLO treated mica, dynamic imaging: Trace of movement of segments of ssDNA from figure 7.42. The protein is marked by the red marker. The DNA fragment marked by markers black, white, and purple are cleaved.

In figure 7.42, another set of images where the protein had caused the ssDNA to cleave off and move the dsDNA away are shown. In frame 1, the DNA molecule encircled was an overhang DNA. This molecule was divided into segments to track the motion of the markers and plot the displacement vs frames of images. Hence, also considering the corresponding figure 7.43, the DNA could be observed moving slightly in frames 2-3 but then immobilized and remained in the same conformation till frame 25. But, from the displacement plot (figure 7.44), it could be observed that a large displacement had occurred over frame 3-4 which corresponded to the bending of a section of DNA, labelled in black, white, grey, blues and greens. Thus, even

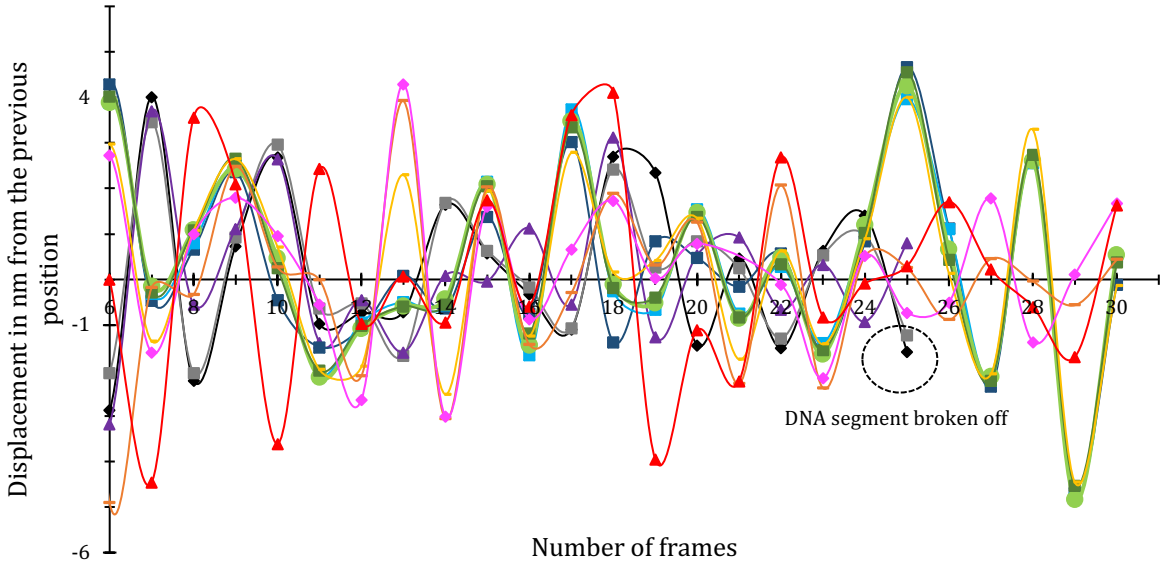


Figure 7.44: Displacement of segments of DNA in the presence of active FEN protein. The x-axis displays the number of frames and y-axis indicates the displacement in nm of DNA fragments of colours corresponding to figure 7.42. The red plot line indicates the motion of protein in figure 7.42 and 7.43. The DNA section marked by black, white, and purple markers was cleaved after frame 25 and hence show the absence of trace points.

though the motion appeared to be subtle, the statistical data emphasized the motion of the DNA molecule. There was another overhang DNA seen in frame 1 (figure 7.42), the branch point of which is shown by an arrow, but in the consecutive frames, the single strand was revealed to not be attached.

The protein solution was added during frame 5, and the molecules first appeared in frame 6, ~40-60 s later (each frame being 30 s). Interestingly, a protein molecule could be seen binding close to the ssDNA in frame 6 and remained bound to it in the following frames. This protein was labelled by red marker and traced to quantify its motion. This interaction caused the other end of DNA (marked in pink and orange) to move closer to the protein (frames 8-10 in figure 7.43) causing the DNA to bend slightly and led to the DNA end possibly binding to the protein in frame 10. The plot lines for pink and orange markers remained out-of-sync from the other marker displacement plots throughout the scan indicating that the end of the DNA was slightly mobile. But this out-of-sync behaviour of plot line for pink marker in frames 4-6 and again in frame 13 was possibly caused by movement of the DNA end under the influence of the protein present. About 10 minutes later, in frame 25, another protein (marked by arrow in frame 25 in figure 7.43) could be seen binding to the branch point/ junction of ssDNA-dsDNA and

immediately, the dsDNA (marked in black, white, and purple) was broken off from the ssDNA and moved to the position marked by the circle and arrows in frames 25-28 (figure 7.42-7.43) and remained there with two protein molecules bound to its ends. The broken ends were thus not traced further than frame 25 (figure 7.44). The ssDNA remained in the same position even after cleaving in frame 25 (figure 7.42), while the dsDNA had moved and did not immobilize in the later frames. Thus, it could be concluded that the binding and activity of the protein in frame 25 caused the dsDNA to break off the overhang DNA and move away to not immobilize on the PLO surface later.

In figure 7.42, a flap DNA molecule was also seen (encircled in frame 2, blue) but despite apparent protein binding frame 6 onwards, the DNA molecule remained intact and positioned throughout the scan period. Lastly, the overhang DNA marked by an arrow in frame 1 (green) also appeared to bind to protein in frame 7, where the protein bound the end of ssDNA, but here as well, DNA molecule remained intact and immobilized throughout the scan. Thus, these series of images could be used a good example of DNA-protein interaction as well as false positive interactions where it appeared that the protein was interacting with DNA but there was no accompanied movement or breakage of DNA.

Again, in figure 7.45, endonuclease activity of FEN could be seen. The entirety of frames 1-28 are shown in the appendix 13. The DNA molecule encircled in frame 1 is overhang DNA. The protein solution was added 8 minutes before frame 1. The protein molecule could be seen bound to the overhang DNA causing the ssDNA to move (frame 4 shows the conformation clearly). In frame 5, the overhang was cleaved and in the following frames (8-11), the ssDNA was not well immobilized indicated by the movement of the DNA strand. The dsDNA was seen in 12 with the ssDNA missing and remained immobilized from then on. There was another overhang DNA, encircled in frame 2, that appeared to have a protein bound to it and might have been an interaction, but it didn't show any change in shape or disintegration in the subsequent frames. Finally, after frame 20, the protein concentration that had immobilized on the surface was too high to observe any interactions.

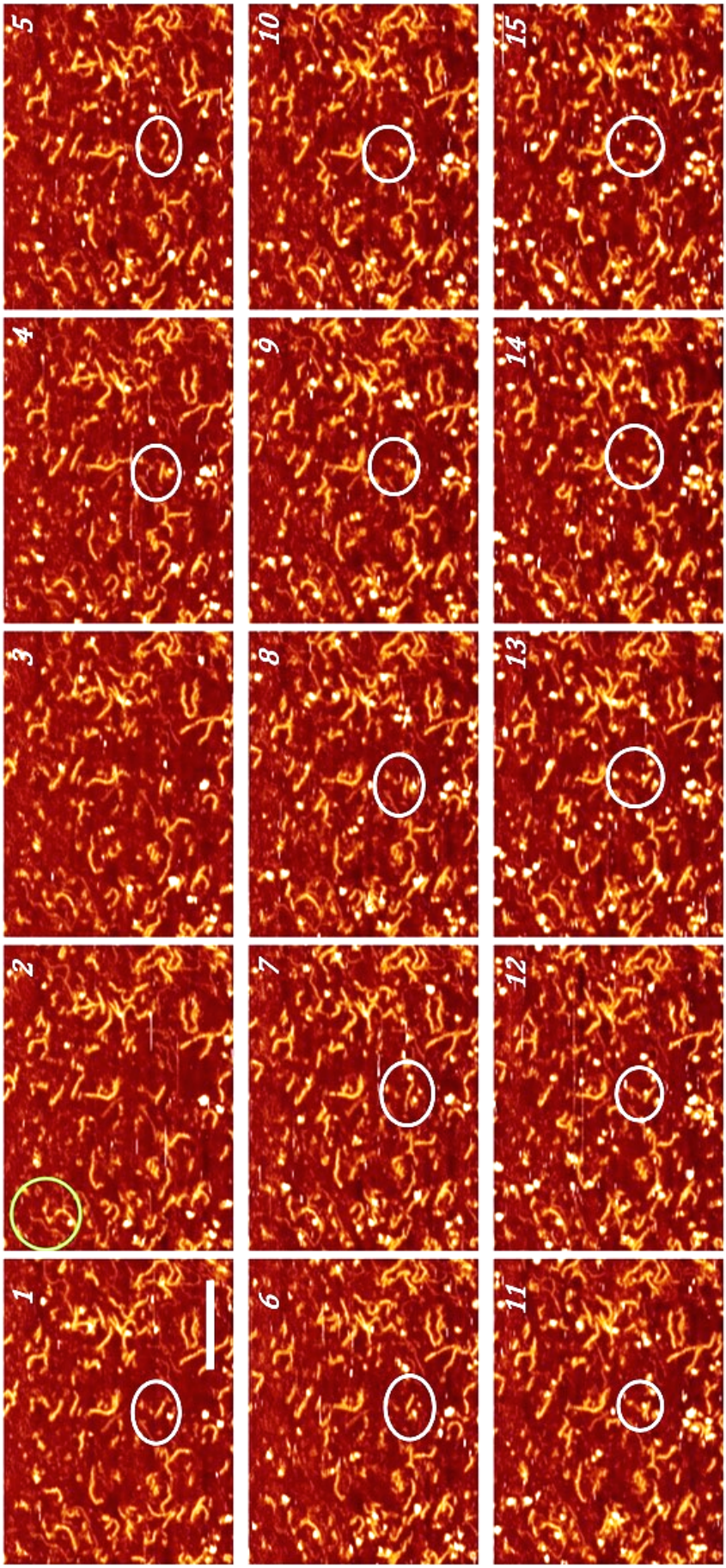


Figure 7.45: 300 bp flap DNA– T_7 (active FEN) imaged in buffer (25 mM HEPES, 2 mM Mg^{2+}, 5 mM K^+) immobilized on PLO treated mica, dynamic imaging: height 7 nm, scale bar 100 nm. These are the first 15 frames of the series of 28 images. The encircled DNA in frame 5 shows cleaving of the single strand under the influence of protein. The ssDNA broken was not seen in the next few frames, indicating that the DNA had come off the surface after breaking off. Refer Movie R in Supplementary Material.

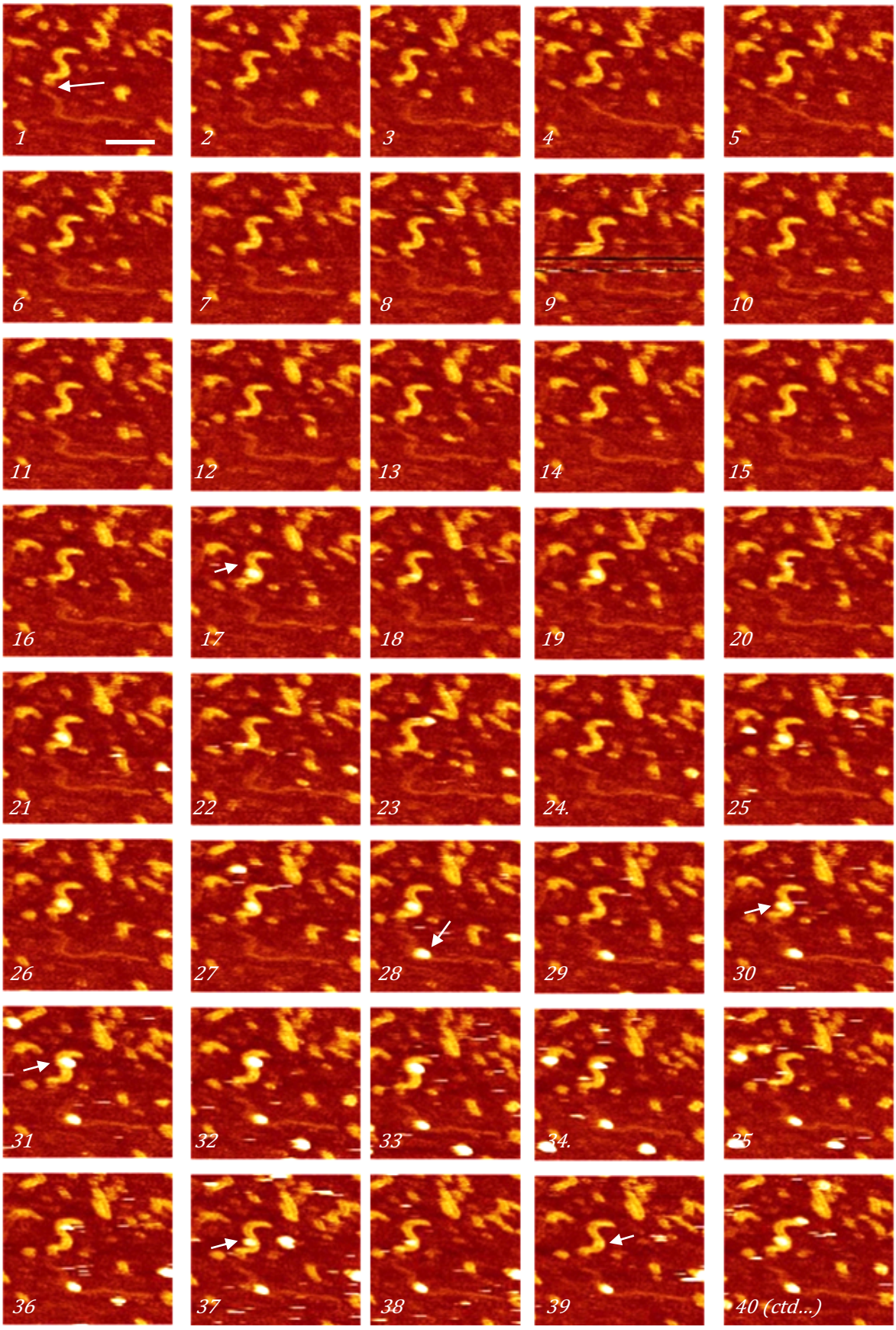


Figure 7.46: 300 bp flap DNA– T_7 (active FEN) imaged in buffer (25 mM HEPES, 2 mM Mg^{2+}, 5 mM K^{+}) immobilized on PLO treated mica, dynamic imaging: height 7 nm, scale bar 50 nm. Refer Movie S in Supplementary Material.

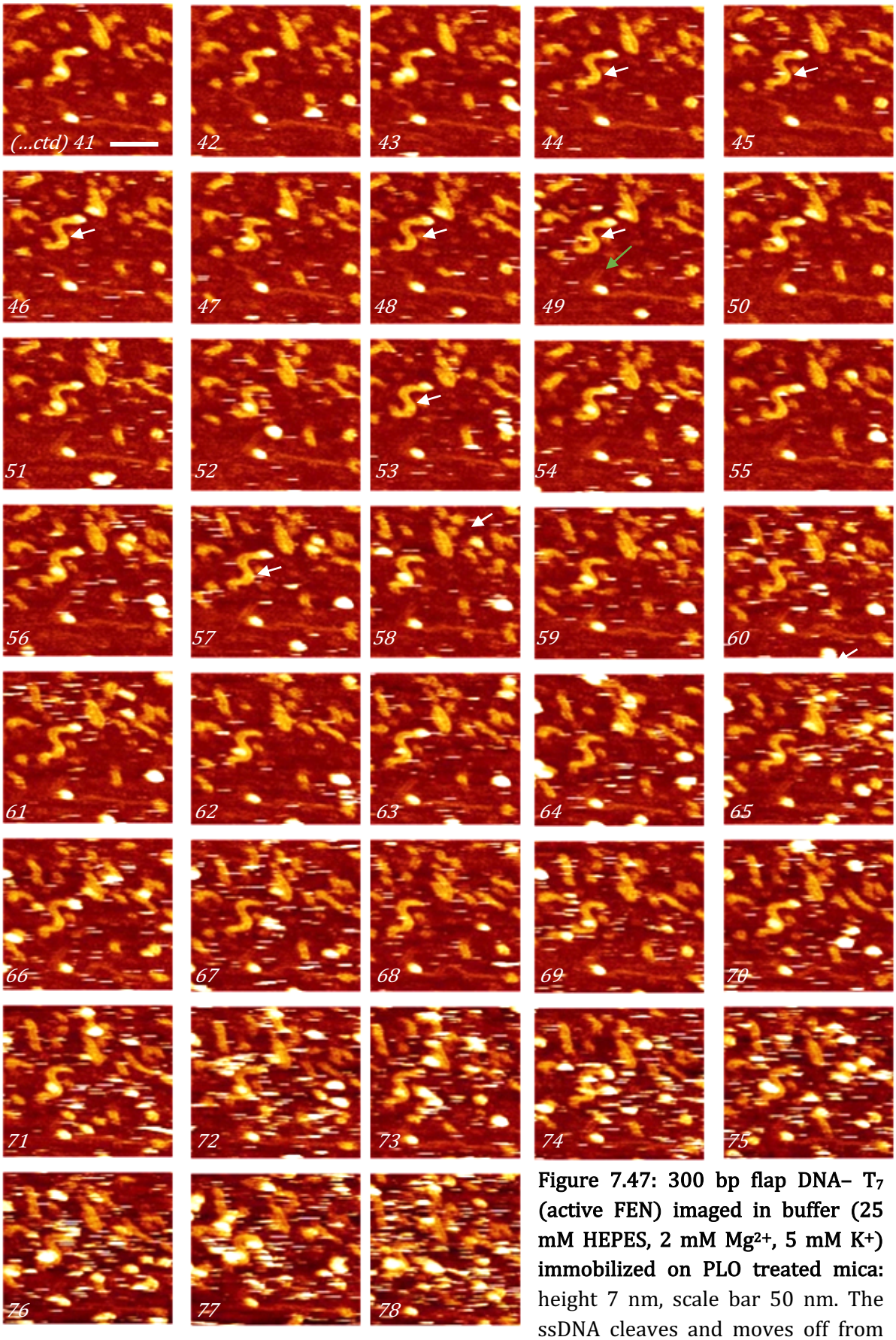


Figure 7.47: 300 bp flap DNA– T_7 (active FEN) imaged in buffer (25 mM HEPES, 2 mM Mg^{2+}, 5 mM K^{+}) immobilized on PLO treated mica: height 7 nm, scale bar 50 nm. The ssDNA cleaves and moves off from the dsDNA segment in frame 49 and remains in that position thereafter. Refer Movie S in Supplementary Material.

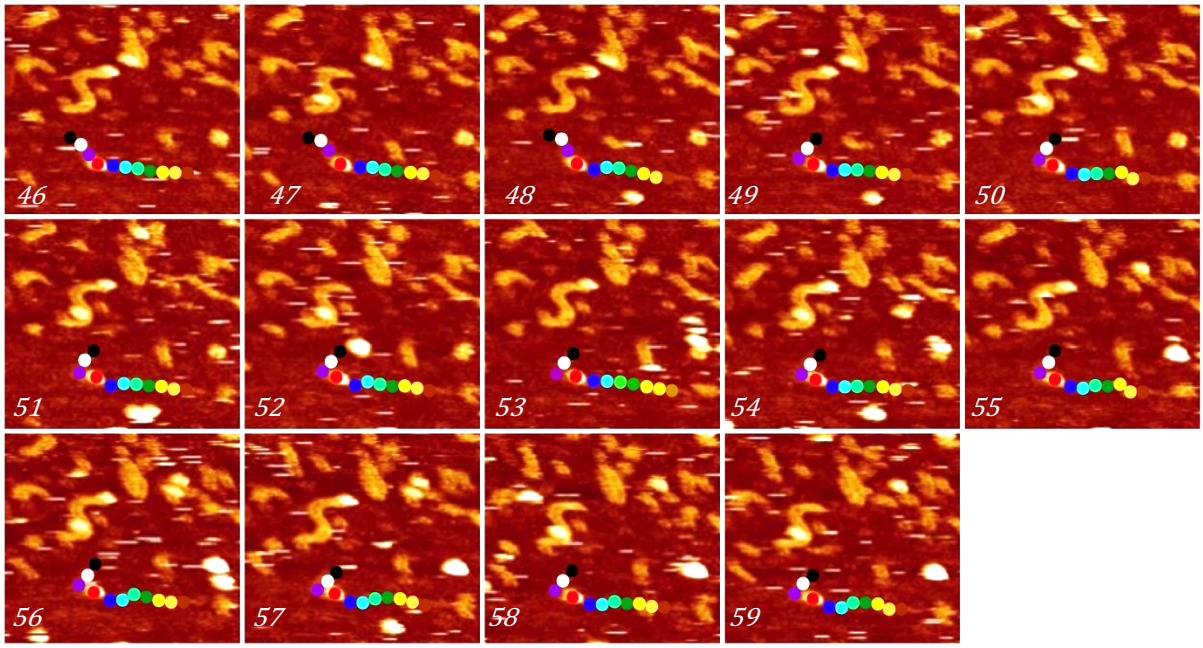


Figure 7.48: 300 bp flap DNA– T_7 (active FEN) imaged in buffer (25 mM HEPES, 2 mM Mg^{2+}, 5 mM K^{+}) immobilized on PLO treated mica, dynamic imaging: Trace of movement of segments of DNA from figure 7.47. The protein is marked by red marker. The ssDNA cleaves and moves off from the dsDNA segment in frame 49 and remains in that position thereafter.

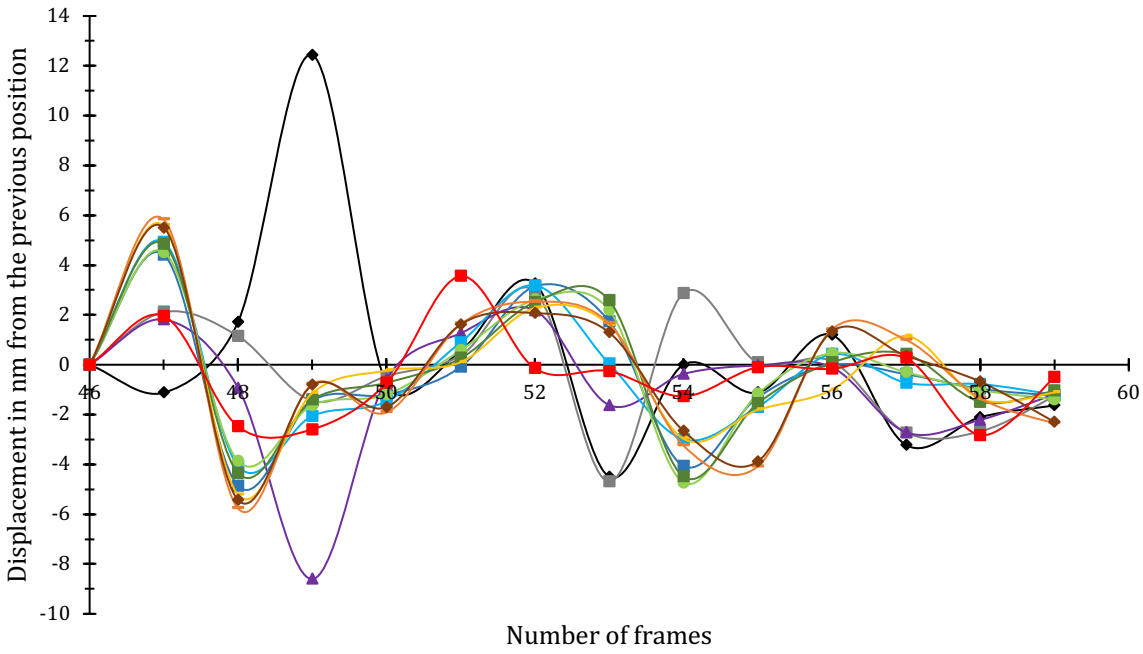


Figure 7.49: Displacement of segments of DNA in the presence of active FEN protein. The x-axis displays the number of frames and y-axis indicates the displacement in nm of DNA fragments of colours corresponding to figure 7.47. The red plot line indicates the motion of protein figures 7.47 and 7.48. The fragment of DNA marked by black, white, and purple markers is displaced in frame 49, thus showing large displacement and the other markers move more or less in synchronisation throughout the imaging.

Lastly, in figures 7.46-48, the overhang DNA could be seen being cleaved by the protein. The protein solution was added in frame 5 while the protein appeared in frame 17, 6 minutes after the protein solution was added (each frame is 30 s). The enzyme bound to the DNA molecule's double strand and remained bound to the same point until frame 30, where it appeared to slide off to a different position. From then on, the protein remained at this position till frame 37 when it moved back to the same position it was initially in frame 30-36. Interestingly, the protein molecule continued to move slightly causing it to not be seen in some frames throughout the scan (*eg.* frames 39, 44-46, 48-49). Another protein molecule bound the ssDNA in frame 28 and remained bound till the end of the images. This protein was labelled by the red marker and its motion was tracked. The shape of the DNA molecule remained consistent till 49 (green arrow), where the ssDNA appeared to move away from the dsDNA. This was probably caused by the FEN cleaving the DNA single strand at the branch point. The rest of the frames didn't show any change in the shape of the DNA molecule. However, it could not be determined if the breaking off ssDNA was the result of the protein in frame 17 bound to the dsDNA or the one marked in frame 28 bound to the ssDNA. Also, the reaction steps were not time resolved.

The DNA in frames 46-59 was marked and tracked to quantify the motion of DNA. As mentioned above, the DNA remained consistent before frame 49 (figure 7.48), where the ssDNA segment marked in black, white, and purple moved away from the dsDNA while the rest of the DNA fragment remained well immobilized. This could be corroborated by the displacement plots (figure 7.49) where the plot lines for black, white, and purple markers were out of sync from the displacement plots of the rest of the ssDNA over frames 47-50, thus indicating motion of the DNA fragment. What was very subtly observable in the figures 7.48-49 was that this DNA end continued to move slightly in the following frames, as seen in the difference in the plot lines of these same markers in frames 52-58. Also, the farther end of DNA, labelled by yellow, orange, and brown markers, also bent marginally in the frames 54-59, and this observation could be seconded by the displacement lines for these markers out-of sync in the corresponding frames.

Thus, from the above images, it was concluded that FEN could be observed interacting with overhang or flap DNA causing a change in the conformation of the latter.

7.5. Mechanism of Binding and Action of FEN on Flap/overhang DNA: A Discussion

The aim of the experiments so far had been not only to develop a method for dynamic imaging of FEN and flap DNA, but also to understand the mechanism of action of the enzyme when it encounters DNA. Work has been done to study the different kinds of substrates[11] that FEN recognises and it has been found that it binds the DNA in a sequence non-specific manner by the detection of structures different from the double helix[12]. From the current work, we have shown that FEN can recognise not only flap DNA but also single stranded overhangs, in both the cases, by possibly identifying the 3'OH at the branch point.

From previously reported research, described in section 2.2.3, the most commonly accepted theory of protein binding is the Threading model (figure 7.50) according to which FEN recognizes the branching point [13,14] and not the 5' flap end of DNA. The protein bends the DNA junction, threads 5' flap and recognises the 3'OH at the branch point [15–20]. It then encloses a single 3' nucleotide, allowing the cleaved product (resultant nicked DNA) to be ready for ligation and directs the 5' ssDNA flap through a conserved helical arch by threading mechanism [16,17,21–31]. Thus, the substrate recognition mechanism of 'fly-casting, thread, bend and barb' (figure 7.51) has been proposed [32] where the binding of the enzyme to the branch point of flap DNA aids the threading of the ssDNA through the hole in the protein to allow the ssDNA to reach the divalent cations in the active site and cause the cleavage reaction to occur.

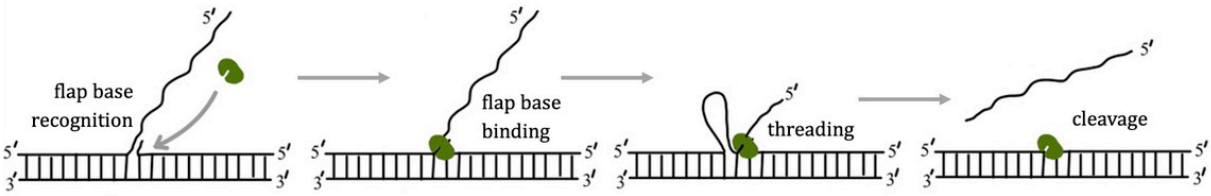


Figure 7.50: Threading model for flap DNA recognition and biding by FEN[14]: FEN binds the branch point, followed by bending and threading of the 5' end and cleaving of the ssDNA. The diagram is not drawn to scale.

In the AFM imaging experiments results reported in the current work, TaqPolI could be seen binding to DNA anywhere along the length of DNA (and not necessarily to the branch point), and cause subsequent motion of the surrounding section of DNA in the time span around protein binding. This signified that the surface charges of the DNA Polymerase I interacted with DNA causing the latter to overcome the attractive forces of PLO binding it to the mica

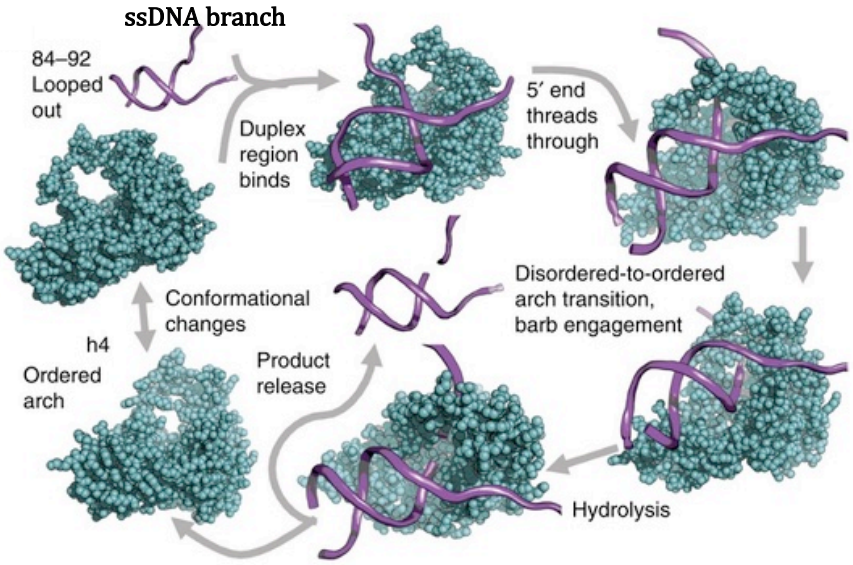


Figure 7.51: Overview of 'fly-casting, thread, bend and barb' mechanism of substrate recognition and binding by FEN [32].

surface. This resulted in the observed change in the DNA conformation as a result of protein's presence in its vicinity. However, the FEN domain alone was also capable of interacting with the entire length of DNA (ss or ds) and not just the branch point, as proposed in accordance to the threading model [13,14] of FEN recognition of substrate. The interaction of FEN with DNA caused the latter to loosen from the surface, allow the protein to bind, move (slightly or a lot, depending upon how easily the ssDNA can anchor the PLO treated mica surface) and immobilize again. In the cases where the FEN was *appearing* to bind the ssDNA end without actual binding, there was no movement of the ssDNA thereby showing no actual binding. It was noteworthy that no nuclease activity occurred when FEN was not bound close to the branch point but at farther positions on the DNA. There were brisk episodes of FEN binding directly to the branch point when the ssDNA broke off the overhang DNA.

The proposed threading of ssDNA through FEN could be observed in figures 7.39-40, where FEN bound the overhang DNA branch point and caused the ssDNA end to bend towards it and thread through (or along as it could not be resolved) the protein. This set of images supported the proposed 'fly-casting, thread, bend and barb' model of FEN activity. Interestingly, it could be seen that the ssDNA end did not bend and bind FEN immediately, but flapped around trying to bind it and became successful after multiple attempts. This showed that the protein generated some force that was attracting the ssDNA end and forced the latter to continue to attempt to bind the protein. As seen in many other images as well, the protein binding caused the single strand to bend towards the protein. This indicated that the ssDNA end was somehow attracted

to the protein with a force more than the charge forces binding the DNA to the PLO treated surface, enabling the DNA to mobilize intermittently. However, what caused the ssDNA end to be attracted to FEN from a distance to bend towards it and thread, could not be understood from these images.

Hence, though the results reported here support the Threading model of DNA substrate recognition by FEN, that the protein bound to the branch point of DNA to perform the nuclease activity, we propose the following additional steps in the recognition mechanism (figure 7.52):

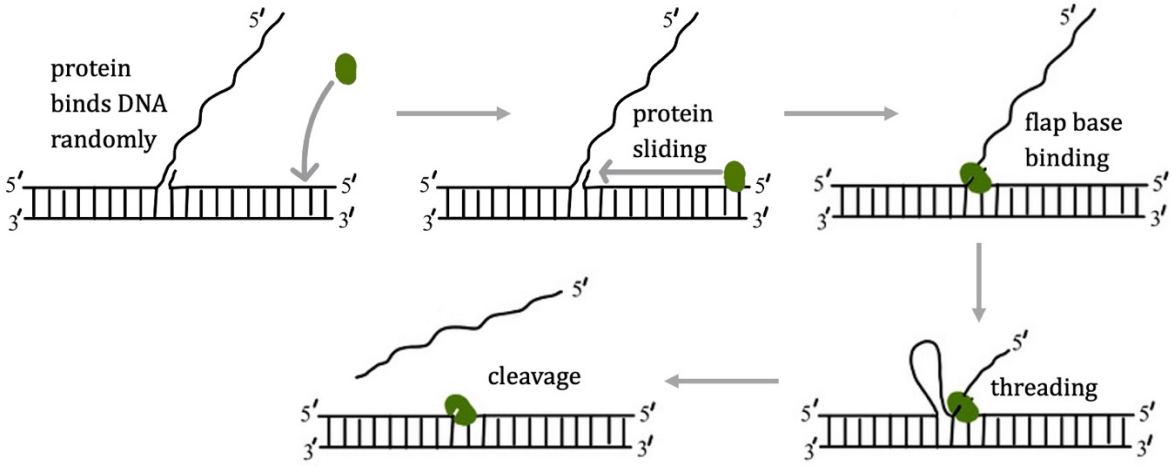


Figure 7.52: Proposed Threading model for flap DNA recognition and biding by FEN (modification of the Threading model taken from Gloor *et al* (2010)[14]): FEN or DNAPolI binds randomly to DNA, the protein slides to find the branch point, anchors to the branch point, followed by bending and threading of the 5' end and cleaving of the ssDNA. The diagram is not drawn to scale.

Firstly, the surface residues of FEN (and DNA Polymerase I, since FEN does not occur as isolated domain in the cell) are capable of attracting and binding the negatively charged DNA anywhere along its length.

Secondly, the protein is also capable of threading or sliding along the length of DNA, irrespective of the position it bound to on the DNA.

Thirdly, the sliding is a slower process than the direct recognition of the branch point and the nuclease activity of FEN.

Thereby, in order to explain how DNA Polymerase I or FEN recognizes overhang or flap DNA for nuclease activity in the cell, it could be proposed that these proteins can randomly bind DNA at any position. This binding is followed by sliding of the protein along the length of DNA, until it encounters the branch point of DNA, to which it binds and subsequently causes

the ssDNA bending and threading, as already proposed by the Threading model. Also, the nuclease activity is the faster step in the catalysis and the enzyme activity is limited by the encounter of the branch point, either randomly in the cell, or by random binding to DNA followed by sliding to reach the branch point.

7.6. Discussion: Dynamic AFM

From the current work, it could be seen that AFM can be used to image the process of FEN interacting with flap/overhang DNA. Images of 500 nm scale were obtained in a duration of 30 s per image in an environment of HEPES buffer surrounding the DNA and protein sample in the presence of Mg^{2+} and K^{+} to allow FEN activity and DTT to stabilize the protein. To analyse the motion of DNA due to interaction with protein, a method of video analysis was developed based on motion tracking. In each of the frames, the DNA was segmented into differently coloured labels and tracked using the software 'Tracker: Video Analysis and Modelling Tool'. Displacement-frame plots were used to understand the motion of sections of DNA due to the vicinity and timing of protein binding.

It was observed that the inactive protein bound to DNA causing it to mobilize from the surface, move and anchor to the surface again. Most of the motion of DNA occurred in locations that were close to the site where the protein was bound, indicating that the protein caused an attractive force that was stronger than the charge interactions binding the DNA to the surface. The further away sections were strongly anchored to the surface and did not show any movement. Also, the motion of DNA was restricted to very few frames (2-4) or ~ 2 minutes around the protein's presence in the vicinity of the DNA. The DNA then anchored to the surface again after a few frames of protein interaction, indicating that the effect of the protein causing DNA mobility had been removed. Since the time and spatial resolution of AFM is limited, the conformation details in those 2 minutes could not be further observed.

The active protein caused the DNA to not only move, but also be cleaved off and thread through FEN. This interaction assured that the sample immobilization for AFM was adequate for the DNA to be bound strongly enough to be detected by the cantilever yet loose enough to interact with the protein.

Many images before and after the introduction of protein into the DNA solution while imaging showed that the force of imaging was sufficient to not damage the sample and that the

cantilever tip was not interfering with the ability of the proteins to interact with the DNA. There was a limitation to the duration for which imaging could be continued, not because of the tip causing damage to the sample due to continuous tapping but owing to the crowding of the proteins on the sample surface inhibiting the view of any possible conformation changes. This crowding also contributed to the noise in the later images as the tip became sticky. Interestingly, in these instances, the cantilever could be gently washed with detergent, isopropanol and HPLC water and reused for at-least one more series of interaction images.

Nevertheless, there are some limitations to the current set of experiments. Firstly, the DNA and protein samples interact in 3-D space and the observation of the reaction on AFM required the transfer to 2-D space because the sample is immobilized on the surface. It therefore becomes imperative to assume that the reaction is not exactly mimicking the way it happens in the cell. The speed of sliding of the protein on DNA observed here using AFM would be slower than the actual sliding in the 3-D environment of the cell. There was also the obvious limitation to the movement of FEN due to the attractive forces of the PLO treated mica and the subsequent immobilization of the protein on the surface.

Secondly, the spatial resolution of AFM in general is not good enough to see the various motifs of the protein or the individual nucleotides of the DNA strand, which is important to understand the conformational changes during DNA-protein interaction. Thirdly, the time resolution of dynamic imaging needs to be improved to observe fast reactions of the cell, especially the ones during DNA replication and protein expression. And lastly, flap or overhang DNA sample is difficult to observe by an AFM owing to the presence of two different kinds of DNA-single strands and double strands present together. Improvements in the technique, especially in the aspects of time and spatial resolution would prove to be highly beneficial for observing dynamic processes.

7.7. Conclusion

In the current work, we have proposed that the FEN or DNA Polymerase I can bind anywhere on the flap or overhang DNA substrate, not just at the branch point, as reported by previous research works. With the aid of the AFM images, we have shown that the protein is able to move along/slide on the DNA which allows it to ‘search’ the DNA for branch points or 3' OH that it can recognise and anchor, and subsequently begin the catalytic step of nuclease activity. Due to the charged surface residues of the protein, the latter is able to attract the ends of DNA

such that they overcome the electrostatic forces of attraction binding the DNA to the PLO treated mica surface and become mobile in the vicinity of the protein, to thread through it and consequently be cleaved in the active site of FEN.

High resolution AFM could be used to image protein molecules to understand their different conformations. A method of AFM image analysis pertaining to DNA motion tracking has been developed using Tracker: Video Analysis and Modelling Tool, that has allowed quantification of the changes in conformation of DNA due to the interaction with the protein. Further experimentation and developments in AFM are required to obtain resolution such that the details of protein structure and possibly the various motifs and domains could be seen. Future experiments could also be done with purer DNA and protein sample to reduce the noise. Also, developments in AFM to obtain better spatial and time resolution, especially with soft biological samples could greatly assist the understanding of biochemical processes.

Bibliography

1. Kim Y, Eom SHSH, Wang J, Lee D-SDS, Suh SWSW, Steitz TATA. Structure of Taq Polymerase. Nature. http://www.rcsb.org/structure/1TAQ. Published August 1995.
2. Garforth SJ, Patel D, Feng M, Sayers JR. Unusually wide co-factor tolerance in a metalloenzyme; divalent metal ions modulate endo-exonuclease activity in T5 exonuclease. Nucleic Acids Res. 2001;29(13):2772-2779.
3. Yang W, Lee JY, Nowotny M. Making and Breaking Nucleic Acids: Two-Mg2+-Ion Catalysis and Substrate Specificity. Mol Cell. 2006;22(1):5-13.
4. Tock MR, Frary E, Sayers JR, Grasby JA. Dynamic evidence for metal ion catalysis in the reaction mediated by a flap endonuclease. EMBO J. 2003:995-1004.
5. Feng M *et al*. Roles of divalent metal ions in flap endonuclease-substrate interactions. Nat Struct Mol Biol. 2004;11:450-456.
6. Tomlinson G *et al*. Neutralizing Mutations of Carboxylates That Bind Metal 2 in T5 Flap Endonuclease Result in an Enzyme That Still Requires Two Metal Ions. J Biol Chem. 2011;286:30878-30887.
7. Syson K *et al*. Three metal ions participate in the reaction catalyzed by T5 flap endonuclease. J Biol Chem. 2008;283:28741–28746.
8. Sorel I *et al*. The EcoRI-DNA complex as a model for investigating protein-DNA interactions by atomic force microscopy. Biochemistry. 2006;45(49):14675-14682.
9. Klenow H, Henningsen I. Effect of Monovalent Cations on the Activity of the DNA Polymerase of Escherichia coli B. Eur J Biochem. 1969;9(1):133-141.

10. Nagai M, Yoshida A, Sato N. Additive Effects of Bovine Serum Albumin, Dithiothreitol, and Glycerol on PCR. Biol Int. 1998;44(1):157-163.

11. Nazarkina ZK, Lavrik OI, Khodyreva SN. Flap endonuclease 1 and its role in eukaryotic DNA metabolism. Mol Biol. 2008;42(3):357-370.

12. Beddows A, Patel N, Finger D, Atack JM, Williams DM, Grasby JA. Interstrand disulfide crosslinking of DNA bases supports a double nucleotide unpairing mechanism for flap endonucleases. Chem Commun. 2012;48:8895–8897.

13. Xu Y, Potapova O, Leschziner AE, Grindley NDF, Joyce CM. Contacts between the 5' Nuclease of DNA Polymerase I and Its DNA Substrate. J Biol Chem. 2001;276:30167-30177.

14. Gloor JW, Balakrishnan L, Bambara RA. Flap Endonuclease 1 Mechanism Analysis Indicates Flap Base Binding Prior to Threading. J Biol Chem. 2010;285(45):34922-34931.

15. Chapados BR *et al*. Structural Basis for FEN-1 Substrate Specificity and PCNA-Mediated Activation in DNA Replication and Repair. Cell. 2004;116(1):39-50.

16. Tsutakawa SEE *et al*. Human Flap Endonuclease Structures, DNA Double-Base Flipping, and a Unified Understanding of the FEN1 Superfamily. Cell. 2011;145(2):198-211.

17. Blanchard MS *et al*. The 3'-Flap Pocket of Human Flap Endonuclease 1 Is Critical for Substrate Binding and Catalysis. J Biol Chem. 2009;284:22184-22194.

18. Tsutakawa SE *et al*. Phosphate steering by Flap Endonuclease 1 promotes 5 0-flap specificity and incision to prevent genome instability. Nat Commun. 2017;8:15855.

19. Rashid F *et al*. Single-molecule FRET unveils induced-fit mechanism for substrate selectivity in flap endonuclease 1. Elife. 2017;6:1-23.

20. Algasaier SI *et al*. DNA and Protein Requirements for Substrate Conformational Changes Necessary for Human Flap Endonuclease-1-catalyzed Reaction. J Biol Chem. 2016;291(15):8258-8268.

21. Devos JM, Tomanicek SJ, Jones CE, Nossal NG, Mueser TC. Crystal Structure of Bacteriophage T4 5 Nuclease in Complex with a Branched DNA Reveals How Flap Endonuclease-1 Family Nucleases Bind Their Substrates. J Biol Chem. 2007;282(43):31713-31724.

22. Bornarth CJ, Ranalli TA, Henricksen LA, Wahl AF, Bambara RA. Effect of Flap Modifications on Human FEN1 Cleavage. Biochemistry. 1999;38:13347-13354.

23. Murante RS, Rust L, Bambara RA. Calf 5' to 3' Exo/Endonuclease Must Slide from a 5' End of the Substrate to Perform Structure-specific Cleavage. J Biol Chem. 1995;270:30377-30383.

24. Craggs TD, Hutton RD, Brenlla A, White MF, Penedo JC. Single-molecule characterization of Fen1 and Fen1/PCNA complexes acting on flap substrates. Nucleic Acids Res. 2014;42:1857-1872.

25. Ceska TA, Sayers JR. Structure-specific DNA cleavage by 5′ nucleases. Trends Biochem Sci. 1998;23(9):331-336.

26. Tsutakawa SE, Tainer JA. Double strand binding–single strand incision mechanism for human flap endonuclease: Implications for the superfamily. Mech Ageing Dev. 2012;133(4):195-202.

27. Ceska TA, Sayers JR, Stier G, Suck D. A helical arch allowing single-stranded DNA to thread through T5 5'-exonuclease. Nature. 1996;382(6586):90-93.

28. Shaw SJ, Finger LD, Grasby JA. Human Exonuclease 1 Threads 5′-Flap Substrates through Its Helical Arch. Biochemistry. 2017.

29. Liu R, Qiu J, David Finger L, Zheng L, Shen B. The DNA-protein interaction modes of FEN-1 with gap substrates and their implication in preventing duplication mutations.

30. Qiu J, Liu R, Chapados BR, Sherman M, Tainer JA, Shen B. Interaction Interface of Human Flap Endonuclease-1 with Its DNA Substrates S. J Biol Chem. 2004;279(23):24394-24402.

31. Henricksen LA, Tom S, Liu Y, Bambara RA. Inhibition of Flap Endonuclease 1 by Flap Secondary Structure and Relevance to Repeat Sequence Expansion. J Biol Chem. 2000;272:16420–16427.

32. AlMalki FA *et al*. Direct observation of DNA threading in flap endonuclease complexes. Nat Struct Mol Biol. 2016.

Chapter 8

Imaging *parS* DNA-ParB protein Interactions

8.1. Introduction

In the previous chapters, the objective of the experiments was to observe single molecule images of DNA-protein interactions of flap endonuclease with flap or overhang DNA. The long-term purpose of the experimentation, besides understanding the mechanism of action of FEN, was to establish the methods for dynamic DNA-protein imaging using AFM that could be expanded to numerous protein-nucleic acid interactions in general. A step in this direction is the current chapter, based on the imaging of ParB protein interaction with DNA molecules containing *parS* protein-binding sites (PBS).

As already described in chapter 2, ParB binds *parS* sites on the DNA as an imperative step in the process of plasmid partitioning (separation of the duplicated bacterial genome into the daughter cells). Various molecular interactions, conformation changes and enzymatic activity occurs during ParB-*parS* DNA recognition and binding. Despite a lot of research and years of experimentation, there are many gaps in the understanding of the details of the process. Recently, the involvement of CTP in the binding of ParB to DNA has been discovered [1–3]. To understand this, experiments have been performed to analyse the change in behaviour of ParB in accordance with varying concentrations of CTP in the buffer. Additionally, the change in ParB binding specifically to *parS* site vs to DNA in a sequence non-specific manner has also been found to be affected by the presence of CTP. To understand the above reactions, high-resolution AFM has been used to observe the protein and DNA conformations during their interactions and the effect of CTP on the process.

8.2. Imaging single parS site containing Plasmid

The first step in the experimentation of protein-DNA interaction was obtaining DNA control images.

Plasmid DNA (~3 kb) containing *parS* site was immobilized on PLO treated mica and imaged in air, water and buffer environment (as described in section 4.2.8.1,3 and 4) (figures 8.1-3). The images revealed the DNA plasmid to be convoluted and twisted into various shapes, and multiple points were visible where the strands coiled. Figure 8.1 shows some of the DNA shapes observed when imaged in air. The DNA molecules, though closed, were seen to be broken at many places on the plasmid (marked by green arrows) while some DNA molecules remained intact (figure 8.1 e-f). This could have occurred because of the DNA breaking during immobilization steps or the extraction and purification experiments. In almost all the case, the DNA strands appeared to cluster at one or more points on the molecule (marked by pink arrows), giving the appearance that some section of the DNA is binding/linking to other strands. Interestingly, these kinds of images should be expected when imaging plasmid during interaction with ParB. The expectations in this case would be plasmid DNA having more knots while interacting with ParB than without. Hence, it was worth comparing these negative control images with the ones of DNA with the protein in the forthcoming sections.

In some sections of DNA molecules (figure 8.1 e and f, marked by brackets), the double helix could be faintly resolved, while in some molecules, the strands were supercoiled (figure 8.1 b, in brackets and c, whole molecule). Also, there were many sections of the molecules that were relaxed and spread out, like in figure 8.1 e and f. The significance of these conformations was that it would be interesting to compare these molecular arrangements in the presence of protein.

When a slightly longer plasmid sample (exact length unknown) was imaged, a beadiness on the DNA could be seen in air (figure 8.2 a-d and f), possibly indicating the helical structure of dsDNA. In water (figure 8.2 e,g-i), however, the DNA resolution was compromised owing to the slight movement of DNA causing noise in the images. Here as well, the DNA was broken for some molecules (marked by green arrow in figure 8.2 f) but for most, they remained intact. Almost all of the molecules were coiled together and there were very few instances where some sections were relaxed into spread-out shapes (figure 8.2f). The molecular shapes were similar while imaging in HEPES buffer (figure 8.3) but the images were better resolved. Incidentally, in this DNA sample, unlike the slightly shorter one used before (figure 8.1), the molecules were

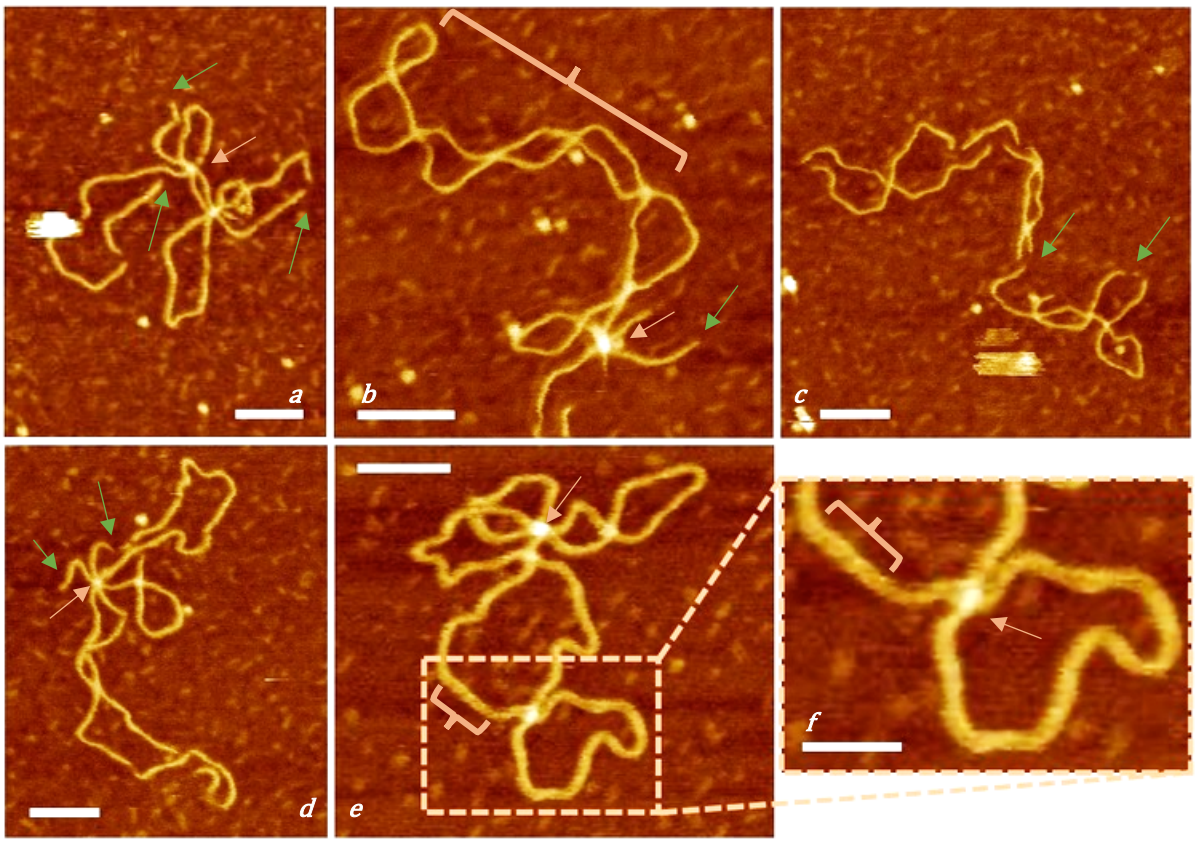


Figure 8.1 : *parS* DNA plasmid (~3 kb) in air: a-e: height 6 nm, scale bar 50 nm, f. scale bar 30 nm, height 6 nm.

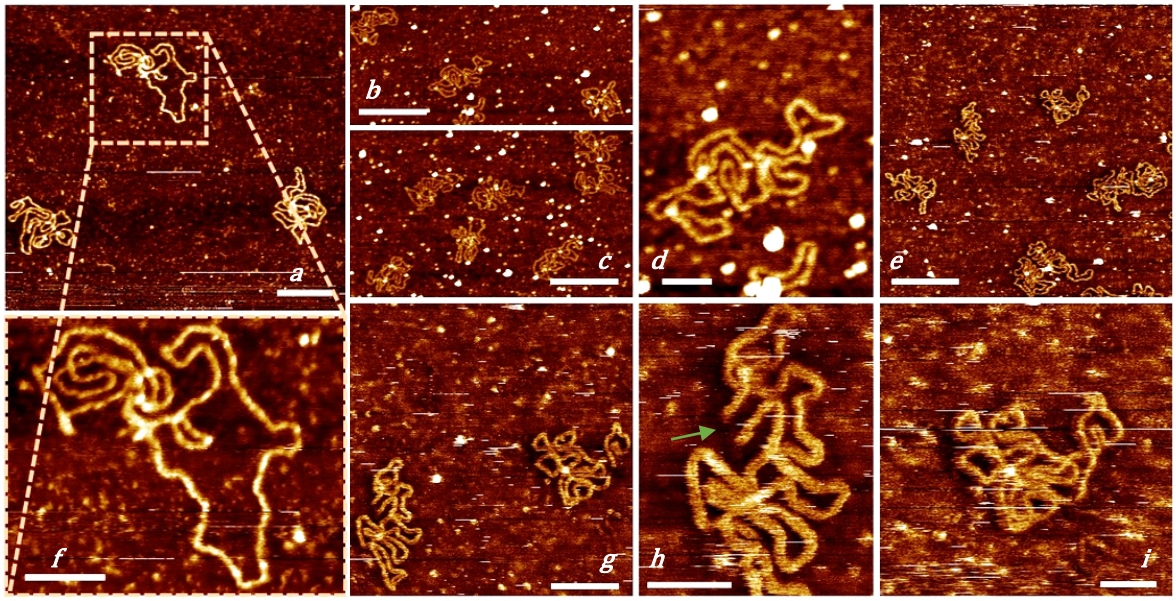


Figure 8.2: *parS* DNA plasmid (longer than 3 kb plasmid in figure 8.1)
in air: a. height 1.4 nm, scale bar 100 nm, b and c. height 2.5 nm, scale bar 200 nm, d. height 1.8 nm, scale bar 50 nm, f. height 1.7 nm, scale bar 40 nm,
in water: e. height 3.1 nm, scale bar 200 nm, g. height 3.3 nm, scale bar 100 nm, h. height 3.5 nm, scale bar 50 nm, i. height 3.3 nm, scale bar 50 nm.

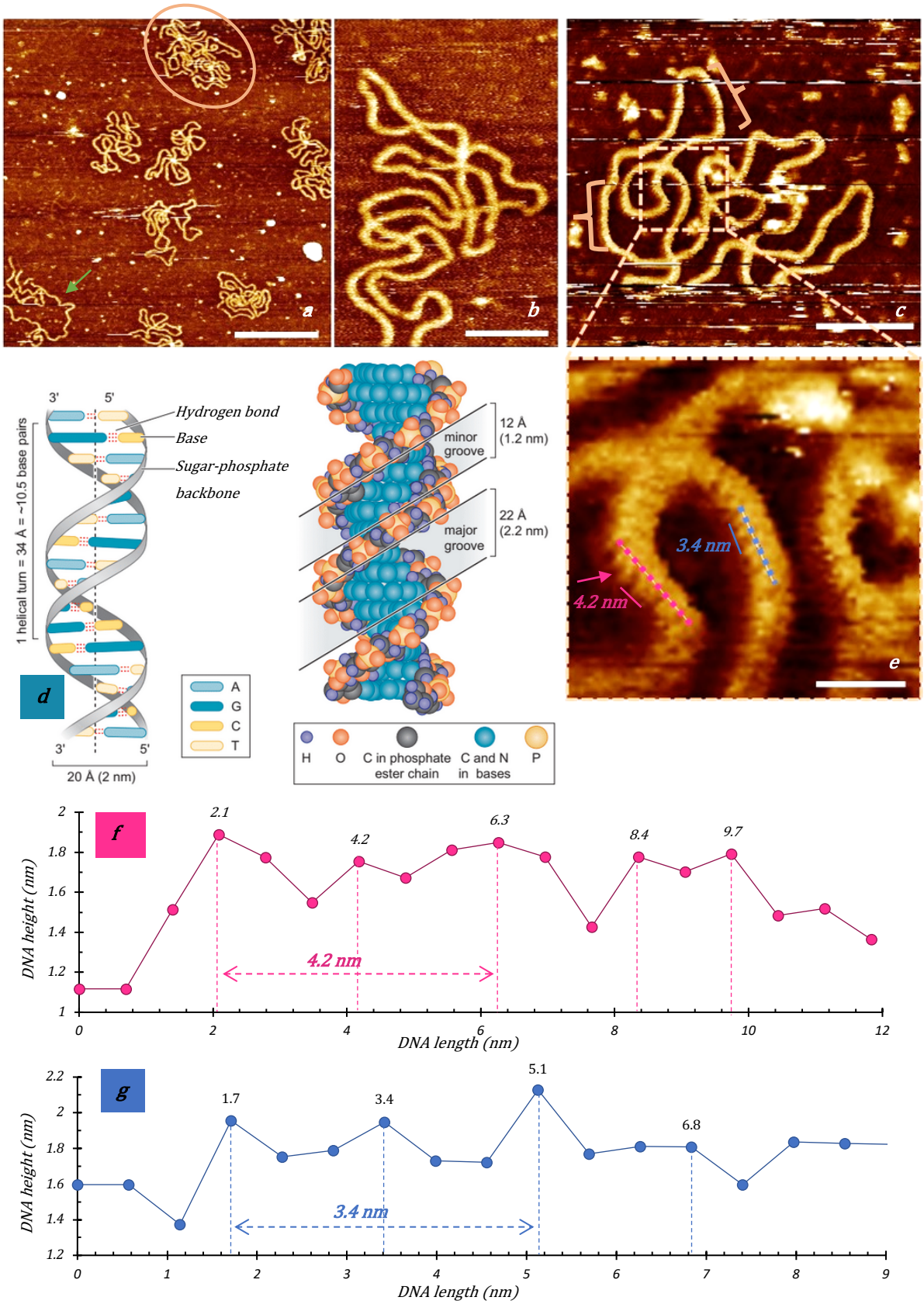


Figure 8.3: *parS* DNA plasmid (~ 3 kb) in HEPES buffer (pH 6.5-7): a. height 5 nm, scale bar 200 nm, b. height 5 nm, scale bar 50 nm, c. height 4.8 nm, scale bar 50 nm, e. height 5.8 nm, scale bar 10 nm, **d. schematic model of DNA double-helix,** reproduced from 'Molecular Biology of the Gene" by Watson et. al [4]. **f and g. cross section of the DNA double helix seen in e.** In f, the

peaks can be seen differing by 2.1 nm while in g, the peaks are differing by 1.7 nm. These peaks represent the sugar-phosphate backbone which are 1.2 nm apart at the minor groove and 2.2 nm apart at the major groove. The gap between major each alternate peak corresponds to 1 helical turn, measured about 4.2 nm and 3.4 nm respectively for the 2 cross-sections.

not only coiled more, but sometimes few molecules intertwined to form a group from which individual molecules could not be made out (encircled in 8.3 a). The coiling of the molecules would make it impossible to see any possible interaction with the protein in the further experiments as there were scarcely any molecules with a relaxed conformation (like the one marked by green arrow in figure 8.3 a). These convolutions were similar, but much less complex, than the ones observed in case of dsDNA formed from M13 ssDNA , discussed in chapter 6 (section 6.3), where the length of DNA was even longer and thus the shape of immobilized DNA more intricate.

Owing to the better resolution of the images in buffer, the double helix of DNA could be seen. In figure 8.3 c and e, the major and minor grooves of some sections of DNA could be seen clearly (marked by trace of lines). The cross-section of the strand revealed repetition of the peaks at consistent intervals. For example, in one of the cross sections (figure 8.3 f), the peaks appear every 2.1 nm, corresponding to the sugar-phosphate backbone of the DNA double helix (shown as chain of orange molecules in figure 8.3 d[4]). The pitch of the double helix would be the difference between the consecutive major grooves, or consecutive minor grooves or every alternate sugar-phosphate backbone, which in this case was observed to be 4.2 nm. It could be seen from the image that the DNA was slightly elongated in this section as compared to the rest of the molecule (figure 8.3 c) as the strand was bending here. Hence, this could be regarded as the possible cause of the slightly higher pitch (4.2 nm) observed as compared to the theoretical 3.4 nm expected from the crystallography data for DNA double helix [4,5]. Another section of DNA was cross-sectioned (figure 8.3 g) and revealed similar recurrence of peaks corresponding to the sugar-phosphate backbone. But here the pitch was measured to be 3.4 nm. These variations in the pitch lengths were probably due to the bending of DNA molecule at various points causing the helix to be slightly relaxed at the outside of the bends (8.3 f) and slightly constrained on the inside of the bends (8.3 g). Nevertheless, it was interesting to observe the minor groove (arrow in 8.3 e) being resolved on the images which, contrary to the major groove, is difficult to resolve using AFM.

8.3. Methods to Spread Plasmid DNA during Immobilization

8.3.1. Immobilization of parS DNA

To overcome the problem of convolution of large sized DNA plasmids, a method of immobilization focussing on gentle transfer of DNA from the 3-dimensional environment in solution to a 2-dimensional surface was adopted [6] (figure 8.4) (see section 4.2.8.8). Here, the DNA was diluted in a HEPES buffer containing Mg^{2+} and K^{+}, added onto Ni^{2+} -treated mica surface and gently washed with a large volume of the same buffer so as to avoid harshly squishing the DNA onto the flat surface. This ensured that the molecules did not clump but were allowed to spread. The sample was then washed with Ni^{2+} containing buffer, *aka* imaging buffer (in place of Mg^{2+}) to 'seal' the bound DNA as the Ni^{2+} enhance the binding of DNA to the mica surface more than Mg^{2+}. It is worth emphasizing that this method was proposed by the authors for DNA varying in length between 200 and 2 kB, while the DNA sample in the current work was > 3 kb. Yet attempts were made to experiment with this method of immobilization.

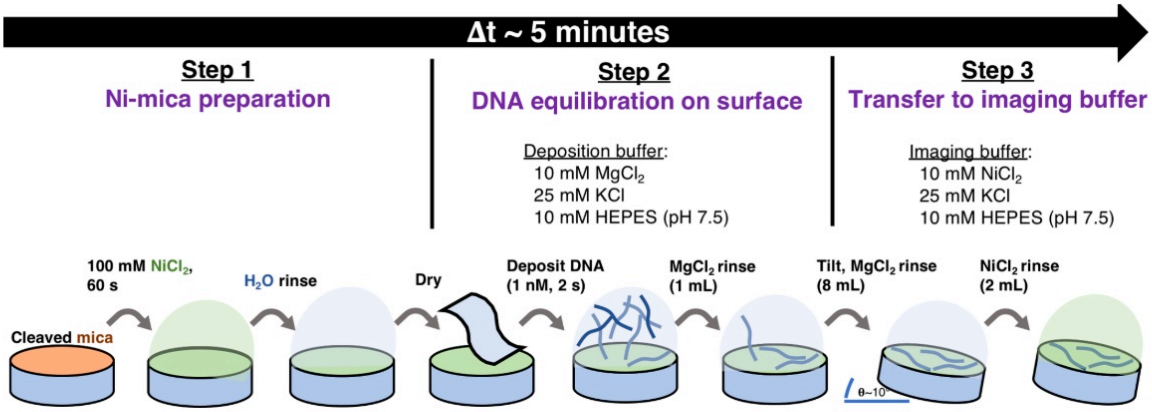


Figure 8.4: Schematic of sample preparation method for AFM by Heenan and Perkins (2019)[6].

Figure 8.5 a shows DNA molecules immobilized by a slight modification of the Heenan-Perkins' method. DNA in ParB reaction buffer (50 mM TRIS (pH 7.5), 100 mM NaCl and 5 mM $MgCl_2$) was added on PLO treated mica and washed with deposition buffer (HEPES, Mg^{2+} and K^{+}) followed by washing with HEPES buffer and imaged in HEPES buffer without the presence of salts. The images revealed most of the DNA molecules quite clumped together, even if the protein was not present in the solution. Interestingly, many molecules interlinked to form complex meshes (*e.g.* the few ones marked by blue arrows) of various density and could not be distinguished into individual molecules. Such complex meshwork in large amount had

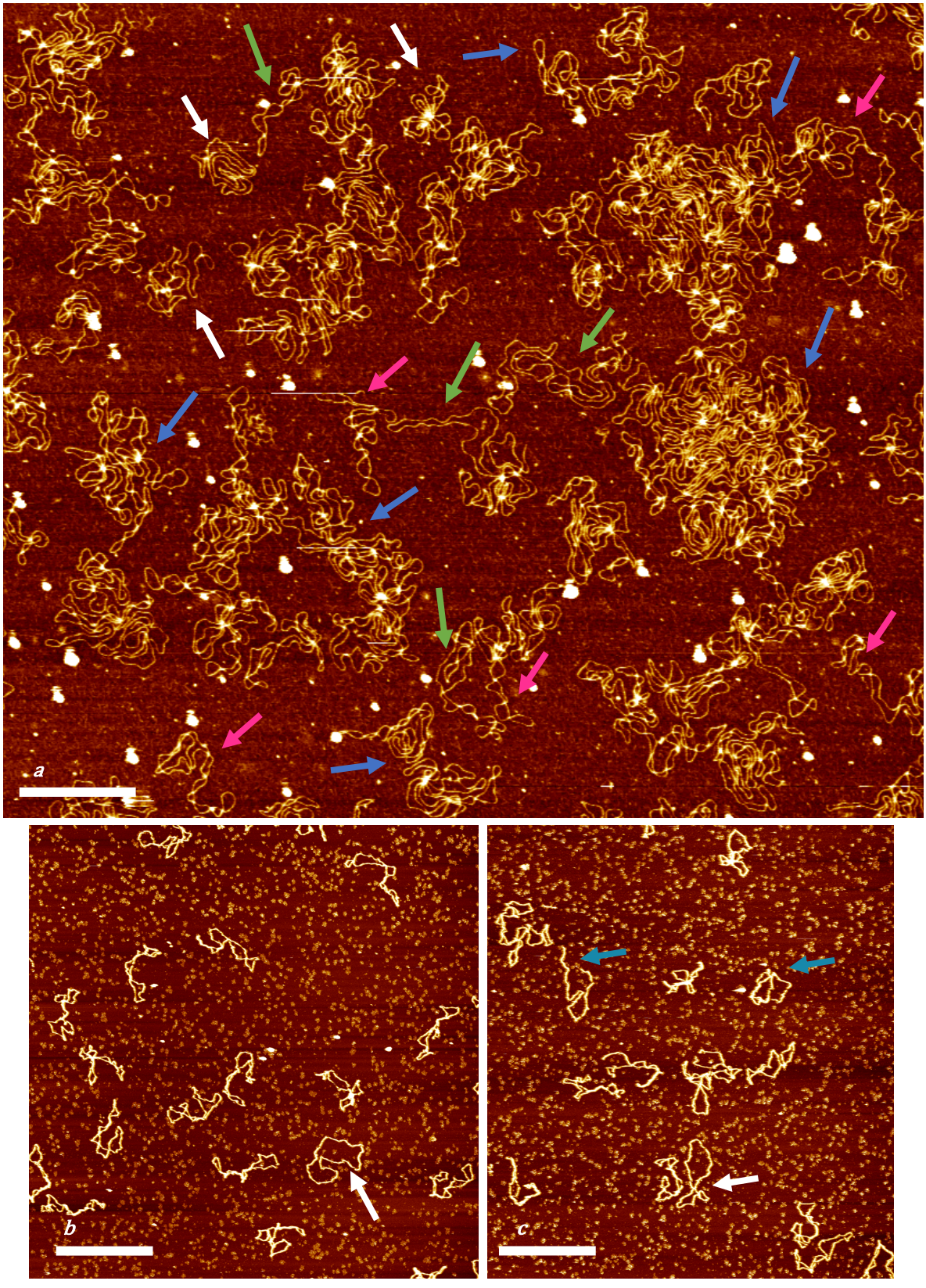


Figure 8.5: *parS* DNA plasmid (~ 3 kb): a. DNA in ParB reaction buffer (50 mM TRIS (pH 7.5), 100 mM NaCl and 5 mM $MgCl_2$) immobilized on PLO, washed with deposition buffer (HEPES buffer containing Mg^{2+} and K^{+}) followed by 2 mL HEPES, imaged in 10 mM HEPES. height 4.5 nm, scale 200 nm. b and c. DNA immobilized by method detailed by Heenan and Perkins (2019)[6], imaged in imaging buffer containing HEPES, Ni^{2+} and K^{+}, height 3.5 nm and 4.5 nm respectively, scale 400 nm.

not been observed by the previously used immobilization methods, or the ones used later. Some molecules were individually convoluted (white arrows) while some DNA molecules were comparatively relaxed and spread out (green arrows) at some parts of the molecule while the other parts were convoluted. Some molecules (pink arrows) were relaxed and showed the supercoiled conformation (entirely or partially on the molecule) seen previously when the plasmids were simply immobilized on PLO and washed with water (section 8.2). Since this method was a combination of the immobilization technique used previously and that proposed by Heenan and Perkins, it was understandable why some molecules would depict the shapes similar to those obtained when immobilized on PLO.

In comparison, Heenan and Perkins' method was adopted exactly to see how the 3 kb plasmid immobilized when allowed to spread on the mica surface. Figure 8.5 b and c show the DNA immobilized thusly. The images revealed most of the molecules coiled into twists, some more than others, while some rare molecules were spread (white arrows) and some were broken into open DNA and not twisted into coils (blue arrows). The rest were too coiled to resolve any helical conformation and would be difficult to observe interacting with ParB protein molecules in the future experiments. Interestingly, there were very scarce clusters of plasmids seen that could be more than 1 plasmid intertwined, and the interlinking was not complex like the meshwork of plasmids seen in case of immobilization on PLO (like figure 8.5 a). This indicated that the method was partially successful to spread DNA, though was not completely suitable for 3 kb DNA sample used in the current work as the DNA did not completely spread on the surface to allow the observation of individual strand conformations.

8.3.2. Immobilization of parS DNA and ParB protein reaction mixture

To compare the immobilization of DNA in the presence of protein, DNA and protein were mixed in the reaction buffer (50 mM TRIS (pH 7.5), 100 mM NaCl and 5 mM $MgCl_2$) in various ratios and immobilized by Heenan and Perkins' method on mica. The sample was then imaged in the imaging buffer containing Ni^{2+}. The images revealed that the DNA looked similar to the images without the presence of any protein, as in the previous section. There were some interlinking points seen on the DNA that could have either been DNA twisting (figure 8.6 and 8.7 c) or DNA strands interlinked by the binding of the ParB molecule. Since many DNA molecules showed similar conformations even in the absence of the protein (figure 8.5 b and c), it can be safely assumed that most of the twists seen here could have been DNA alone. In figure 8.7 b, the DNA clumped into conformations that could have been 2 molecules

intertwined but too convoluted to distinguish, again similar to the images obtained without protein. Even at different ratio of DNA-protein, figures 8.6-8.10 were very similar to images obtained for DNA negative control, mostly coiled or convoluted DNA strands with some very rare instances of DNA spread out on the surface (encircled in figure 8.8).

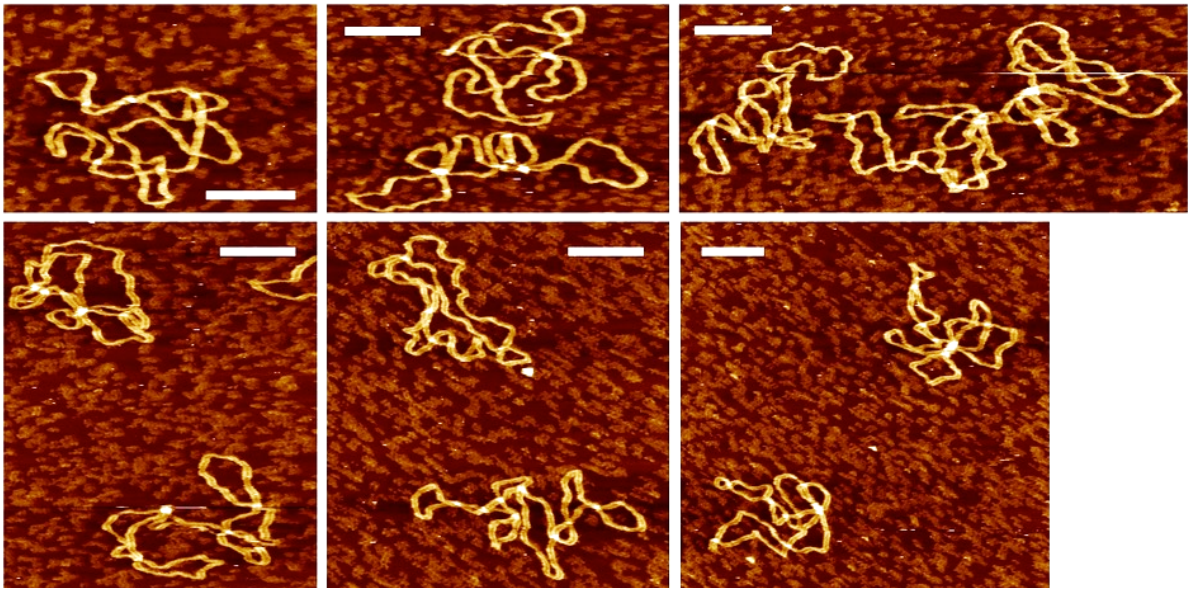


Figure 8.6: *parS* DNA plasmid (1 ng/µL) with ParB protein (1 nM) in Imaging buffer: height 4.5 nm, scale bars 100 nm. DNA-protein sample immobilized by Heenan and Perkins' method did not show any spreading of DNA.

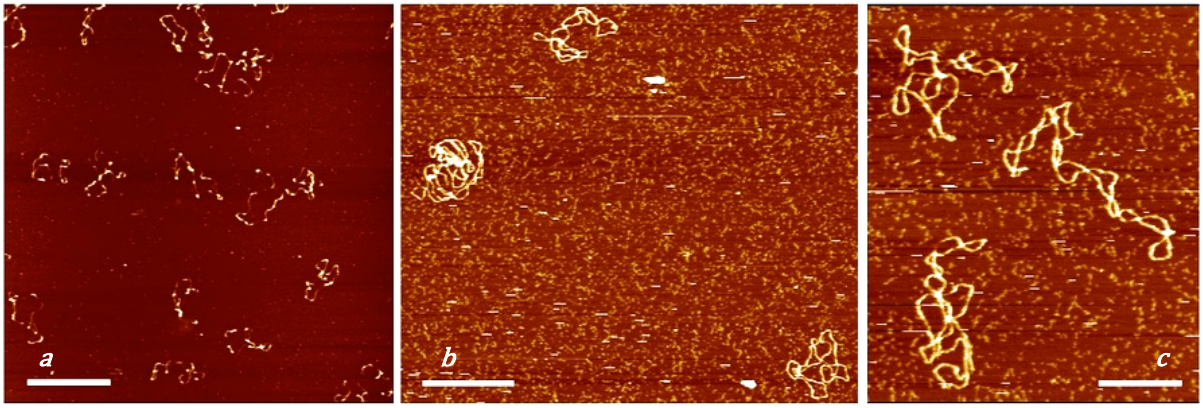


Figure 8.7: *parS* DNA plasmid (1 ng/µL) with ParB protein (2 nM) in Imaging buffer: a. height 5 nm, scale bar 400 nm, b. height 4 nm, scale bar 200 nm, c. height 4 nm, scale bar 100 nm. a had been imaged possibly at a higher force, thus causing damage to the molecules.

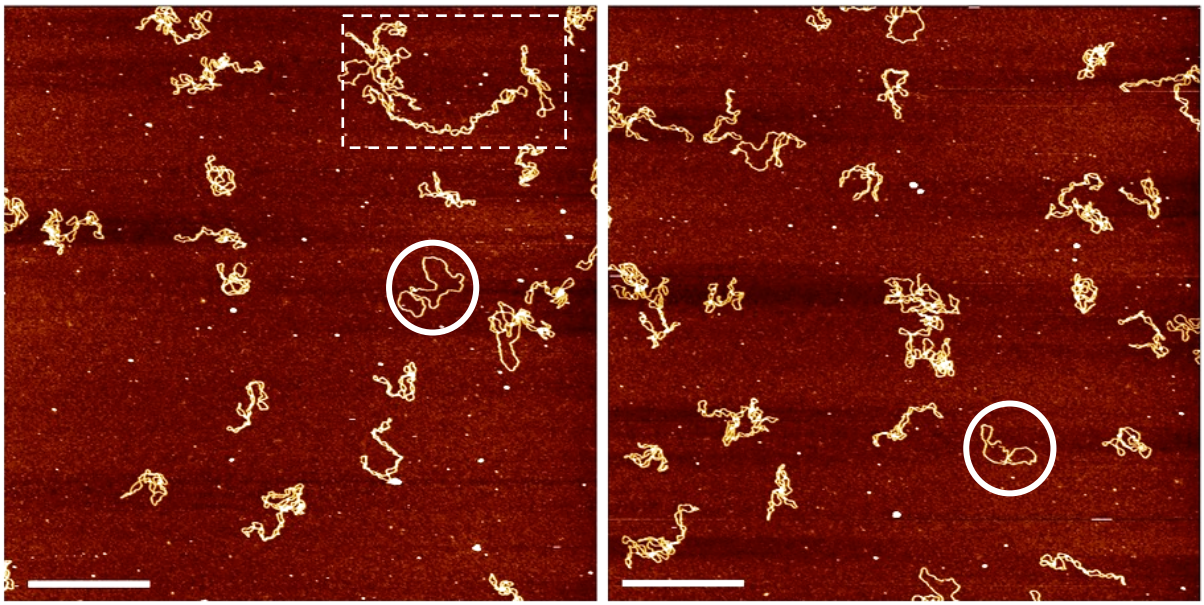


Figure 8.8: *parS* DNA plasmid (2 ng/µL) with ParB protein (0.5 nM) in Imaging buffer: height 4.8 nm, scale bar 400 nm. DNA-protein sample immobilized by Heenan and Perkins' method did not show any spreading of DNA. The box section of the image is zoomed in figure 8.9.

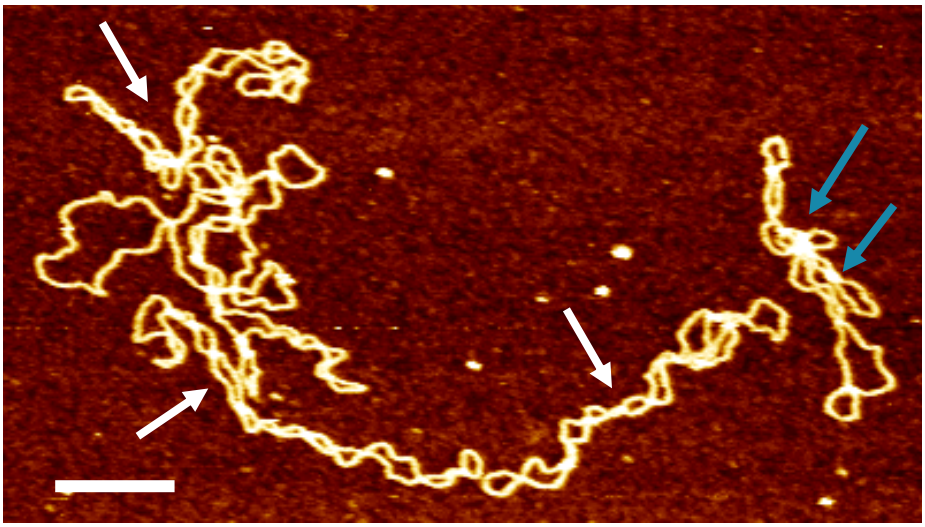


Figure 8.9: *parS* DNA plasmid (2 ng/µL) with ParB protein (0.5 nM) in Imaging buffer: height 4.8 nm, scale bar 100 nm. Frame zoomed from figure 8.8. The white arrows show some of the supercoiled DNA segments. The blue arrows mark the positions that could have been protein causing the convolution of DNA, but there was no way to be sure.

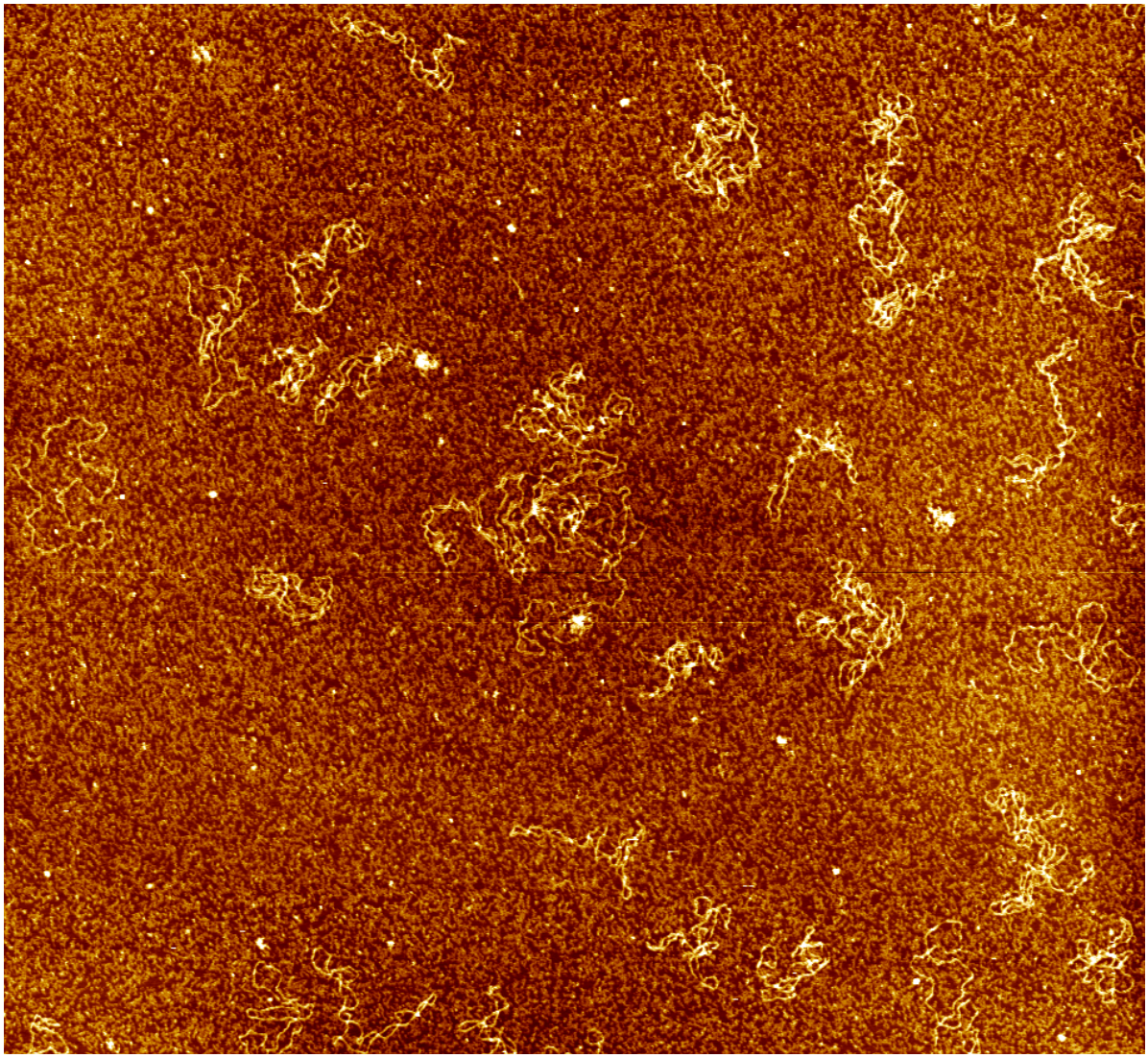


Figure 8.10: *parS* DNA plasmid (2 ng/µL) with ParB protein (1 nM) in Imaging buffer: height 4 nm, scale bar 200 nm. The encircled DNA are spread out on the surface as compared to the other molecules that are very convoluted.

This method of immobilization was thus rejected for the plasmid samples, because

a. the length of the *parS* plasmid was larger than the range of DNA recommended by the authors (200-2000 bp), and
b. the imaging buffer containing Ni^{2+} could be unsuitable to study the DNA-protein interaction as the absence of Mg^{2+} would disrupt the protein activity.

Therefore, another approach for the immobilization of DNA molecules was required to allow the DNA to be seen without any coiling or clumping before adding the protein to the reaction solution.

8.4. Imaging Short Open DNA containing various parS sites interacting with ParB Protein

From the experiments so far, it was realized that the length of the DNA plasmids was too long to be immobilized on the mica surface without being convoluted. The lack of spreading of the DNA prohibited the observation of ParB interaction with DNA. Methods to immobilize the DNA in a way to spread the molecules were unsuccessful as well. Hence, it was necessary to synthesize short DNA fragments containing *parS* sites that could be immobilized on the PLO treated mica surface without coiling in the absence of protein. This would nevertheless pose the problem that ParB proteins have been known to slide along the DNA's length and closed conformation of DNA prevented the protein from coming off the molecule [1]. Open DNA strands, with the additional characteristic of being short, would possibly cause the protein to not be observed bound to DNA as the former would slide off the DNA strands. Nevertheless, the interactions were observed anyway.

Plasmid DNA containing a varied number of *parS* sites (0-9) were used as templates with suitable primers (M13 primers in this case) to clone DNA molecules by PCR to obtain lengths containing the *parS* sites. The lengths varied from 68 to 155 nm for DNA that ranged from 200 to 455 bp (table 8.1) (see appendix 15-17 for details of the DNA sequence). The molecules were then purified on an agarose gel and the concentration determined using a spectrophotometer (see section 4.2.7 for details). DNA sample was diluted with reaction buffer (50 mM TRIS (pH 7.5), 100 mM NaCl and 5 mM $MgCl_2$), and immobilized on PLO treated mica to image the negative control sample. As already described in section 4.2.8.2, the protein solution was added to this reaction mixture to obtain the DNA-protein reaction mixture, which was incubated at 21ºC for 5-10 minutes to allow the protein-DNA interaction. This solution was then further diluted to obtain an AFM suitable concentration before immobilization on PLO treated mica. The samples were then imaged in air and water.

	Actual lengths (bp)	Theoretical length (nm)
0 *parS*	200	68
1 *parS*	215	73
2 *parS*	250	85
9 *parS*	455	155

Table 8.1: Lengths of *parS* site containing DNA.

8.4.1. parS-ParB interactions in the absence of CTP

Firstly, the interactions were studied in the absence of CTP in the reaction buffer. Negative control DNA sample solution and protein-DNA reaction mixture were imaged in air and water. The lengths of DNA were traced using NeuronJ software and plotted to compare the lengths. The argument here was that the DNA would remain in the extended form in the absence of protein, but the binding of ParB would either cause a slight bend or substantial folding. Either of these instances would cause the DNA to shorten. Hence, a percentage of shortening could be used as a parameter for protein binding.

The lengths of the DNA were traced using NeuronJ and plotted for comparison using MS Excel. The difference in the measured lengths of DNA from the theoretical lengths (calculated as the number of bp x 0.34 nm, assuming the DNA is in B-form) was calculated and plotted on the x-axis. The y-axis denoted the frequency of DNA with the corresponding difference in length from the theoretical number. The statistical significance of the differences in the data distribution for DNA imaged in air, with and without the protein was calculated using the t-Test for two-samples assuming Unequal Variances (because the variances of the sets of data were very different from each other, see appendix 18 for details). The P-values obtained in each case were less than 0.05, signifying that the data was indeed different and the two experimental setups of the absence and presence of ParB caused the DNA to behave differently.

The images (figure 8.11, 13, 15 and 17) revealed that

a. DNA containing 0 *parS* sites immobilized, as expected, as more or less straight strands without any sharp folding and coiling. The conformation remained the same in air as well as in water (figure 8.11 a-b). When the protein containing solution was imaged in air (figure 8.11 c), the DNA were indistinguishable from those of DNA negative control, with no possible protein molecules bound to the DNA, while some protein molecules had immobilized on the PLO surface (*e.g.* the ones marked by green arrows). On the contrary, when the sample was imaged in water (figure 8.11 d), there was a lot of noise observed on the images, which was deduced to be protein molecules which were somehow not visible in such abundant quantities and large size while imaging in air. It is possible that the addition of water had somehow disrupted the equilibrium of the protein reaction buffer and caused the molecules to aggregate, thus contribute to noise. The DNA were seen scarcely between

the crowd of protein molecules (figure 8.11 d), were partially eclipsed and this inhibited the measurements of length to quantify the protein binding to DNA.

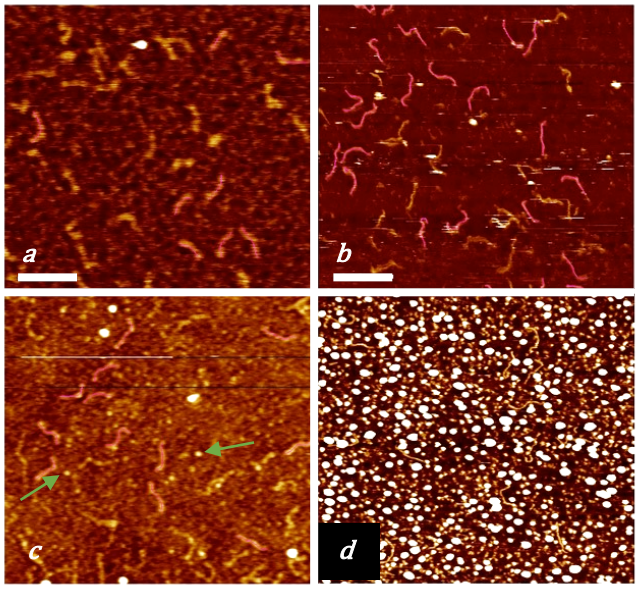


Figure 8.11: 0 *parS* DNA in buffer without CTP: DNA imaged in a. air (height 2.5 nm, scale 100 nm), b. water (height 5 nm, scale bar 100 nm), c. air with ParB protein (height 3 nm, scale bar 100 nm) and d. water with ParB protein (height 4 nm, scale bar 100 nm). There was no binding of protein to DNA observed, due to the absence of sequence non-specific binding.

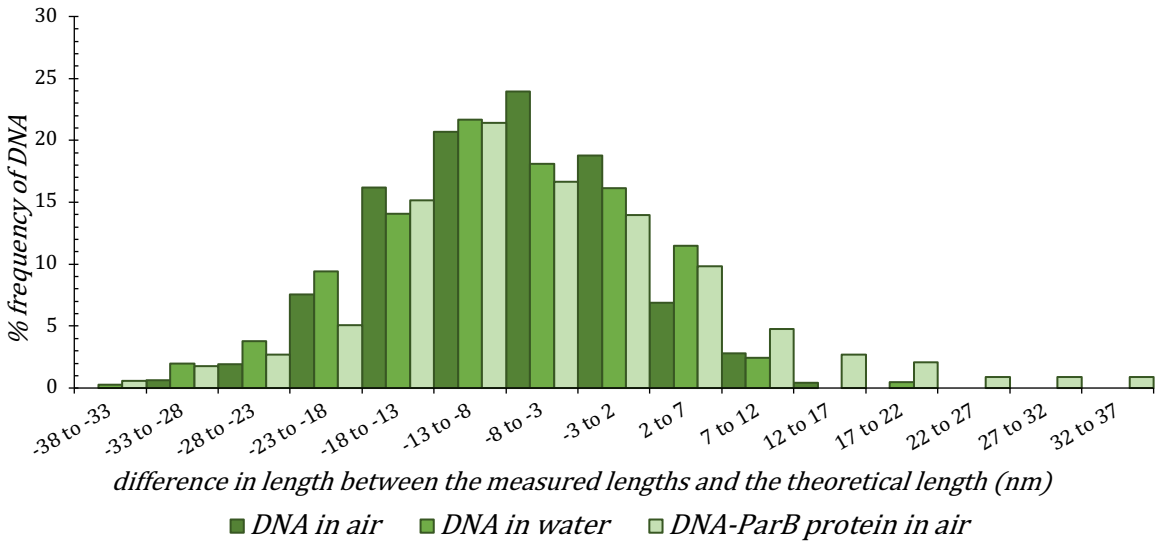


Figure 8.12: Comparison of lengths of 0 *parS* site containing DNA in non-CTP buffer. P-value for comparison between DNA with and without protein in air is 0.0042. There was no difference in the lengths of DNA seen when imaged in air in the presence and absence of protein, indicating that the protein had not bound to DNA in the absence of a DNA binding site.

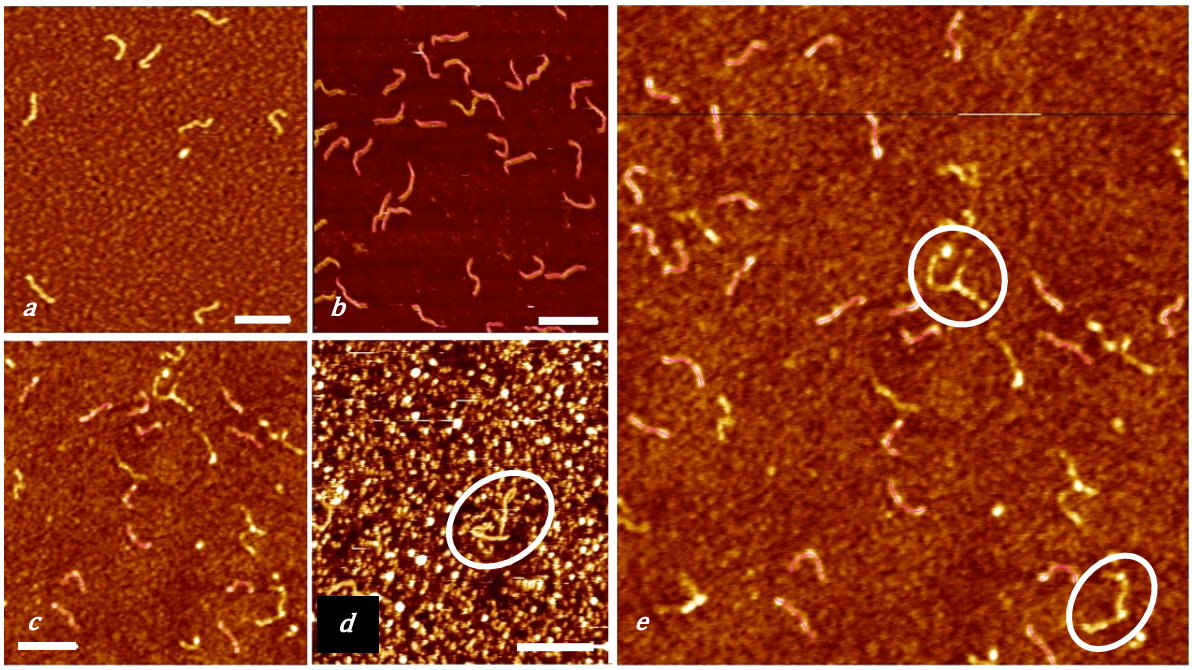


Figure 8.13: 1 *parS* DNA in buffer without CTP: DNA imaged in a. air (height 2.6 nm, scale 100 nm), b. water (height 5 nm, scale bar 100 nm), c and e. air with ParB protein (height 3 nm, scale bar 100 nm) and d. water with ParB protein (height 4 nm, scale bar 100 nm). There is beadiness seen on the DNA molecules in the presence of ParB and a prominent shortness of length as compared to DNA sample alone imaged in air.

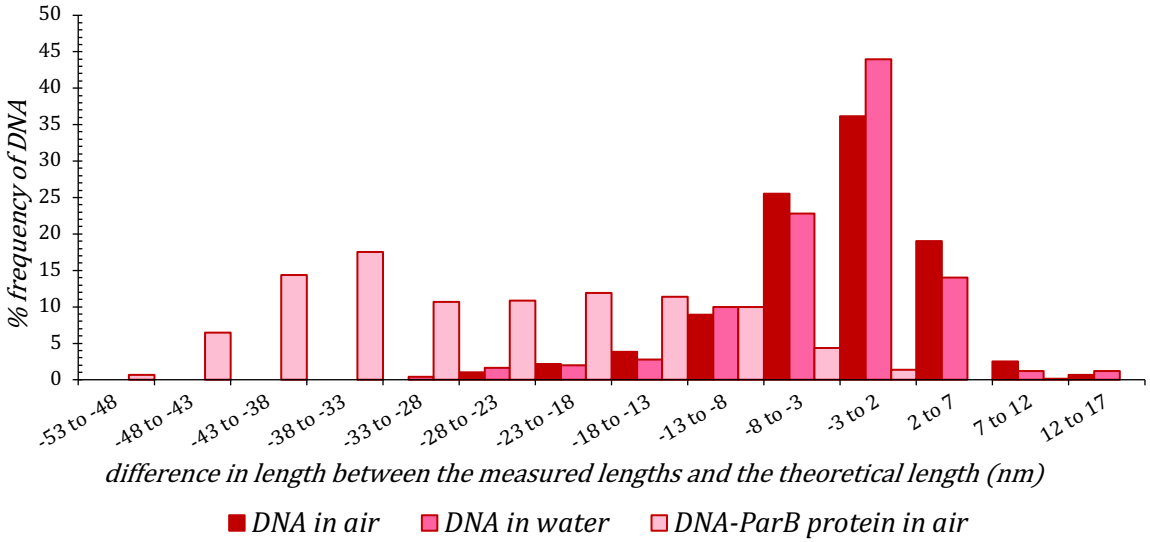


Figure 8.14: Comparison of lengths of 1 *parS* site containing DNA in non-CTP buffer P-value for comparison between DNA with and without protein in air is 6.0E-214. There was a significant flattening and shifting of the curve for DNA length differences in the presence of protein, indicating that ParB had bound the DNA.

The histograms of DNA in air, water and in the presence of protein in air (figure 8.12) showed that there wasn't much difference in the length distribution of the DNA in each of

the cases. The p-value of the data for DNA with and without protein imaged in air was also only about 12 times less indicating a closeness in their behaviour. The similarity signified that in the absence of any DNA binding site, the protein did not bind to the DNA in the absence of CTP in the buffer. This was in reconciliation with the proposed theory that CTP is necessary for the ParB to bind the DNA in a sequence non-specific manner and that in the absence of *parS* sites the protein did not bind to DNA.

b. DNA containing 1 *parS* site when imaged in air and water showed images (figure 8.13 a and b) similar to DNA with 0 *parS* site. The measured length of DNA was marginally more than DNA with no *parS* site, which could be seen on the plots comparing the lengths of DNA (figure 8.12 vs 8.14).

   The images of DNA-protein sample (figure 8.13 c and e) in air revealed the DNA to have protein 'beads' attached onto them. Many molecules were shorter than the observed control length of ~ 70 nm and measured to an average of ~ 47 nm (see table 8.2). This shortening of length indicated that the protein had bound to DNA and subsequently caused its bending or folding. There were also many instances of DNA molecules cross-linking (encircled in figure 8.13 e), which could be due to the binding of a protein molecule to multiple DNA molecules' *parS* sites and interlinking them.

   However, when the sample was imaged in water (figure 8.13 d), the protein molecules aggregated and crowded the image, similar to the 0 *parS* DNA sample, but interestingly, the meagre amount of DNA molecules that could be distinguished (*e.g.* the ones encircled) did not show any beadiness from the protein binding and the length of DNA was comparably same as the DNA imaged in water without any protein present. Owing to the crowded protein molecules, the lengths of DNA were not traced on the images in water.

   The images showed that there was protein binding to 1 *parS* site containing DNA, and that caused the shortening of the DNA lengths. However, the conformation of the bound DNA could not be seen.

   The histograms of the comparison of lengths of DNA in air and water and in the presence of protein (figure 8.14) showed that while the plots for DNA in air and water were very similar, there was a significant shift for DNA-protein complexes imaged in air. A small fraction of the molecules had the lengths corresponding to the length without any protein bound, but about half of the DNA had length drastically reduced. A large section's length was even less than the shortest DNA measured without the protein (~ 45-15 nm reduction in length for DNA with protein as compared to about -15 nm to +7 nm variation in length

for most of the DNA measured without protein). Also, the p-value measured for the data set (DNA in air with (out) protein) was extremely low, hence indicating that the behaviour of DNA was very different due to the presence of the protein.

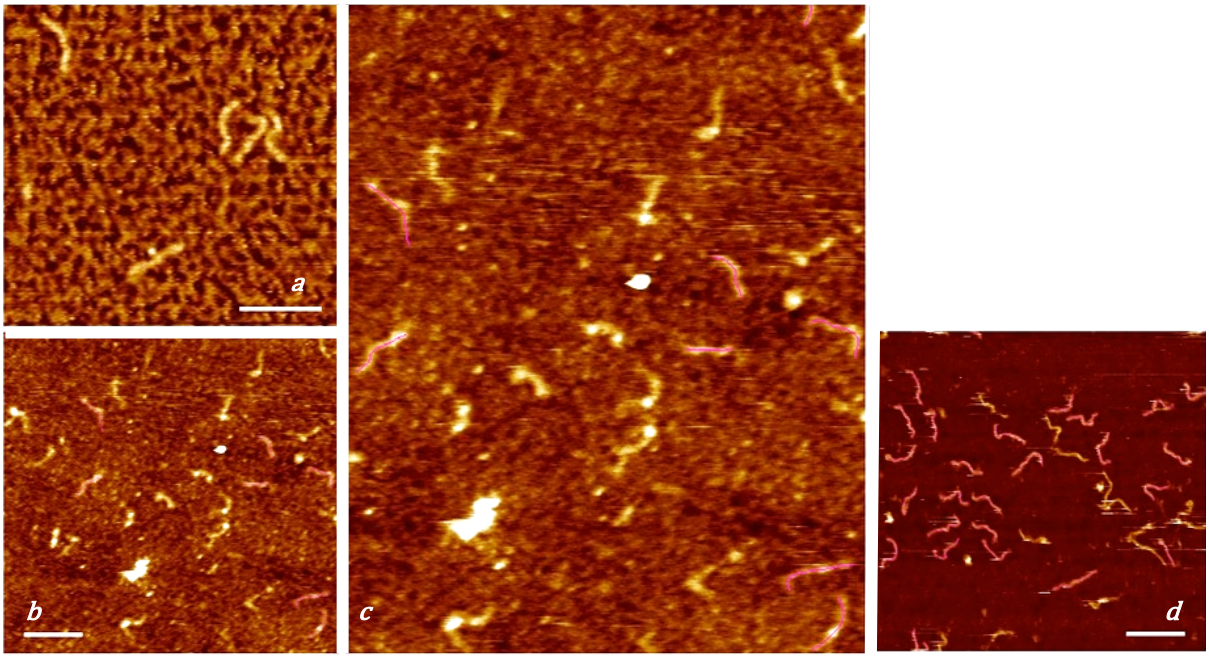


Figure 8.15: 2 *parS* DNA in buffer without CTP: DNA imaged in a. air (height 3 nm, scale 100 nm), b and c. air with ParB protein (height 2 nm, scale bar 100 nm) and d. water (height 5 nm, scale bar 100 nm). The DNA molecules with protein show some beadiness due to protein binding to it, but the lengths are much reduced as compared to the images of DNA in air.

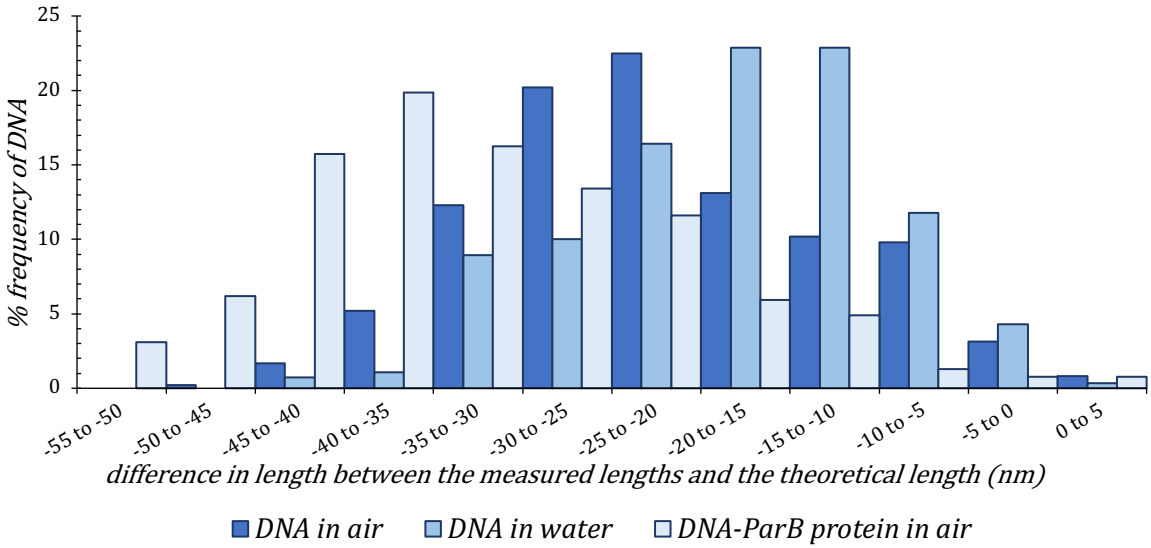


Figure 8.16: Comparison of lengths of 2 *parS* site containing DNA in non-CTP buffer. P-value for comparison between DNA with and without protein in air is 2.0E-40. There was a shift in the plots in the presence of protein towards the shorter length of DNA but the shift was not as prominent as the case of 1 *parS* site DNA.

c. DNA with 2 *parS* sites imaged in air and water (figure 8.15 a and d) depicted strands without any beadiness or bends and had lengths averaging at ~ 64 nm, which was contradictory as it was shorter than the lengths observed for 1 *parS* site DNA and also the theoretical length expected. In comparison, DNA molecules imaged with protein in air depicted a very different conformation than without protein. The DNA were visibly shorter here (figure 8.15 b and c) and showed the presence of protein molecules bound to them. A measurement of the lengths of DNA (figure 8.16) revealed that there was a shift in the peak of DNA length difference by about 10 nm when the protein bound to it. However, the shift was not as pronounced as the 1 *parS* site DNA's interaction with the protein and the p-value, though extremely low, was far more than the p-value for 1 *parS* DNA length comparisons. It was not possible to observe any conformation changes occurring during the interaction, hence any possible explanation for this anomalous behaviour could not be proposed.

   The length was more in case of images in water for DNA sample, probably due to the tip-sample dynamics being different for imaging in liquid. Lengths were not measured for DNA with protein imaged in water as the images had a lot of noise, like the 0 and 1 *parS* site DNA.

d. 9 *parS* site containing DNA imaged in air and water (figure 8.17 a and b) revealed DNA strands of lengths substantially longer than the DNA observed so far. This agreed with the theoretical length of 9 *parS* DNA of 155 nm, more than double the length of 0 *parS* DNA. Here, the length of DNA was measured to average at 127 nm from the images in air and 133 nm from the images in water.

   Interestingly, the DNA-protein sample images displayed some DNA strands which were of a length comparable to the DNA without protein (marked by white arrow in figure 8.17 c). These were probably the DNA without any protein bound and their length measurements contributed to the increase in the average length of the DNA sample. However, there were other DNA strands that showed the protein 'beads' stuck to them accompanied by a decrease in the length of DNA (like the DNA marked by green arrow). These measurements overall gave the average length of DNA in the presence of protein to be ~103 nm, which was less than the lengths observed for DNA without the protein bound.

   The comparison of the difference in lengths of DNA from the theoretical lengths (figure 8.18) revealed that the plots for DNA in air and water were similar, with the images in water showing the lengths to be larger. In comparison, there was a significant shift and flattening

of the histogram for the images of the DNA-protein complex. Almost half of the DNA had lengths in the same bracket as DNA without protein, but the other half measured to be even shorter than the shortest DNA in the absence of ParB. This reduction showed that protein-DNA interaction had occurred in case of 9 *parS* site DNA.

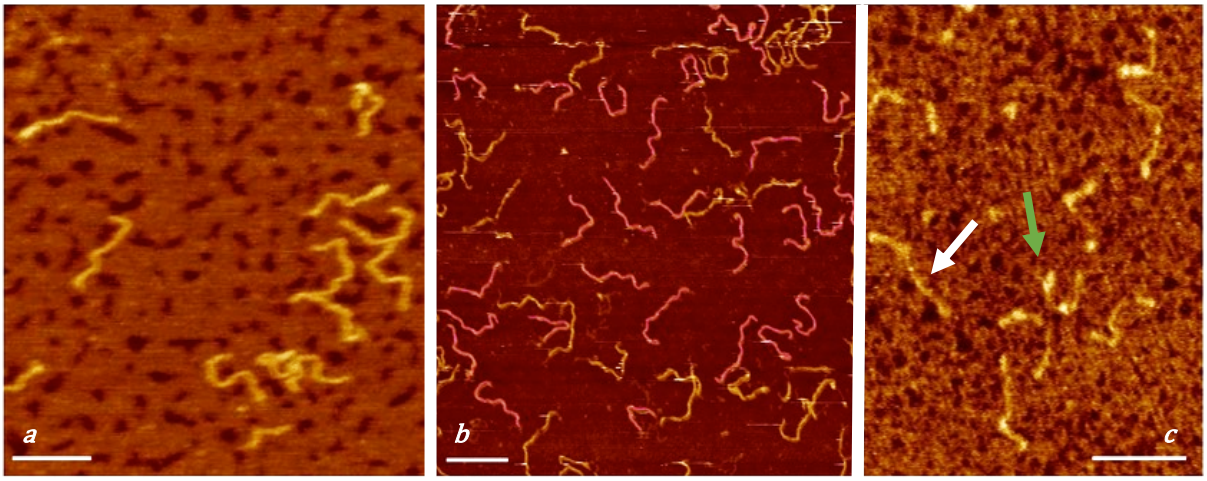


Figure 8.17: 9 *parS* DNA in buffer without CTP: DNA imaged in a. air (height 3 nm, scale 100 nm), b. water (height 7 nm, scale bar 100 nm), c. air with ParB protein (height 3 nm, scale bar 100 nm). There were some protein molecules bound to DNA showing the beadiness on the DNA.

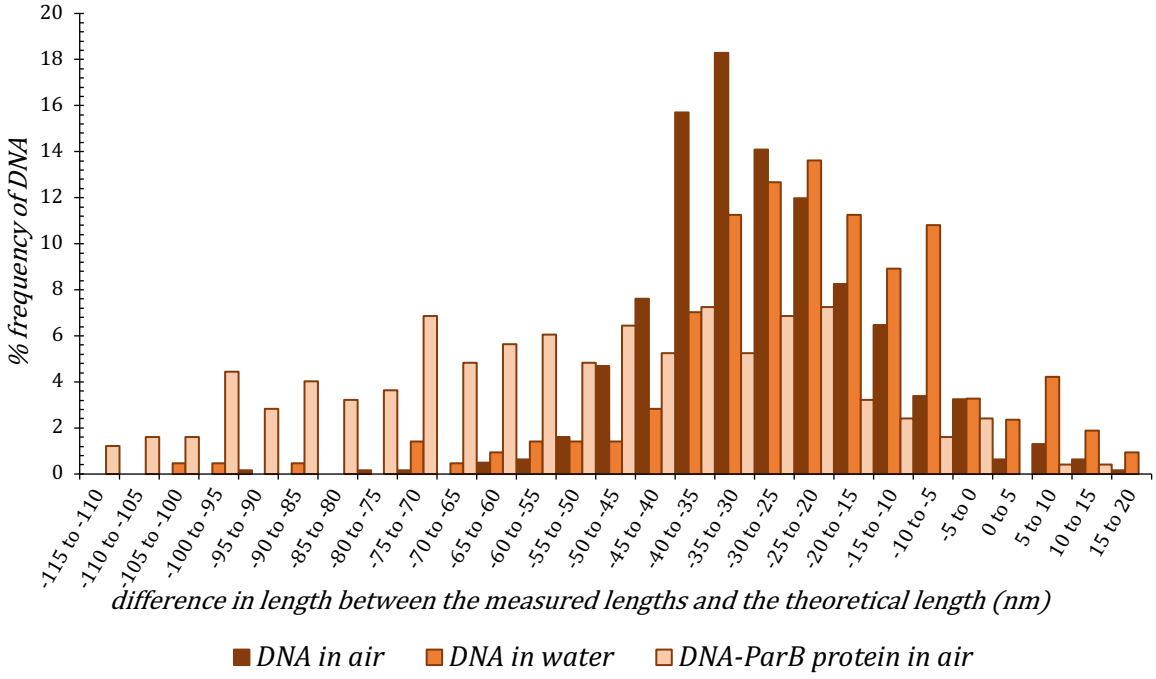


Figure 8.18: Comparison of lengths of 9 *parS* site containing DNA in non-CTP buffer. P-value for comparison between DNA with and without protein in air is 6.9E-30. There was a significant flattening of the curve for DNA length shifts in the presence of protein, indicating that ParB had bound to DNA. Some fraction of the DNA retained the length measurements same as DNA without any protein while almost half of the DNA had shorter lengths.

The overall lengths of DNA were compared with and without the presence of protein when imaged in air. The numbers have been tabulated in table 8.2 where the percentage reduction in the length of DNA without the protein vs with the protein had been calculated using the equation:

$$\%\, reduction = \frac{(average\ length\ of\ DNA\ without\ protein) - (average\ length\ of\ DNA\ with\ protein)}{average\ length\ of\ DNA\ without\ protein} X\ 100$$

The percentage revealed that the maximum amount of reduction in length by protein binding occurred for DNA with 1 *parS* site, followed by 9 and then 2, with DNA containing 0 *parS* site showing no significant change in the DNA lengths. This indicated that there was no sequence non-specific binding of the protein in the absence of CTP in the buffer and the protein required *parS* sites to bind. For 9 *parS* site containing DNA, it was expected that multiple protein molecules would bind and cause a much larger percentage reduction in length, but this was not the case. In this and the instance of 2 *parS* sites, the proximity of the sequences may have caused the protein to interfere with the binding of other protein molecules near the site where the protein was already bound. Thus, it could be deduced that the steric hindrance of the protein reduced its sequence non-specific binding to DNA.

The results were attempted to be repeated in presence of CTP in the buffer and have been discussed in the following section.

	Actual Lengths in bp	Theoretical Length in nm	DNA in air	DNA in water	DNA-protein in air	% Decrease in length in air
0 *ParS*	200	68	61	60	63	-3.51 %
1 *ParS*	215	73	71	71	47	33.80 %
2 *ParS*	250	85	64	68	54	15.63 %
9 *ParS*	455	155	127	133	103	18.90 %

Table 8.2: Comparison of lengths of *parS* site containing DNA in non-CTP buffer.

8.4.2. parS-ParB interactions in the presence of CTP

It has been reported that in DNA of length about 200 bp with the ends blocked, CTP enhanced the non-specific binding of ParB to the entire DNA segments and increased the *parS* dependent spreading of ParB [1]. This effect of CTP on the binding of ParB to DNA to the *parS* sites specifically and in a sequence non-specific manner was examined using AFM. The above-described set of experiments for *parS*-ParB interaction were repeated with CTP (1 μM or 0.5 μM) in the reaction buffer and imaged in air and water. The variation in the CTP concentration did not produce any AFM image observable modifications in the results.

The images obtained (figures 8.19, 21, 23 and 25) revealed that

a. DNA containing 0 *parS* site immobilized on the surface as bent strands (figure 8.19) and not straight DNA as in the case of absence of CTP. However, the lengths of the DNA were comparable (71 nm) to the theoretical length expected (68 nm) (table 8.3). Contrary to the images in buffer without CTP, the DNA when imaged in the presence of protein showed a significant reduction in the length and displayed the slight 'beadiness' characteristic of protein binding to DNA. The observation was supported by the measurement of lengths of DNA (figure 8.20 and table 8.3) depicting the reduction in the average length of DNA from 71 nm without protein to 55 nm in the presence of protein, equating to a ~ 23 % reduction in the length of DNA. This was a significant decrease compared to the negligible change in the DNA length when the reaction buffer did not contain CTP. This could be explained by considering that the presence of CTP in the buffer allowed the non-specific binding of ParB to DNA even in the absence of any *parS* sites. This was in contradiction to the previously reported work [2,3] that *parS* sites are required for the recognition of DNA by ParB.
   The images for DNA-protein interactions imaged in water were similar to the images obtained previously (figures 8.11 d and 8.13 d) with protein aggregation and crowding inhibiting the observation of any DNA conformation during interaction with the protein. Hence, those images were not traced to measure DNA lengths.
   The plots of the frequency of DNA vs the difference between the observed lengths and theoretical length showed that there was a flattening of the histogram and shift towards the shorter lengths in case of DNA-protein sample, as compared to the DNA sample alone. This further strengthened the proposition that the presence of CTP allowed non-specific binding of ParB to DNA. (see appendix 19 for details for the t-test).

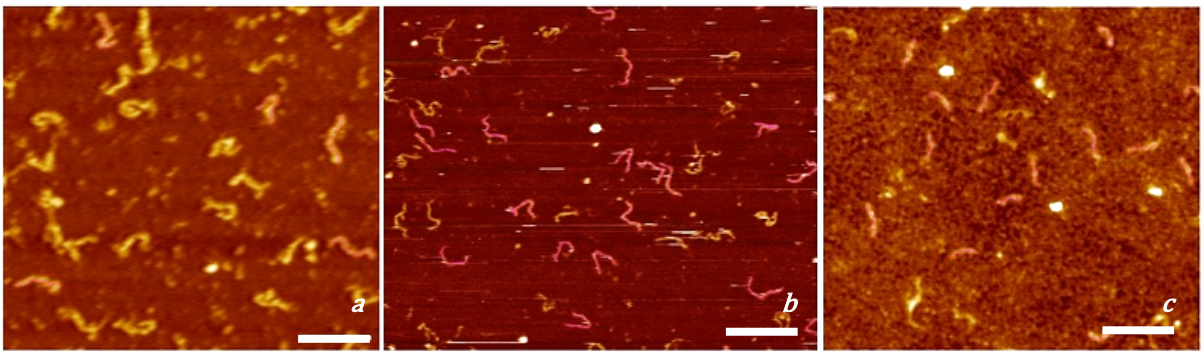


Figure 8.19: 0 *parS* DNA in buffer with CTP: DNA imaged in a. air (height 4 nm, scale 100 nm), b. water (height 7 nm, scale bar 100 nm), c. air with ParB protein (height 3 nm, scale 100 nm). DNA were seen as bent strands and the protein bound to them in a sequence non-specific manner, contrary to when the buffer did not contain CTP.

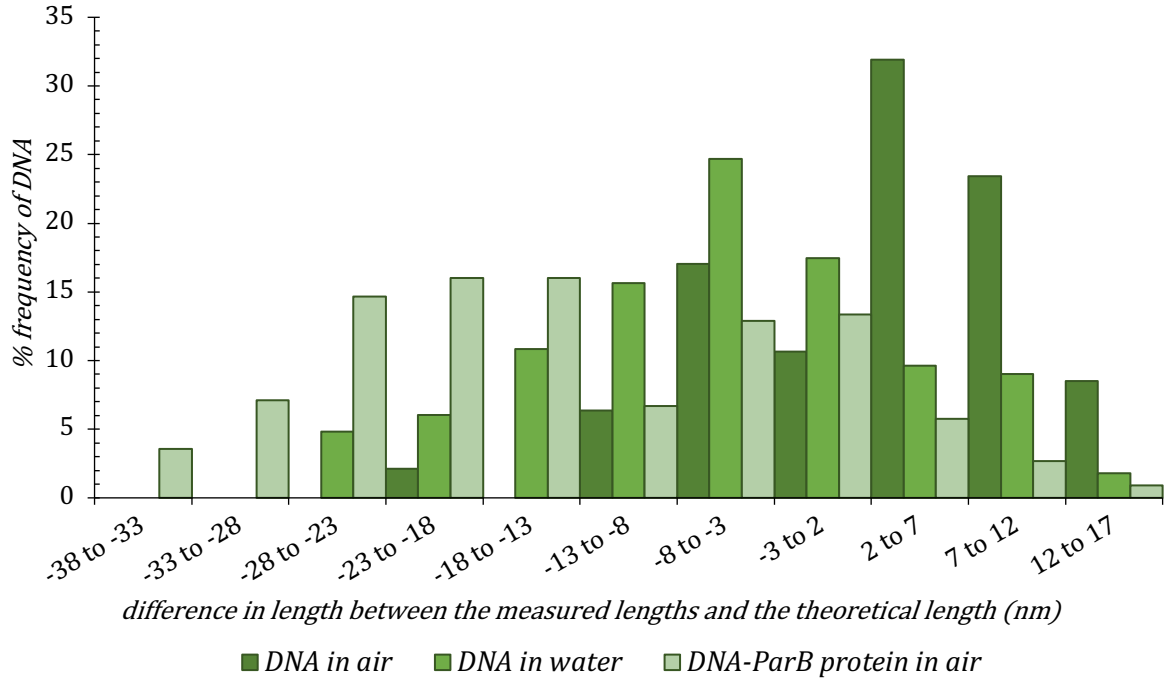


Figure 8.20: Comparison of lengths of 0 *parS* site containing DNA in CTP buffer: P-value for comparison between DNA with and without protein in air is 5.4E-22. There was a significant flattening and shifting of the curve when the DNA was bound to the protein, indicating the binding of protein to DNA, even in the absence of any DBS.

b. DNA with 1 *parS* site possessed elongated structures (figure 8.21 a-b) like in the reaction buffer without CTP. The images in water as well were similar to the images with CTP absent and the lengths averaged at 70 nm, comparable to the length obtained in reaction buffer without CTP (71 nm) (table 8.2). Interestingly, the DNA-protein sample (figure 8.21 c) also displayed characteristics similar to the images without CTP. The 'beadiness' due to protein binding to DNA could be observed and the DNA molecules were significantly shorter in the case of DNA-protein sample than DNA alone. There were instances where many DNA molecules crosslinked (encircled), as was seen in the images without CTP. This was probably due to the protein molecules binding to multiple DNA molecules' *parS* sites.

   Measurements of DNA lengths indicated a shortening of DNA in DNA-protein images from an average of 73 nm to 55 nm (table 8.3). From the length difference plots (figure 8.22) as well, it was observed that there was a prominent shift in the length peaks from about 5 nm reduction from theoretical length to most of the DNA having about 8-28 nm reduction in lengths from the theoretical value. This showed that the protein bound to DNA and caused the observed shortening.

   Interestingly, the percentage reduction in the average lengths of DNA from the theoretical value (tables 8.2 and 8.3) illustrated that there was more protein binding in the absence of CTP (~ 34 % reduction) than in its presence (~ 25 % reduction). This is in accordance with the propositions that CTP caused the switch from specific to non-specific binding and hence allowed the sliding of ParB on DNA. In the absence of something blocking the DNA ends, like in the case of the current work, the protein might have fallen off. Thus there was more

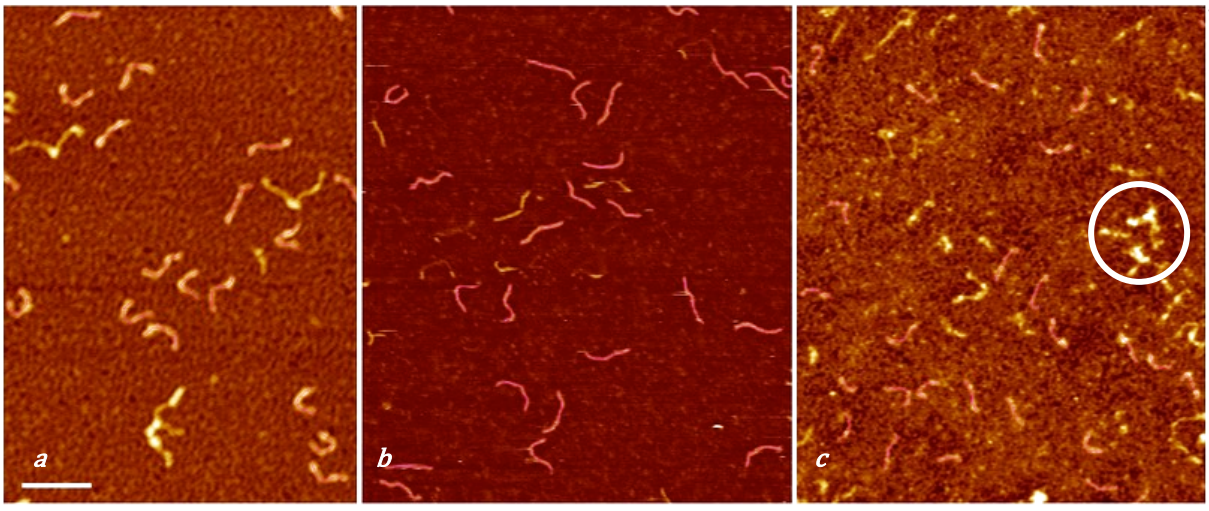


Figure 8.21: 1 *parS* DNA in buffer with CTP: DNA imaged in a. air (height 4 nm, scale 100 nm), b. water (height 7 nm, scale bar 100 nm), c. air with ParB protein (height 3 nm, scale 100 nm).

protein retained in the bound state in the absence of CTP in the buffer, than in the buffer containing CTP.

Another observation worth noting was that the percentage reduction in the average lengths of DNA from the theoretical value in case of DNA with 0 *parS* site was almost identical to the percentage for 1 *parS* site. Since the lengths of both DNA are very similar, this showed that there was a comparable amount of protein bound in both the cases. This showed that DNA without any *parS* site behaved as having a single DBS in the presence of CTP in the buffer.

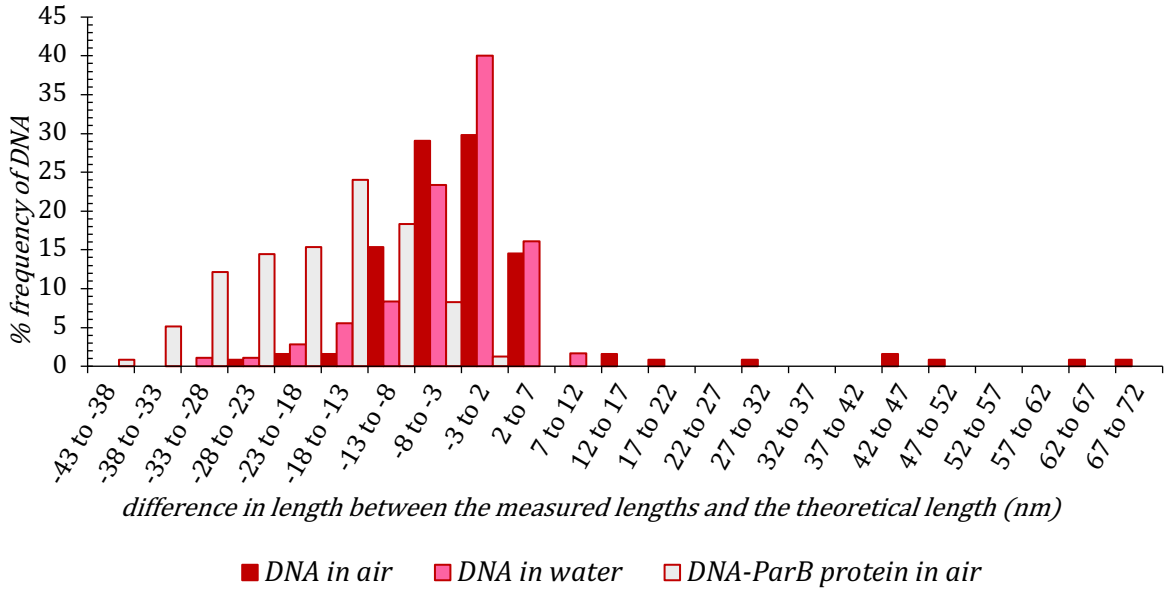


Figure 8.22: Comparison of lengths of 1 *parS* site containing DNA in CTP buffer: P-value for comparison between DNA with and without protein in air is 7.4E-29. A prominent shift in the peak indicated a reduced length of DNA in the presence of protein. There were many outliers with large lengths in case of DNA alone imaged in air and water, which were probably 2 or more DNA molecules close together measured as 1.

c. DNA with 2 *parS* sites also depicted the elongated structure of DNA (figures 8.23 a and b) as seen in the images with 1 *parS* site, but the lengths were very short (table 8.3) as compared to the theoretical lengths (102 nm). Incidentally, the exact same average length of DNA was obtained when 2 *parS* site containing DNA was imaged in air, immobilized in the reaction buffer without CTP (table 8.2), compared to the presence of CTP here. Also, the lengths obtained while imaging in water in both the instances of with and without CTP were close to 70 nm, and not to the theoretical value of 85 nm. This implied that 2 *parS* site containing DNA was somehow folded or coiled causing the DNA to immobilize as significantly shorter strands than the 0, 1 and 9 *parS* site containing counterparts which had lengths consistent with the theoretical values.

   DNA when imaged interacting with protein (figure 8.23 c) showed DNA with protein bound to them as 'beads on a string', though the lengths were not drastically different from the DNA negative control sample. The trace measurements confirmed the average lengths of DNA to be 59 nm long in the presence of protein (table 8.3) and 64 nm without the protein. Hence, the percentage decrease in the size was ~ 8 %. As was the case for the absence of CTP in the reaction buffer, the decrease % was least for DNA with 2 *parS* sites than the 1 and 9 *parS* site DNA counterparts. Hence the supposed protein interaction was lesser in case of 2 *parS* sites containing DNA than 1 *parS* site DNA. Also, like the case of 1 *parS* DNA, the percentage reduction was more in case of the absence of CTP than the presence, indicating that the CTP had probably caused the switch to ParB spreading and subsequent sliding off, causing a decrease in the percentage of binding.

   From the plots of difference of DNA lengths from the theoretical length vs the frequency of DNA with the corresponding lengths (figure 8.24), it could be observed that the DNA length distributions were very similar in case of DNA with and without protein. This further strengthened the argument that there was very little protein binding.

   Incidentally, if the length of DNA immobilized on the surface was close to the theoretical length, as in case of the other DNA strands with 0, 1 and 9 *parS* sites, the % reduction in the length from ~ 85 nm to 59 nm would have been ~ 30 % in the presence of CTP and from 85 nm to 54 nm would have been ~ 35 % in case of absence of CTP. Thus, from these observations, either of the following conclusions could be possible:

   1. The DNA (negative control sample) had immobilized in folded or bend conformation hence giving the short length measurements that were very different from the theoretical length expected. This had affected the percentage reduction in the length of DNA on

protein interaction, even though substantial ParB-*parS* binding had occurred, as the resultant comparative reduction in length was not much. Or,

2. Despite repetitive attempts of experimentation and contrasting results obtained for other *parS* site containing DNA, either the sample had not immobilized well as DNA negative-control, or the DNA-protein interactions had not occurred well.

In either case, it could be agreed that the results obtained in case of 2 *parS* site were an anomaly and required further experimentation.

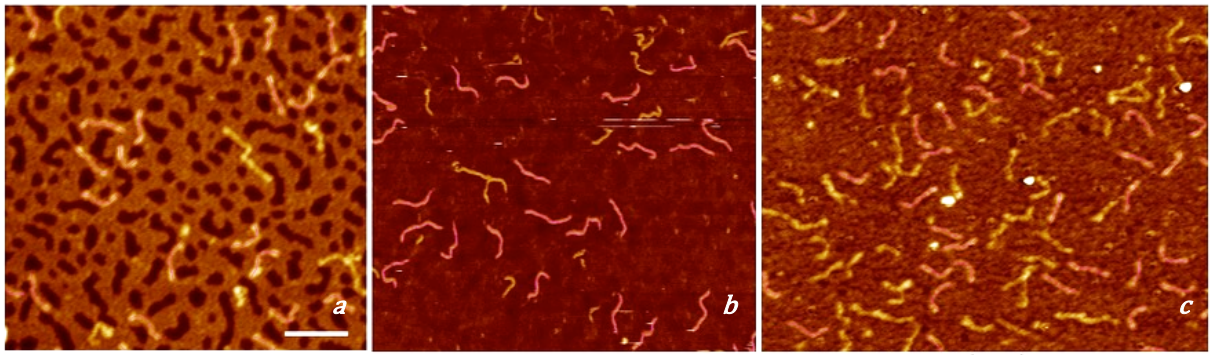


Figure 8.23: 2 *parS* DNA in buffer with CTP: DNA imaged in a. air (height 4 nm, scale 100 nm), b. water (height 7 nm, scale bar 100 nm), c. air with ParB protein (height 3 nm, scale 100 nm). The DNA lengths seen were very similar in the case of absence and presence of protein, showing that not much interaction had occurred.

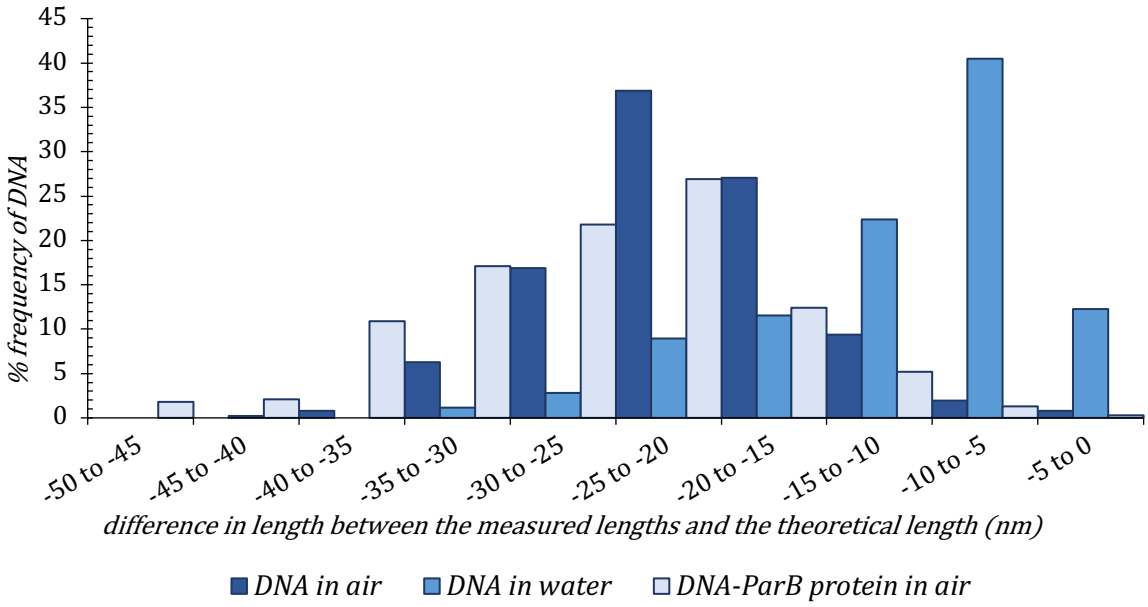


Figure 8.24: Comparison of lengths of 2 *parS* site containing DNA in CTP buffer: P-value for comparison between DNA with and without protein in air is 3.1E-21. There was not a significant shift in the peaks for DNA-protein solution imaged in air as compared to DNA sample alone, indicating only a small amount of DNA-protein interactions.

d. DNA with 9 *parS* sites immobilized as long strands (figure 8.25 a and c) with many bends and loops. The lengths averaged at 141 nm in air, which was close to the theoretical length expected (155 nm) (table 8.3). The DNA-protein sample revealed that a large percentage of DNA did not have any protein bound to them and remained elongated. Very few molecules had the appearance of proteins bound on them (marked by arrows). Nevertheless, random molecules were traced and their lengths measured to reveal an average of 137 nm, ~ 3 % decreased length from the negative control sample (figure 8.26 and table 8.3). On comparison with the data from the sample prepared in buffer without CTP, where the percentage reduction in the length of DNA in the presence of protein was ~ 19 %, it could be concluded that more interactions occurred in the absence of CTP, than with. This may suggest that CTP increased the sliding of ParB on DNA, causing the protein to slide off and hence show less interactions.

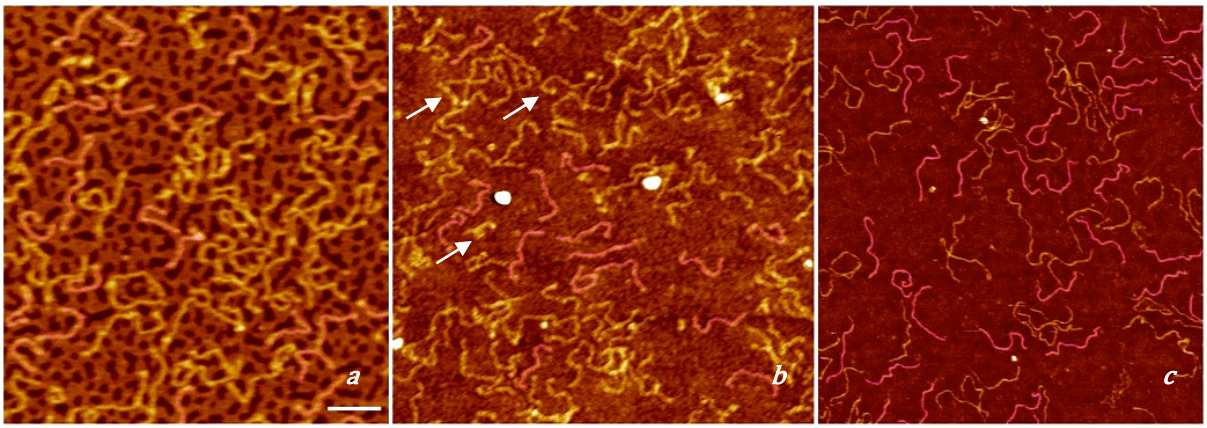


Figure 8.25: 9 *parS* DNA in buffer with CTP: DNA imaged in a. air (height 4 nm, scale 100 nm), b. air with ParB protein (height 3 nm, scale bar 100 nm), c. air (height 7 nm, scale bar 100 nm).

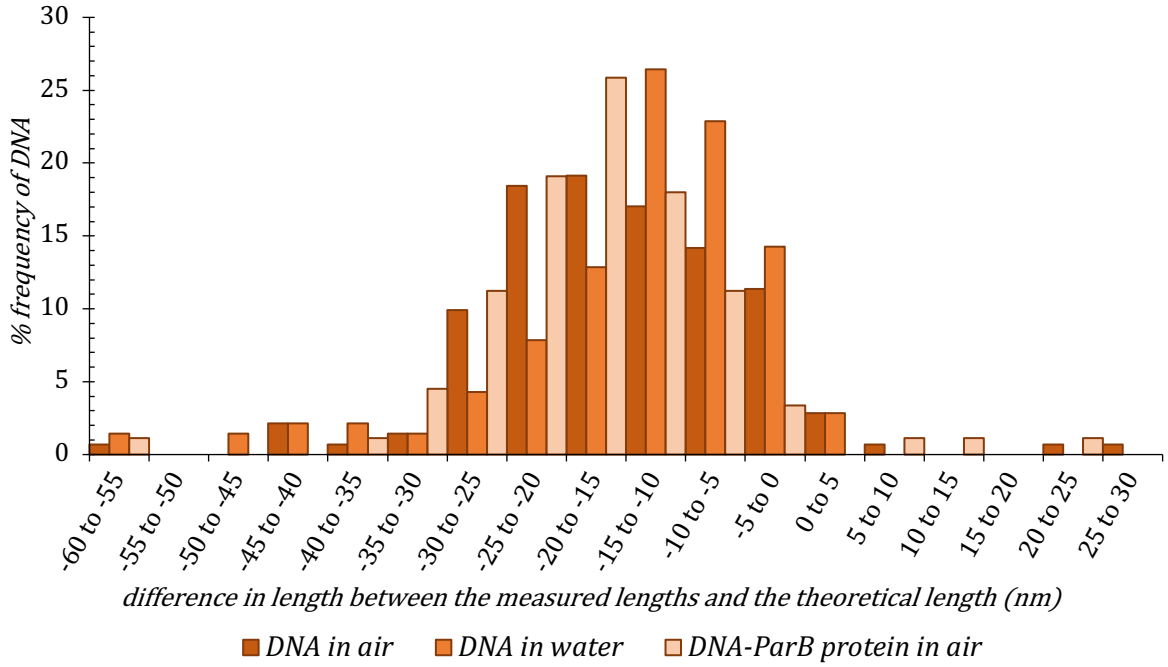


Figure 8.26: Comparison of lengths of 9 *parS* site containing DNA in CTP buffer: P-value for comparison between DNA with and without protein in air is 0.04. Not much variation in the plots for DNA and DNA-protein sample imaged in air, indicating a lack of significant protein binding.

	Actual Lengths in bp	Theoretical Length in nm	DNA in air	DNA in water	DNA-protein in air	% Decrease in length in air
0 *ParS*	200	68	71	63	55	22.54 %
1 *ParS*	215	73	73	70	55	24.66 %
2 *ParS*	250	85	64	74	59	7.81 %
9 *ParS*	455	155	141	141	137	2.84 %

Table 8.3: Comparison of lengths of *parS* site containing DNA in CTP buffer.

8.5. General Discussion and Conclusion

From the results so far, it can be concluded that DNA plasmid can be visualized using AFM to reveal helical resolution, though the DNA immobilized on the surface in a coiled and twisted conformation that made it difficult to observe protein binding to DNA and the associated conformational changes. Though sample immobilization protocols have been proposed to allow the DNA to spread onto the mica surface, those methods were not suitable for the ~ 3 kb plasmids in our case. The required length of 200-2000 bp was very difficult to achieve for *parS* plasmids and could be possible by synthesizing overhang DNA of that length containing *parS* sites followed by circularization to form closed DNA by end ligation.

To make things simpler, short open DNA could be immobilized as elongated and relaxed molecules on the PLO treated mica surface. Due to protein binding, the DNA molecules displayed a 'beadiness' and reduction in their length in comparison to the length of DNA negative control. The binding efficiency of protein was maximum to DNA with 1 *parS* site, followed by 9 and 2 *parS* sites. In the absence of CTP in the buffer, the protein mainly bound specifically to the *parS* sites. There was no protein bound to the DNA with no *parS* site, showing the absence of a sequence non-specific protein binding.

The non-specific binding was enhanced by the presence of CTP in the buffer: 0 *parS* DNA did not bind with protein in the absence of CTP but its presence caused the ParB molecules to non-specifically bind DNA. 1 *parS* site DNA showed protein binding which was reduced in the presence of CTP, most probably because the presence of CTP also facilitated the switch from specific binding to non-specific binding causing the protein to slide off the DNA [1–3]. 2 *parS* site DNA displayed an anomalous behaviour and immobilized with a length shorter than the theoretical length of DNA. Though protein binding had occurred to some extent and could be seen in some molecules on the images, the percentage reduction in length was not comparable to the other kind of DNA sample used. This could be a contribution of the decreased length of the DNA control sample or an error in the experimentation, despite repeated attempts. Lastly, the DNA with 9 *parS* sites exhibited more efficiency of protein binding in the absence of CTP than the presence of CTP in the reaction buffer, which could also be contributed to the sliding of protein from the DNA in the presence of CTP.

Thus, the current work substantiated the previously reported theories of ParB interaction with *parS* DNA that the protein was able to bind DNA in a sequence non-specific manner in the

presence of CTP and that the latter enhanced the sliding of ParB on the DNA molecule [1–3]. It is noteworthy that the results reported in this chapter were consistent with the Light Scattering experiments of these *parS* DNA-ParB protein samples [7].

However, there is a contradiction highlighted in the current work with the previous research. It has been reported that the presence of *parS* site is required for the initial recognition of DNA by ParB, even if the presence of CTP later facilitates the switch to a non-specific binding[2,3]. Yet, it was seen in the AFM experiments of DNA-protein interaction in the presence of CTP in the buffer (section 8.4.2) that there was protein binding to DNA with no *parS* site in the presence of CTP. To resolve this conflict and also to understand the mechanism of the interaction better, further experiments could be performed.

To obtain results with protein binding to the DNA with multiple binding sites varying from 3-8, more types of DNA could be synthesized in order to obtain a more comprehensive data set. Also, there is the argument that some protein molecules would slide off the DNA strands since the DNA used here were short, open and did not have any other protein bound to them to stop them sliding off. Hence, to counter that uncertainty, circular DNA could be synthesized with size ~ 1 kb, probably by circularization of sticky end DNA ~ 1 kb long containing a varied number of *parS* sites. This would prevent the loss of efficiency of protein binding by ParB sliding off the DNA. Secondly, immobilization of DNA on the mica surface and allowing the protein to interact with it on the surface (dynamic imaging) would also avoid the sliding off of the protein as many points on the DNA would be anchored to the surface. This technique would also allow the visualization of the actual binding process.

Better resolution AFM would allow the observation of various motifs of the protein and the conformation changes occurring during the DNA-protein interaction. Attempts could also be made to reduce the image noise to obtain AFM images in reaction buffer, which would be the more suitable environment for the proteins to react with DNA.

Bibliography

1. Jalal ASB, Tran NT, Le TBK. ParB spreading on DNA requires cytidine triphosphate in vitro. Elife. 2020;9:1-24.
2. Wu LJ, Errington J. Coordination of Cell Division and Chromosome Segregation by a Nucleoid Occlusion Protein in Bacillus subtilis. Cell. 2004;117(7):915-925.
3. Jalal ASB, Tran NT, Stevenson CE, *et al*. Diversification of DNA-binding specificity via permissive and specificity-switching mutations in the ParB/Noc protein family. bioRxiv. 2019:724823.
4. James D. Watson *et al*. The Structure of DNA. In: Molecular Biology of the Gene. 7th ed. Pearson; 2014:77-105.
5. Berg J, Tymoczko J, Stryer L. Biochemistry. 2002.
6. Heenan PR, Perkins TT. Imaging DNA Equilibrated onto Mica in Liquid Using Biochemically Relevant Deposition Conditions. ACS Nano. 2019;13(4):4220-4229.
7. Chodha SS. Investigating the mechanistic features of parabs-mediated Vibrio cholerae chromosome 2 segregation.

Chapter 9

Conclusion and Future Perspectives

In the present scenario of the prevalence of pathogen infections and numerous disorders related to improper cell division, it becomes imperative to develop and understand the mechanism of DNA replication and cell division, and their interplay. Contributing to this research, in the current work, attempts have been made to comprehend the process of catalysis by FEN and ParB using AFM. Additionally, an AFM sample immobilization and imaging protocol has been formulated that is suitable for dynamic imaging of DNA-protein interactions, especially in case of immobilizing flap DNA sample with both ssDNA and dsDNA present together in one molecule. Experiments have also been performed to obtain DNA and proteins specially tailored to be imaged using AFM. Finally, the experimental results have been analyzed and quantified by DNA molecule motion tracking and length measurements to examine the change in the conformation of DNA due to interactions with the protein.

9.1. DNA Immobilization

In the current research work, different immobilization methods, as well as protein and DNA modification methods, have been applied to observe the DNA molecules interacting with FEN on modified mica surface. It was observed that PLO treated mica proved to be the most suitable surface for DNA immobilization during imaging in buffer conditions, while experimentation was performed to modify the surface using varied concentrations of divalent cations. However, the PLO method provided some difficulty in the visualization of short ssDNA branching from dsDNA in flap DNA. The imaging was simpler and more efficient if the length of the DNA strands (both the double-stranded section and the ssDNA flap) was greater (*e.g.* ~ 300 bp dsDNA and ~ 100 nt for ssDNA). Longer DNA strands were hence synthesized of length suitable for the flap to immobilize on the PLO surface to allow adequate resolution.

The flap or overhang DNA, when imaged in air, could be seen as strands of very comparable heights. On the contrary, when imaged in liquid, dsDNA molecules were seen as molecules with a much larger height (~ 1.5 nm) (as they were thicker) than the ssDNA branches (~ 500 pm). This made it easy to distinguish the ssDNA from the double-stranded ones. The flap DNA could be seen as Y-shaped molecules while the nicked DNA was seen as a dumbbell-shaped

molecule while imaging in air. The break in the phosphodiester bond formed the neck of the dumbbell, but it relaxed into the strand form during liquid imaging.

The DNA double helix could be resolved when the sample was imaged in HEPES buffer of pH~ 8 and the major grooves could be visualized, especially in the case of long DNA molecules like plasmids. The pitch of the DNA measured from the images was consistent with the B-form of DNA and dehydration of the sample during drying by N_2 gas did not cause the conformation of DNA to convert to A-form, showing that the conformation of DNA depended on the form it was immobilized in and did not change in the subsequent steps of imaging.

When immobilized on divalent cation treated mica, DNA appeared to be shorter than when they were immobilized on PLO treated mica, due to folding of DNA molecules in the z-plane or immobilization at the ends. The DNA molecules did not immobilize at a low concentration of divalent cations and could be seen as blobs or strands that moved around in the consecutive frames. Yet, some instances where the DNA were immobilized well (on Ca^{2+} treated mica), a good helical resolution could be observed where the molecules showed the major grooves of DNA. This was noteworthy as the helical resolution was difficult to resolve for such short DNA fragments of length of 100 bp and such results have not been reported previously. But, even in the images of the best dsDNA resolution, there were no or very scarce ssDNA seen, signifying the immobilization technique was not suitable for flap DNA.

At a high salt concentration, the salt precipitates could be seen as noise on the mica surface, though the DNA immobilized well at such high concentrations. Ca^{2+} and Ni^{2+} proved to be more effective than Mg^{2+} for immobilization of dsDNA, with Ni^{2+} showing the least amount of mobility of DNA in the consecutive frames. Further experiments could be pursued to achieve the optimized ion concentration for immobilization using $Ca^{2+}$for DNA molecules as short as 100 bp, while longer strand length could make the experimental results more feasible.

9.2. DNA Synthesis and Purification as a sample for AFM

The problem with immobilizing 100 bp flap DNA sample on modified mica was that 100 bp dsDNA was either too short to be immobilized and give helical resolution or stay immobilized during the consecutive scans. Additionally, the single-stranded flap was 50 nt long and was difficult to recognize when it was randomly folded or closely aligned along the dsDNA. Even when numerous immobilization methods were attempted by treatment of mica with divalent

cations, the DNA was not sufficiently immobilized (chapter 5). Hence, attempts were made to synthesize longer DNA strands that could be used for imaging on PLO treated mica.

Various methods of DNA synthesis, either flap DNA synthesized by polymerization of M13 ssDNA, annealing asymmetric PCR products to produce overhang DNA or annealing and ligating oligonucleotides to produce flap DNA, were attempted (chapter 6) to obtain molecules of suitable length. In the first case, the M13 DNA was very long causing the dsDNA or flap DNA produced from it to be long as well, hence making it difficult to see any flap at all, if present. The second method produced a large amount of interlinked and unwanted DNA complexes that contributed to noise.

The last method of assembly of long DNA oligonucleotides was the most adequate method of synthesis of flap DNA for high-speed AFM to study DNA-protein interactions. The length of DNA was not too short that it immobilized, nor was it too long that the DNA folded and convoluted to inhibit observation of details of DNA. The length of DNA was also suitable to immobilize the DNA well, such that they did not mobilize during consecutive scans. The other methods of synthesis of flap DNA, like the synthesis of ~ 300 bp long overhang DNA fragments by the assembly of PCR products and polymerization of single-stranded M13 DNA in the absence of exonuclease fragment of DNA Polymerase I (*i.e.* the klenow segment alone) to produce closed dsDNA with a flap, both proved to be ineffective. While in the former case the DNA segments assembled to form unrequired and complex shapes unsuitable for observation of catalysis, the latter produced DNA that coiled on the surface and could not be studied for any conformation changes in the presence of the protein.

9.3. Dynamic Imaging DNA—Protein Interactions

High-resolution AFM has been applied to image DNA Polymerase I and FEN molecules to visualize their different conformations. Mostly the molecules appeared as blobs, but on closer inspection, images of DNAPolI revealed somewhat elongated structures folded together. In some very rare cases, the protein appeared in an extended form where the domains were clearly visible: the elongated sphere-shaped FEN domain and the Klenow domain with a hand, thumb and finger shape.

In the current research work, it was possible to image FEN interacting with flap or overhang DNA, causing the latter to undergo different conformation changes. A method of motion tracking and quantification of the movement of sections of DNA was devised using Adobe

Illustrator, Tracker (Video Analysis and Modelling Tool) and MS Excel. It allowed the segmentation of DNA and tracking of each of the segments to recognise the sections of DNA most affected by the protein.

The binding of inactive FEN domain-containing DNA Polymerase I caused the mobilisation of DNA on the mica surface, indicating that the protein recognition and interaction affected the DNA conformation. The protein caused the DNA to bend in the vicinity and time duration of the binding, and the changes in DNA shape could be observed in the short time span of the protein's appearance close to DNA. It was observed that FEN bound to DNA (both flap and overhang) randomly anywhere, and not at the branch point directly as previously proposed based on crystallography data. The binding of the protein caused the DNA to overcome the electrostatic force of attraction binding it to the PLO treated surface, thereby indicating that the protein disrupted the charges around DNA to cause it to be attracted towards the protein. FEN was also seen sliding on the DNA and consequently bending it along its way. Based on these observations, it can be proposed that DNA Polymerase I containing FEN bind to the DNA randomly and slide on it until it encounters the branch point of ssDNA to which it anchors leading to catalysis.

Active FEN protein was observed intermittently mobilizing the DNA, threading it and cutting the single-stranded branch of DNA. Motion tracking of segments of DNA showed that most of the movement occurred during the short time span when the protein encountered the DNA and was restricted to the sections around the protein. It causes the single strand to bend and thread through the arch of the active site, though the resolution of the images was not sufficient to observe the mechanism. However, it was interesting to observe that the protein generated an attractive force on DNA that was stronger than the force binding the DNA to PLO treated mica and caused the DNA to mobilize and bend towards the protein. The DNA then anchored to the surface again after a few frames indicating that the effect of the protein causing DNA mobility had been removed.

As a result of the dynamic AFM imaging results obtained in the current work, the Threading model of DNA substrate recognition by FEN (that the protein bound to the branch point of DNA to perform the nuclease activity) can be corroborated with the following additional steps in the recognition mechanism: attraction and binding of the negatively charged DNA anywhere along its length by the surface residues of FEN (and DNA Polymerase I, since FEN does not occur as isolated domain in the cell) and the sliding of the protein along the length of DNA,

irrespective of the position it bound to on the DNA. Also, the nuclease activity is the faster step in the catalysis and the enzyme activity is limited by the encounter of the branch point, either randomly in the cell, or by random binding to DNA followed by sliding to reach the branch point.

Future experiments could be performed with improved spatial and time resolution AFM to observe the conformation changes and the mechanism of catalysis, *e.g.* the modification of the structure of FEN active site during the reaction and the changes on the surface residues of FEN that allow it to bend the ssDNA to thread through. The importance of various mutations in the protein and their consequences in disorders like cancer could be studied by dynamic AFM imaging of the mutants interacting with DNA. These experiments could be expanded to understand not only the functioning of FEN, but the entire DNA Polymerase I in the highly regulated process of DNA replication and cell division.

9.4. ParB-parS Interactions

From the results reported in this work, it has been found that DNA plasmid can be visualized using AFM to reveal helical resolution. When the molecule was immobilized on the surface in coiled or twisted conformation, protein binding to the DNA and the associated conformational changes could not be observed clearly. Attempts were made to allow the DNA to spread onto the surface by gently transferring the DNA from the solution to the surface without modifying the buffer composition. However, this method (Heenan and Perkins, 2019) did not allow the spreading of DNA of about 3 kb used in our case, as the method was proposed for DNA of lengths 200-2000 bp.

To overcome this limitation of DNA convolution, short open DNA were synthesized that contained 0-9 *parS* sites and were about 200-455 bp long. These were immobilized on PLO treated mica surface to obtain extended DNA. When the ParB protein was bound to DNA, the molecules displayed a ‘beadiness’ and the length of DNA was reduced in comparison with the length of the DNA negative control. DNA with 0 *parS* sites did not show any reduction in the DNA length, signifying no protein binding, while the DNA with 1-9 *parS* sites displayed shortening of DNA, with the maximum effect observed in case of 1 DBS, then 9 and 2 DBS. On the contrary, in the presence of CTP in the reaction buffer, the protein binding (indicated by the percentage reductions in length in the presence of ParB) was almost equal and the highest in case of 0 and 1 *parS* site containing DNA and decreased for 2 and 9 site DNA. This

signified that the presence of CTP in the buffer caused the protein to bind to DNA in a sequence non-specific manner, as there was no binding in case of no binding sites in the absence of CTP in the buffer. Additionally, because the percentage reduction in lengths for DNA with a varied number of binding sites was lower in the presence of CTP than when CTP was absent, it is proposed that CTP caused the protein to slide off the DNA. These observations were consistent with the previous works, yet how ParB recognised the DNA in the first place in the absence of any *parS* site, could not be known.

To improve the results of the study, further experiments could be performed to obtain results of protein binding to the DNA with multiple binding sites, with the aim to obtain a more comprehensive data set. Additionally, to prevent the possible sliding of protein molecules off the short, open and uncapped DNA strands, circular DNA could be synthesized with a size ~ 1 kb. This could be done by circularization of sticky end DNA ~ 1 kb long containing a variety of *parS* sites and would prevent the decrease in protein binding efficiency by ParB sliding off the DNA. Else, experiments could be performed with DNA immobilization on DNA-origami tiles or modified mica surface before any protein interaction so as to block the ends and avoid the sliding off of the protein. Furthermore, to understand the interaction better, dynamic imaging of ParB-*parS* DNA interaction could be performed using high-speed AFM.

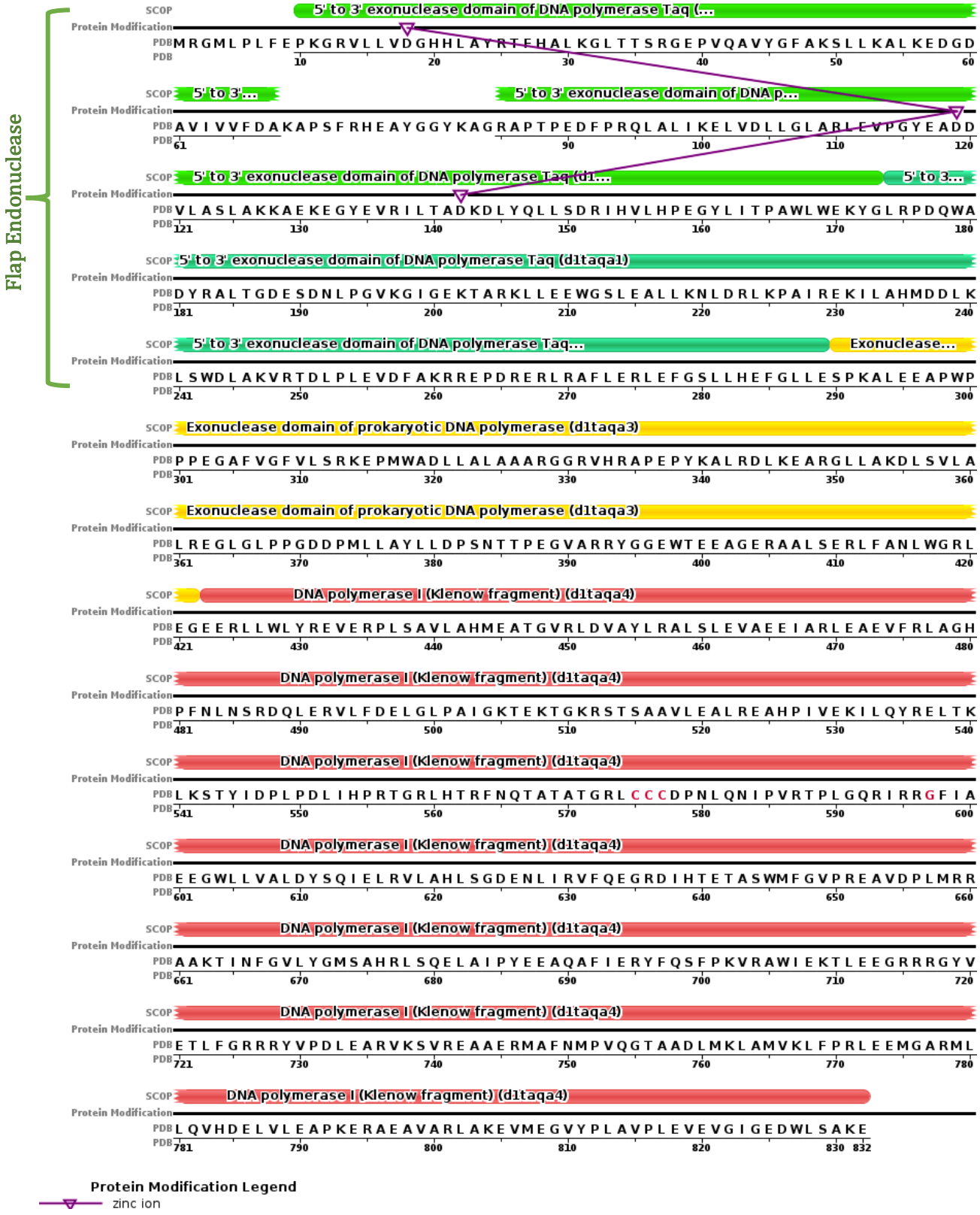
9.5. General Discussion

In summary of the results of this experimental work, it can be concluded that DNA can be successfully immobilized on PLO treated surface for imaging protein-DNA interactions without mobilizing during the scans and allowing good resolution. Flap endonuclease binding and threading on DNA, bending the DNA and nuclease activity can be observed on flap/overhang DNA. Analysis of the conformation changes due to protein interaction has been performed by DNA tracking and measurements of length. FEN catalysis caused the movement of DNA while ParB binding shortened the DNA immobilized on the surface. The variation in the binding of ParB to DNA with a varied number of *parS* sites in buffer with (out) CTP was examined. It was noted that CTP enhanced the sequence non-specific binding and sliding of ParB on DNA, though the mechanism could not be visualized.

Future experimental possibilities should be explored with purer DNA and protein samples to reduce the noise. Further research is also required for imaging ParB-*parS* interaction by experimentation with DNA containing a more extensive variation of *parS* sites and possible

attempts at dynamic imaging of the interaction. Developments in AFM to obtain better resolution and speed of imaging, especially with soft biological samples could allow ground-breaking advances in the understanding of biochemical processes.

Appendices



Appendix 1: Amino acid sequence of *Thermus aquaticus* DNA Polymerase I reproduced from Protein Data Bank.

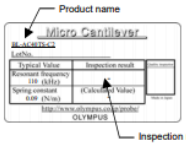
Appendix 2: AFM cantilever AC-40 or Biolever Mini.

Micro cantilever

Product name

BL-AC40TS-C2 (Bio-Lever mini)

Silicon nitride cantilever with silicon tetrahedral tip



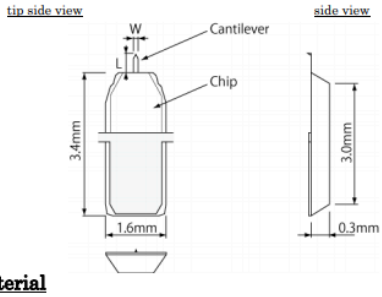
BL - AC 40 T S – C 2

BL : Bio Lever
AC : main application is AC mode measurement
40 : Lever length of 40 μm
T : Tetrahedral tip
S : Gold/Chromium reflex coating (Single side)
C : 24 chips / unit
2 : Chip thickness 0.3 mm

Chip

There is a rectangular cantilever on one side of the silicon chip.

Dimension

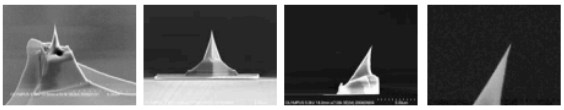


Material

Tip	Silicon (4 - 6 ohm.cm)
Cantilever	Silicon nitride
Metal coating (tip side)	No coating
Metal coating (back side)	Gold/Chromium
Chip	Silicon (4 - 6 ohm.cm)

Tip

A sharpened tetrahedral silicon tip extends from the end of the cantilever. The base of the tip is covered in silicon nitride which attaches it to the cantilever. The effective tip height is the tip apex which is not covered by the silicon nitride.



Perspective View | Front | Side | Magnified Side View

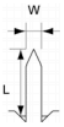
Dimensions

	Typical value	Typical range
Tip height (μm)	7	5 - 9
Effective tip height (μm)	3.5	2 - 5
Tip radius (nm)	8	less than 15
Tip angle (deg.)	(side) less than 35 (front) less than 35	

Cantilever

Dimensions

Cantilever length L (μm)	38 (±3)
Cantilever width W (μm)	16 (±2)
Cantilever thickness t (μm)	0.2 (±0.02)
Thickness of Metal Coat tm (μm)	0.04 (±0.01)



Nominal mechanical properties

	Typical value	Typical range
Resonant frequency in air (kHz)	110	75 – 145
Resonant frequency in water (kHz)	25	17 – 45
Spring constant (N/m)	0.09	0.02 – 0.14

OLYMPUS

BL-AC40TS-C2

Ver. 2.0 June 22, 2009

Appendix 3: AFM cantilever Nunano SCOUT 350 RAI datasheet.

nanO

A Nu approach to AFM Probes

PROBE MODEL: SCOUT 350

The **SCOUT 350** is our general purpose probe for imaging in AC modes (non-contact/tapping), exhibiting exemplary dimensional tolerances and tip sharpness, characteristic of all our AFM probes.

APPLICATIONS

Non-contact and tapping modes in air.

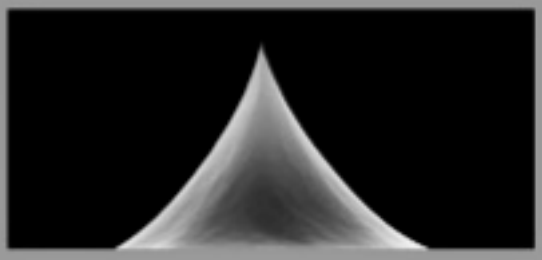
Cantilever Specifications

Parameter	Nominal	Range
Spring constant (N/m)	42	25 – 70
Resonant frequency (kHz)	350	300 – 400
Shape	Rectangular	
Length (µm)	125	123 – 127
Width (µm)	30	29 – 31
Thickness (µm)	4.5	4.0 – 5.0
Material	Silicon (n-type, antimony)	
Resistivity (Ωcm)	0.02	0.015 – 0.025

Tip Specifications

Parameter	Nominal	Range
Radius (nm)	5	< 10
Height (µm)	6	5 – 8
Set back (µm)	7.5	6.5 – 8.5
Shape	Conical	
Cone angle (°)	25	15 - 40
Material	Silicon (n-type, antimony)	
Resistivity (Ωcm)	0.02	0.015 – 0.025

NB. Nominal values for spring constants and resonant frequencies are calculated using well known formulae and based on expected probe dimensions. The ranges are calculated using measured dimensional variations.



Chip Specifications

Length (mm)	3.4
Width (mm)	1.6
Thickness (mm)	0.3
Cantilever anchor	Yes
Alignment grooves	No

Coatings

Frontside / Tip	none
Backside / Reflective	none

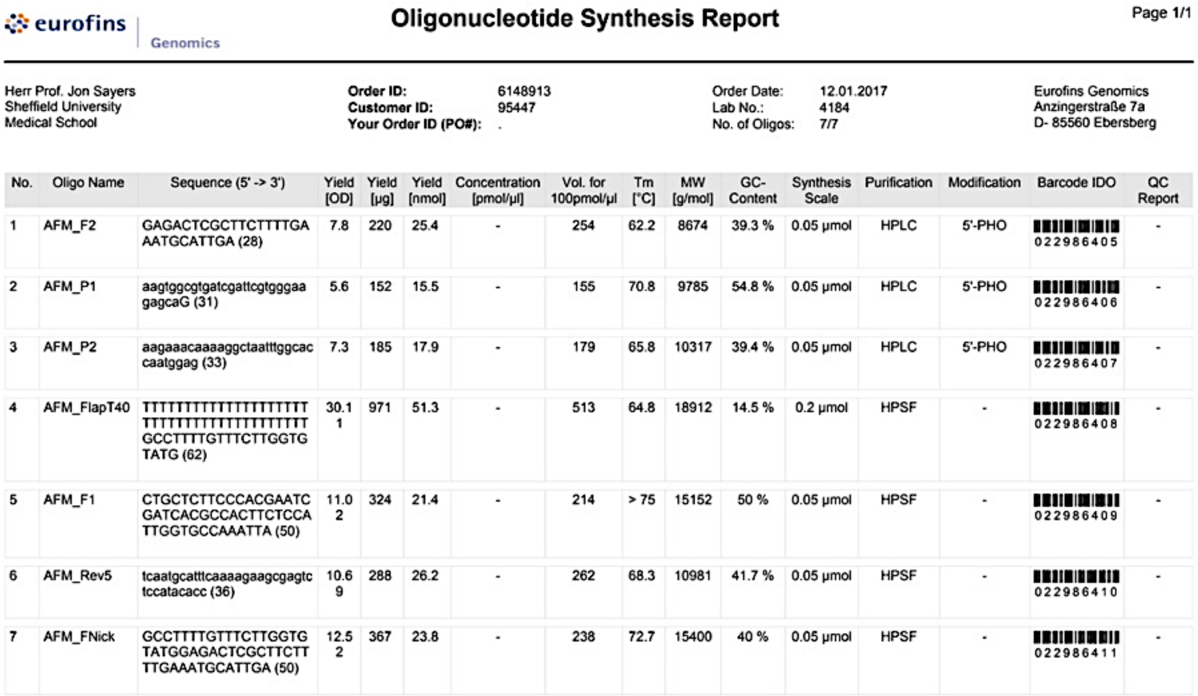
Appendix 4: Sequence of synthetic oligonucleotides to make 100 bp flap and nicked DNA.

T40 tail T
T
T
5' CT GCT CT TCC CACGAATCGATCACGCCAC T T CTCCATTGGTGCCAAATTA GCCT TTTGT TT C TTGGTGTATGGAGACTCGCTTCTTTTGAAATGCATTGA
3' GACGAGAAGGGTG CTTAGCTAGTGCGGTGAAGAGGTAACCACGGTTTAAT CGGAAAACAAAGAACCACA
TACCTCTGAGCGAAGAAAACTTTACGTAACT

1. CTGCTCTTCCCACGAATCGATCACGCCACTTCTCCATTGGTGCCAAATTA
2. TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGCCTTTTGTTTCTTGGTGTATG (phos?)
3. 5’-P-GAGACTCGCTTCTTTTGAAATGCATTGA
4. GACGAGAAGGGTGCTTAGCTAGTGCGGTGAA-P-5’
5. GAGGTAACCACGGTTTAATCGGAAAACAAAGAA-P-5’
6. CCACATACCTCTGAGCGAAGAAAACTTTACGTAACT
7. GCCTTTTGTTT C TT GGTGTATG GAGACTCGCTTCTTTTGAAATGCATTGA   nick

   P=pre-phosphorylated oligo

Appendix 5:Oligonucleotide Synthesis Report for 100 bp flap and nick DNA sample.



eurofins | Genomics

Oligonucleotide Synthesis Report

Page 1/1

Herr Prof. Jon Sayers
Sheffield University
Medical School

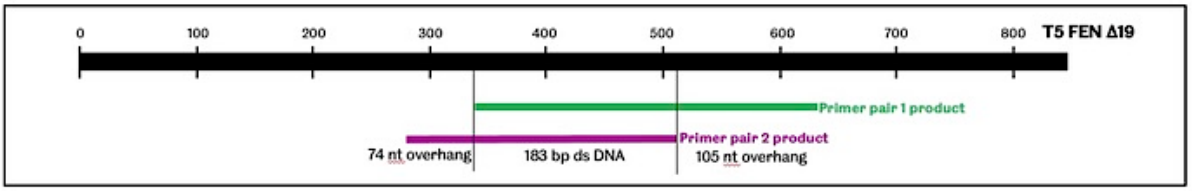
Order ID: 6148913
Customer ID: 95447
Your Order ID (PO#): .

Order Date: 12.01.2017
Lab No.: 4184
No. of Oligos: 7/7

Eurofins Genomics
Anzingerstraße 7a
D- 85560 Ebersberg

No.	Oligo Name	Sequence (5' -> 3')	Yield [OD]	Yield [µg]	Yield [nmol]	Concentration [pmol/µl]	Vol. for 100pmol/µl	Tm [°C]	MW [g/mol]	GC-Content	Synthesis Scale	Purification	Modification	Barcode IDO	QC Report
1	AFM_F2	GAGACTCGCTTCTTTTGA AATGCATTGA (28)	7.8	220	25.4	-	254	62.2	8674	39.3 %	0.05 µmol	HPLC	5'-PHO	022986405	-
2	AFM_P1	aagtggcgtgatcgattcgtgggaa gagcaG (31)	5.6	152	15.5	-	155	70.8	9785	54.8 %	0.05 µmol	HPLC	5'-PHO	022986406	-
3	AFM_P2	aagaaacaaaaggctaatttggcac caatggag (33)	7.3	185	17.9	-	179	65.8	10317	39.4 %	0.05 µmol	HPLC	5'-PHO	022986407	-
4	AFM_FlapT40	TTTTTTTTTTTTTTTTTTTT TTTTTTTTTTTTTTTTTTTT GCCTTTTGTTTCTTGGTG TATG (62)	30.1 1	971	51.3	-	513	64.8	18912	14.5 %	0.2 µmol	HPSF	-	022986408	-
5	AFM_F1	CTGCTCTTCCCACGAATC GATCACGCCACTTCTCCA TTGGTGCCAAATTA (50)	11.0 2	324	21.4	-	214	> 75	15152	50 %	0.05 µmol	HPSF	-	022986409	-
6	AFM_Rev5	tcaatgcatttcaaaagaagcgagtc tccatacacc (36)	10.6 9	288	26.2	-	262	68.3	10981	41.7 %	0.05 µmol	HPSF	-	022986410	-
7	AFM_FNick	GCCTTTTGTTTCTTGGTG TATGGAGACTCGCTTCTT TTGAAATGCATTGA (50)	12.5 2	367	23.8	-	238	72.7	15400	40 %	0.05 µmol	HPSF	-	022986411	-

Appendix 6: Details of the primers for preparation of overhang DNA using PCR.



AATTTATTGAAGAAGAGGAAGCTGAAATGGCTTCCCGTCGTAATCTAATGATTGTCGATGGAACTAACTTAGGCTTTCGCTTCAAACATAACAATAGTAAAAAACCATT
TGCCTCAAGTTATGTTTCAACTATTCAATCTCTGGCAAAATCCTACTCTGCCAGAACTACGATTGTTCTAGGTGATAAGGGAAAATCTGTATTTCGTCTAGAACATCT
ACCAGAGTATAAAGGTAATCGTGATGAAAAGTACGCACAACGTACGGAAGAGGAGAAAGCGCTAGATGAGCAGTTCTTTGAGTATTTGAAGGATGCTTTCGAGTTG
TGTAAAACTACATTCCCAACTTTTACCATTCGTGGTGTAGAAGCAGACGATATGGCAGCTTATATTGTTAAGCTCATCGGGCATCTTTATGATCACGTTTGGCTAA
TATCTACAGATGGTGACTGGGATACTTTATTAACGGATAAAGTTTCTCGTTTTTCTTTCACAACACGTCGTGAGTATCATCTTCGTGATATGTATGAACATCATAAT
GTTGATGATGTTGAGCAGTTTATCTCCCTGAAAGCAATTATGGGAGATCTAGGAGATAATATTCGTGGTGTTGAAGGAATAGGAGCAAAACGCGGATATAATATTA
TTCGTGAGTTTGGTAACGTACTGGATATTATTGATCAGCTTCCACTGCCTGGAAAGCAGAAATATATACAGAACCTGAATGCATCGGAAGAACTGCTTTTCCGAAAC
TTGATTCTGGTTGATTTACCTACCTACTGTGTGGATGCTATTGCTGCTGTAGGTCAAGATGTGTTAGATAAGTTTACAAAAGATATTTTGGAGATTGCAGAACAATGA

Primer pair 1: length 288

	Sequence (5'-3')	Template strand	Length	Start	Tm	GC%
Forward primer	CAT TCC CAA CTT TTA CCA TTC G	Plus	22	334	50.2	40.9
Reverse primer	CTC CTA TTC CTT CAA CAC CAC	Minus	21	621	50.2	47.6

Primer pair 2: length 257

	Sequence (5'-3')	Template strand	Length	Start	Tm	GC%
Forward primer	TACGGAAGAGGAGAAAGCG	Plus	19	260	50.9	52.6
Reverse primer	CACGAAGATGATACTCACGAC	Minus	21	516	50.7	47.6

Appendix 7: Sequence of synthetic oligonucleotides to make 300 bp flap DNA.

T40 tail T
T
T

5'ATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGC
GGGTGTTGGCGGGTGTCGGGGCTGGCTTCTGCTCTTCCCACGAATCGATCACGCCACTTCTCCATTGGTGCCAAATTAGCCTTTTGTTTCTTGGTGTATGGAGAC
TCGCTTCTTTTGAAATGCATTGATCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGA3'

3'TAACATGACTCTCACGTGGTATACGCCACACTTTATGGCGTGTCTACGCATTCCTCTTTCGCCTACGGCCCTCGTCTGTTCGGGCAGTCCCGCGCAGTCGCCCAC
AACCGCCCACAGCCCCGACCGAAGACGAGAAGGGTGCTTAGCTAGTGCGGTGAAGAGGTAACCACGGTTTAATCGGAAAACAAAGAACCACATACCTCTGAGC
GAAGAAAACTTTACGTAACTAGCGCGCAAAGCCACTACTGCCACTTTTGGAGACTGTGTACGTCGAGGGCCT 5'

1. 5’-P-CGGGTGTTGGCGGGTGTCGGGGCTGGCTTCTGCTCTTCCCACGAATCGATCACGCCACTTCTCCATTGGTGCCAAATTA
2. TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGCCTTTTGTTTCTTGGTGTATG (PHOS?)
3. 5’-P-GAGACTCGCTTCTTTTGAAATGCATTGATCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGA
4. TCGGGCAGTCCCGCGCAGTCGCCCACAACCGCCCACAGCCCCGACCGAAGACGAGAAGGGTGCTTAGCTAGTGCGGTGAA-P-5’
5. ~~GAGGTAACCACGGTTTAATCGGAAAACAAAGAA-P-5’~~
6. CCACATACCTCTGAGCGAAGAAAACTTTACGTAACTAGCGCGCAAAGCCACTACTGCCACTTTTGGAGACTGTGTACGTCGAGGGCCT-5’
7. ~~GCCTTTTGTTT C TT GGTGTATG GAGACTCGCTTCTTTTGAAATGCATTGA nick~~

P=pre-phosphorylated oligo

8. ATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAG
9. TAACATGACTCTCACGTGGTATACGCCACACTTTATGGCGTGTCTACGCATTCCTCTTTCGCCTACGGCCCTCGTCTGT-P-5’

Total Length=283 bp with 100-T Flap

Appendix 8: Purification of *Thermus aquaticus* FEN domain.

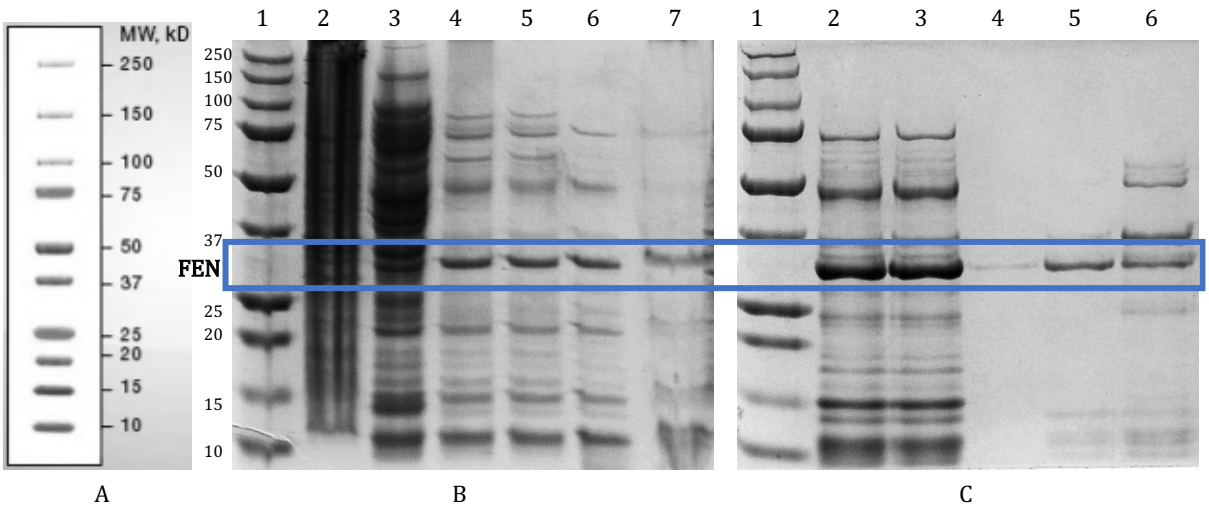


Figure 1: SDS PAGE images for the purification of *Thermus aquaticus* FEN domain: A. Precision Plus Protein™: Protein ladder 10 kDa to 250 kDa. B. Steps for extraction of the protein from the cells: lane 1-ladder. 2-after sonication. 3-before pasteurization. 4-after pasteurization. 5-before adding PEI. 6-after adding PEI. 7-After adding $(NH_4)_2SO_4$. FEN has a molecular weight of about 35 kDa and separates out between the 25 and 37 kDa bands on the protein ladder. C. Purification of the protein on the Q column: lane 1-ladder, 2-before dialysis, 3-load for Q column, 4-6: elution fractions 7-9.

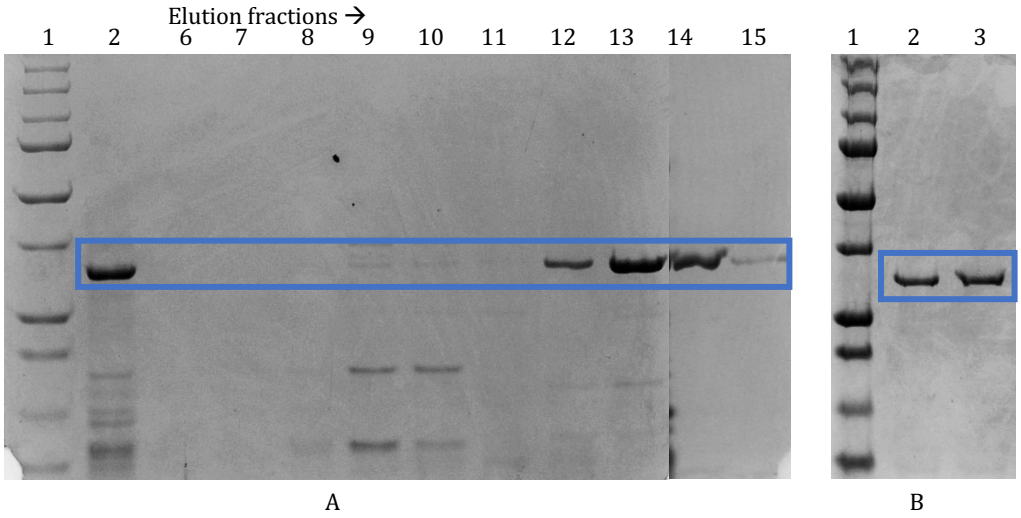


Figure 2: SDS PAGE images for the purification of *Thermus aquaticus* FEN domain: A. Purification of the protein on the Heparin column: lane 1- ladder, 2-load, elution fractions 6 to 15. The protein fractions 12 and 14 were combined and further purified in the heparin column. The fraction 13 was separately purified in the SP and Heparin column and labelled as protein sample iii. B. Purification of the protein on the SP column: fractions 12 and 14 combined were loaded. Lane 1-ladder, 2-load, 3-flow-through. The protein didn't stick onto the SP column and eluted out as the flowthrough, while the impurities stuck onto the SP column.

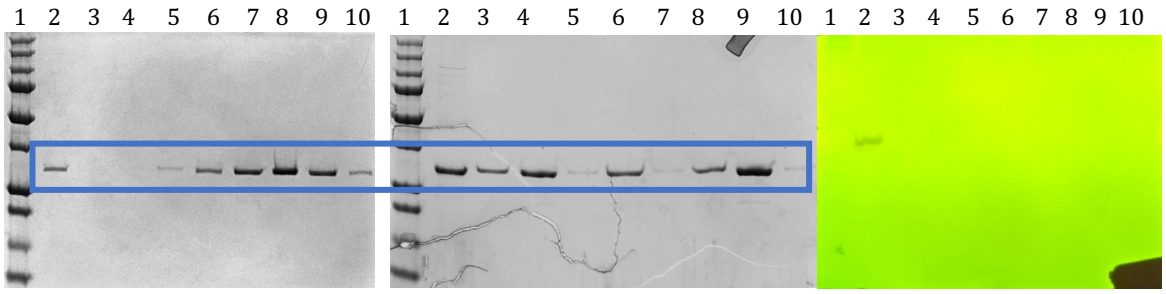


Figure 3: SDS PAGE images for the purification of *Thermus aquaticus* FEN domain: A. Further purification of the protein on the Heparin column: lane 1- ladder, 2-load, 3-7-elution fraction fractions 11 to 15 (5µl), 8-elution fraction 15 (8 µl), 9-elution fraction 16, 10-elution fraction 17. The elution fractions 15 and 16 combined to form protein sample i and fractions 14 and 17 combined to form protein sample ii.
The protein samples i, ii, and iii were obtained as described above and concentrated to ~200µl volume by centrifugation. B. SDS PAGE corresponding to zymogram. C. zymogram: lane 1-ladder, 2-positive control (T_7 FEN) 3µg, 3-load for first heparin column 5µl, 4-first protein sample 3 µg, 5-filtrate of first protein sample 5 µl, 6-second protein sample 3 µg, 7-filtrate of second protein sample 5 µl, 8-load for second heparin column 5 µl, 9-third protein sample 3 µg, 10-filtrate of third protein sample 5 µl. From the zymogram, it was concluded that there is no nuclease activity in the protein sample.

Time (Min)	Positive control	Negative control	Protein aliquot i	Protein aliquot ii	Protein aliquot iii
	Absorbance at 260 nm				
0	0	0	0	0	0
6	0.066	0.017	0.031	0.02	0.022
15	0.133	0.019	0.011	0.019	0.036
24	0.184	0.03	0.02	0.037	0.033
36	0.211	0.024	0.015	0.016	0.018
50	0.219	0.015	0.015	0.017	0.012

Table 1: UV Nuclease Assay.

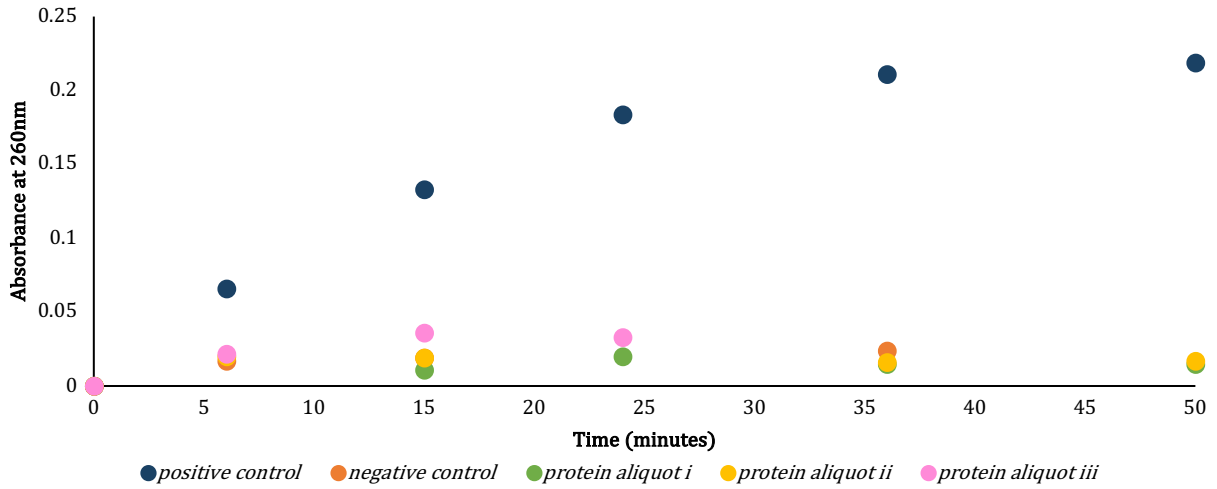


Figure 4: UV nuclease assay for the Taq FEN domain: The positive control is the T_7 nuclease. The negative control is water. These results supplement the zymogram results, that there is no nuclease activity in the FEN sample.

Protein aliquots	Absorbance at 280 nm	Concentration (mg/mL)
i	0.180	1.80
ii	0.028	0.28
iii	0.270	2.70

Table 2: Estimation of concentration of the protein samples using spectrophotometer.

Appendix 9: Purification of *Thermus aquaticus* DNA Polymerase I.

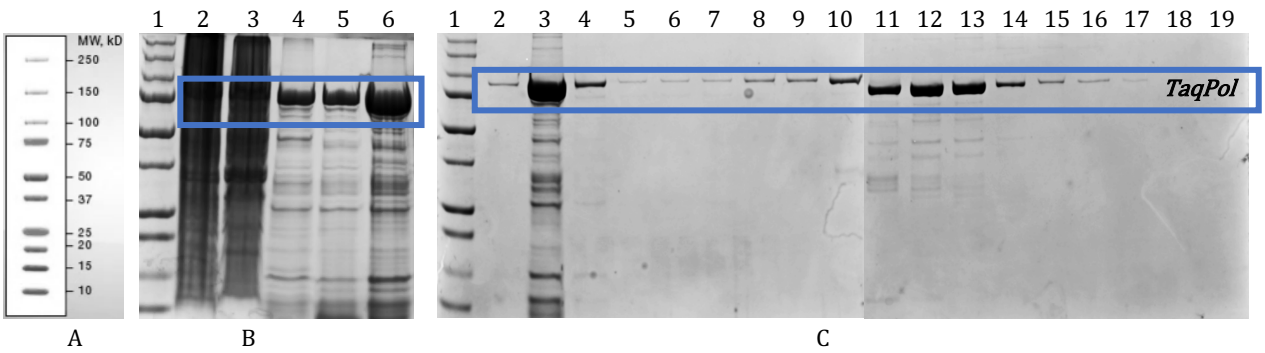


Figure 5: SDS PAGE images for the purification of *Thermus aquaticus* DNA Polymerase I: A. Precision Plus Protein™: Protein ladder 10 kDa to 250 kDa. B. Steps for extraction of the protein from the cells: lane 1-ladder. 2-after sonication. 3-before pasteurization. 4-after pasteurization. 5-after PEI precipitation. 6- before dialysis, after ammonium sulphate precipitate dissolution. *Taq Polymerase* has a molecular weight of about 94 kDa and separates out near the 100 kDa band on the protein ladder. C. Purification of protein on the heparin column: The sample was purified on the heparin column twice. Lane 1-ladder. 2-dialysis buffer (showing some amount of protein leaked into the dialysis buffer), 3-load for 1st heparin column, 4-load for 2nd heparin column, 5-flowthrough, 6-wash with low salt buffer, 7-wash with high salt buffer, lanes 8 to19- fractions 8 to 19. The fractions 10 to 14 were combined and purified further through Q column.

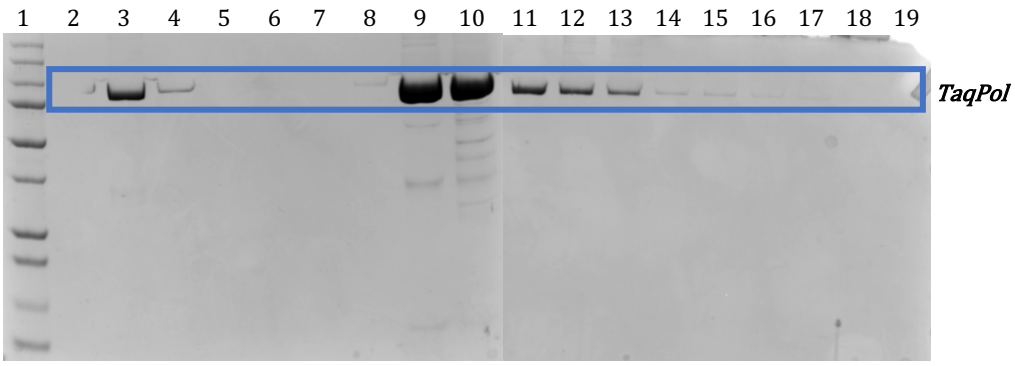


Figure 6: SDS PAGE images for the purification of *Thermus aquaticus* DNA Polymerase I: Purification of fractions 10-14 on the Q column: lane 1-ladder. 2-dialysis buffer, 3-load for Q column, 4-flowthrough, lanes 5 to 19- elution fractions 5 to 19 respectively. Fractions 9 and 10 have a lot of protein compared to 11-13, hence passed the fractions 9 and 10 separately through SP column for further purification.

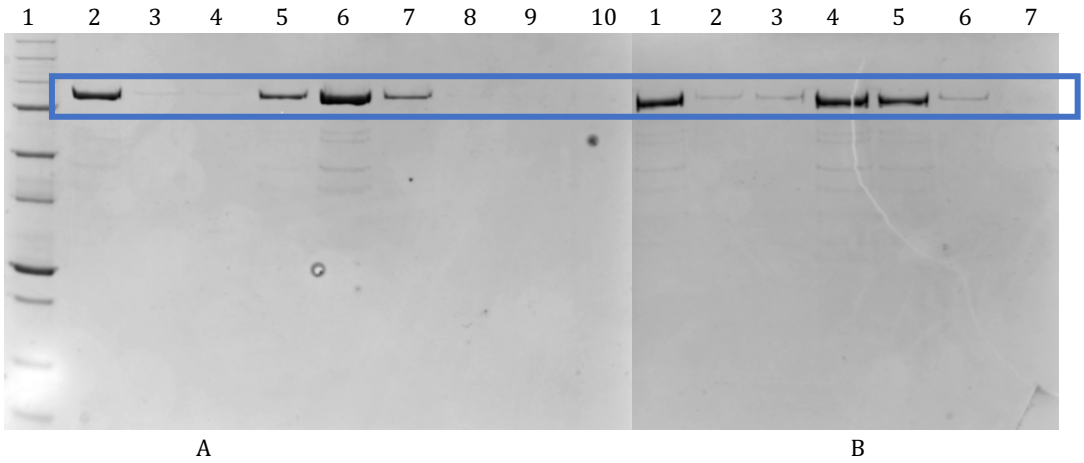


Figure 7: SDS PAGE images for the purification of *Thermus aquaticus* DNA Polymerase I: A. Purification of previous step fraction 10 on the SP column: lane 1-ladder. 2-load, 3- flowthrough, lanes 4 to 10- elution fractions 3-9 respectively. B. Purification of the previous step fraction 9 on the SP column: lane 1-load for SP column, 2-flowthrough, lanes 3-to 7- elution fractions 3-7 respectively. For the next step of purification by pasteurisation, labelled fraction 5 of figure A above as i, fraction 4 and 6 (lanes 5 and 7) combined as ii, fractions 4 and 5 of figure B above combined as iii and fractions 11 to 13 of figure 6 as iv.

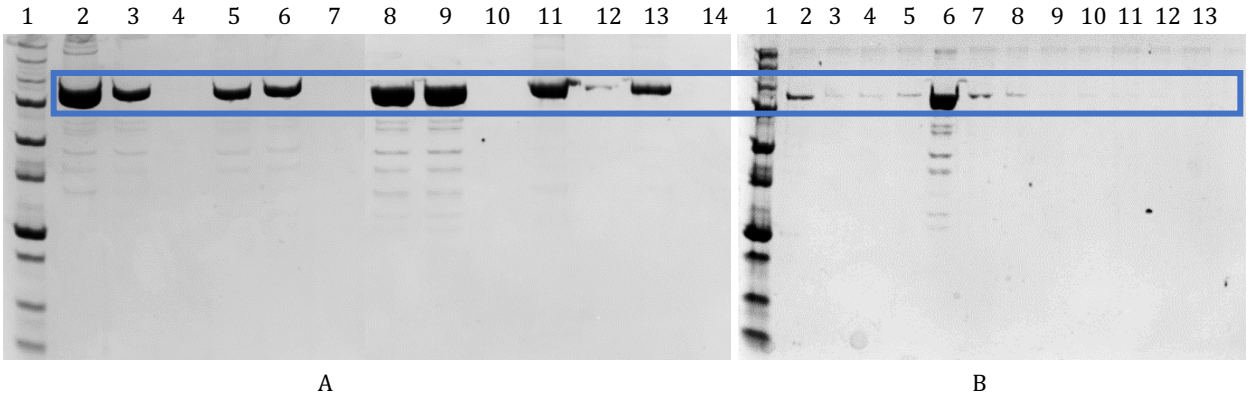


Figure 8: SDS PAGE images for the purification of *Thermus aquaticus* DNA Polymerase I:
A. Purification of the protein by pasteurisation: lane 1-ladder, 2-elution fraction i, 3-elution fraction i after 30 minute pasteurisation, 4-filtrate for i after centrifugation, 5-elution fraction ii, 6-Elution fraction ii after 30 minute pasteurisation, 7-Filtrate for ii after centrifugation, 8-Elution fraction iii, 9-Elution fraction iii after 30 minute pasteurisation, 10-Filtrate for iii after centrifugation, 11-Elution fraction iv, 12-nothing, 13- Elution fraction iv after 30 minute pasteurisation, 14- Filtrate for iv after centrifugation.
B. Purification of the protein by $(NH_4)_2SO_4$ precipitation: the previous step didn't turn very effective to separate out the impurities. Protein fractions i to iv combined and purified together. Lane 1-ladder, 2-Supernatant after 1 M $(NH_4)_2SO_4$ precipitation, 3-Precipitate after 1 M $(NH_4)_2SO_4$ precipitation dissolved in HEPES (no salt) buffer, 4-Supernatant after 2 M $(NH_4)_2SO_4$ precipitation, 5-Supernatant after 2 M $(NH_4)_2SO_4$ precipitation (10µL), 6-Precipitate after 2 M $(NH_4)_2SO_4$ precipitation dissolved in buffer, 7-Supernatant after 2.5 M $(NH_4)_2SO_4$ precipitation, 8-Precipitate after 2.5 M $(NH_4)_2SO_4$ precipitation dissolved in buffer, 9-Supernatant after 3M $(NH_4)_2SO_4$ precipitation, 10-Precipitate after 3 M $(NH_4)_2SO_4$ precipitation dissolved in buffer, 11-Supernatant after 3.5 M $(NH_4)_2SO_4$ precipitation, 12-Precipitate after 3.5 M $(NH_4)_2SO_4$ precipitation dissolved

in buffer, 13-Supernatant after 4 M $(NH_4)_2SO_4$ precipitation, 14-Precipitate after 4 M $(NH_4)_2SO_4$ precipitation dissolved in buffer. The gel images show that the protein precipitated with 2 M $(NH_4)_2SO_4$.

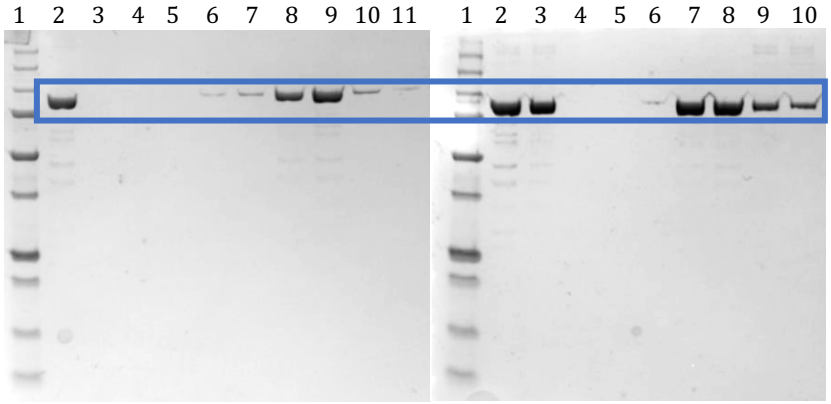


Figure 9: SDS PAGE images for the purification of *Thermus aquaticus* DNA Polymerase I: Purification of dissolved precipitate of 2 M $(NH_4)_2SO_4$: A. One half by SP column at a pH 6: Lane 1-ladder, lane 2- load, lane 3-dialysis buffer, lane 4- flowthrough, lanes 5 to 11- elution fractions 6 to 12 respectively. B. Second half by Q column at pH 9: Lane 1-ladder, lane 2- protein before dialysis, 3-load, lane 4-dialysis buffer, 5-flowthrough, lanes 6 to 10- elution fractions 4 to 8 respectively. Fractions 8 and 9 (lanes 7 and 8) of figure A were combined to form i, fractions 5 to 8 (lanes 7-10) of figure B into ii and fraction 10 (lane 9) of figure A as iii.

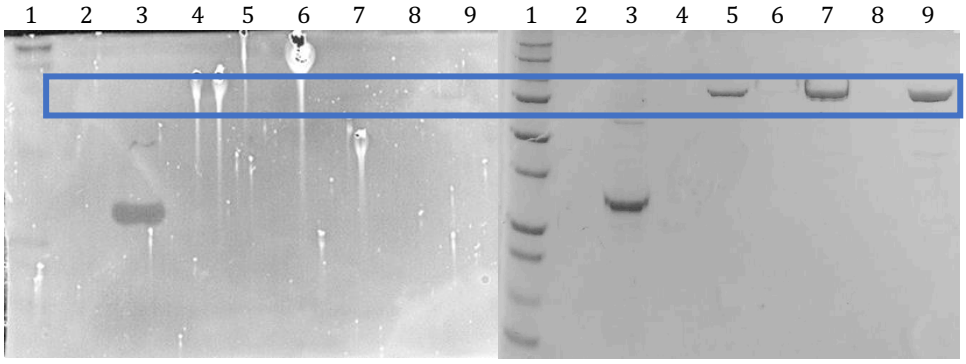


Figure 10: SDS PAGE images for the purification of *Thermus aquaticus* DNA Polymerase I: Protein sample i: fractions 8 and 9 of figure 9A, Sample ii: fractions 5 to 8 of figure 9B, Sample iii: fraction 10 of figure 9A.
A. Zymogram for Taq Polymerase I, B. corresponding SDS-PAGE: Both A and B: lane 1- ladder, lane 3- positive control (T_7 nuclease), lane 5-Protein fraction i, lane 7-Protein fraction ii, lane 9-Protein fraction iii. From the zymogram, it was visible that the protein samples did not have nuclease activity due to the inactive FEN domain.

Time (Min)	Positive Control	Time (Min)	Negative Control	Time (Min)	Protein aliquot i	Time (Min)	Protein aliquot ii	Time (Min)	Protein aliquot iii
0	0	0	0	0	0	0	0	0	0
17.67	0.482	15.07	0.057	12	0.057	11.7	0.059	9.75	0.057
29.67	0.514	20.15	0.066	17.62	0.062	15.95	0.062	13.67	0.066
35.58	0.524	27.52	0.069	25.07	0.072	30.13	0.068	21.77	0.072
45.5	0.527	43.15	0.067	40.34	0.071	38.28	0.066	37.62	0.071
		49.05	0.063			45.37	0.073	44.29	0.077

Table 3: UV Nuclease assay data: absorbance at 260 nm for each of the sample at different times in minutes. The positive control is the T_7 nuclease. The negative control is water.

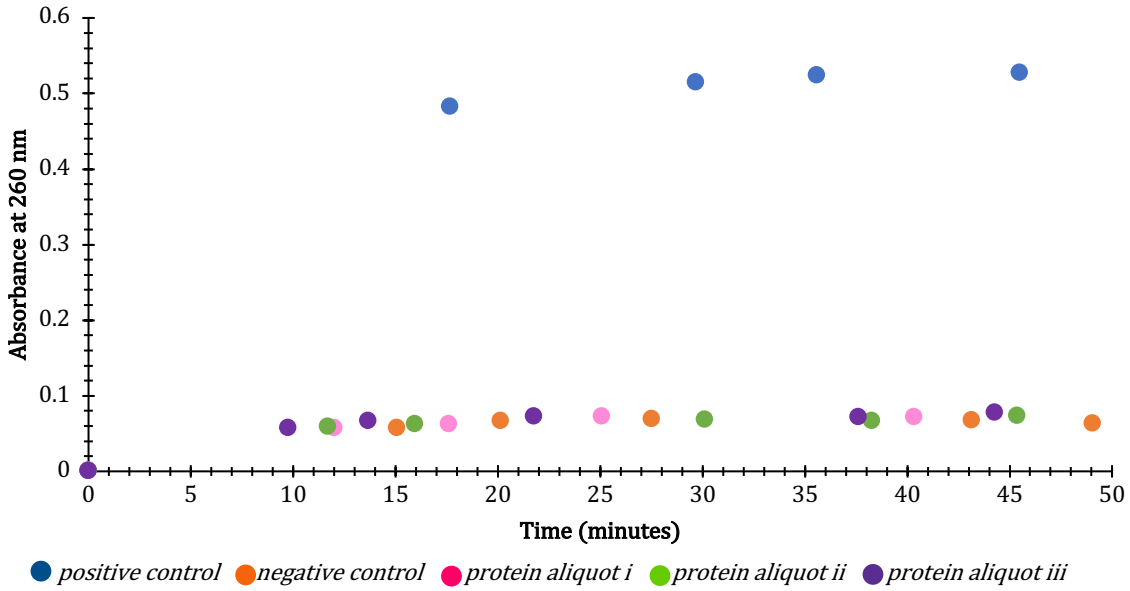
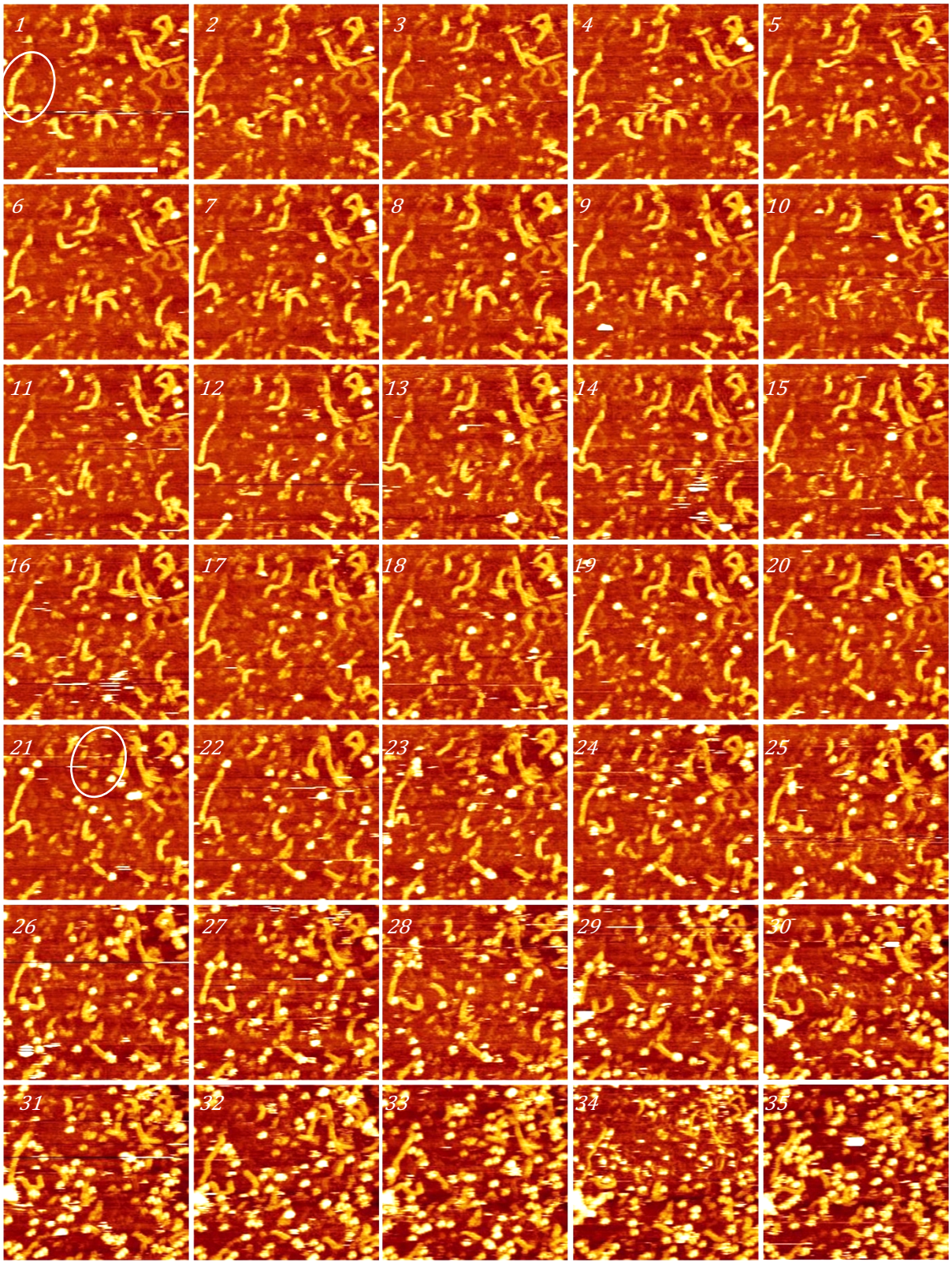


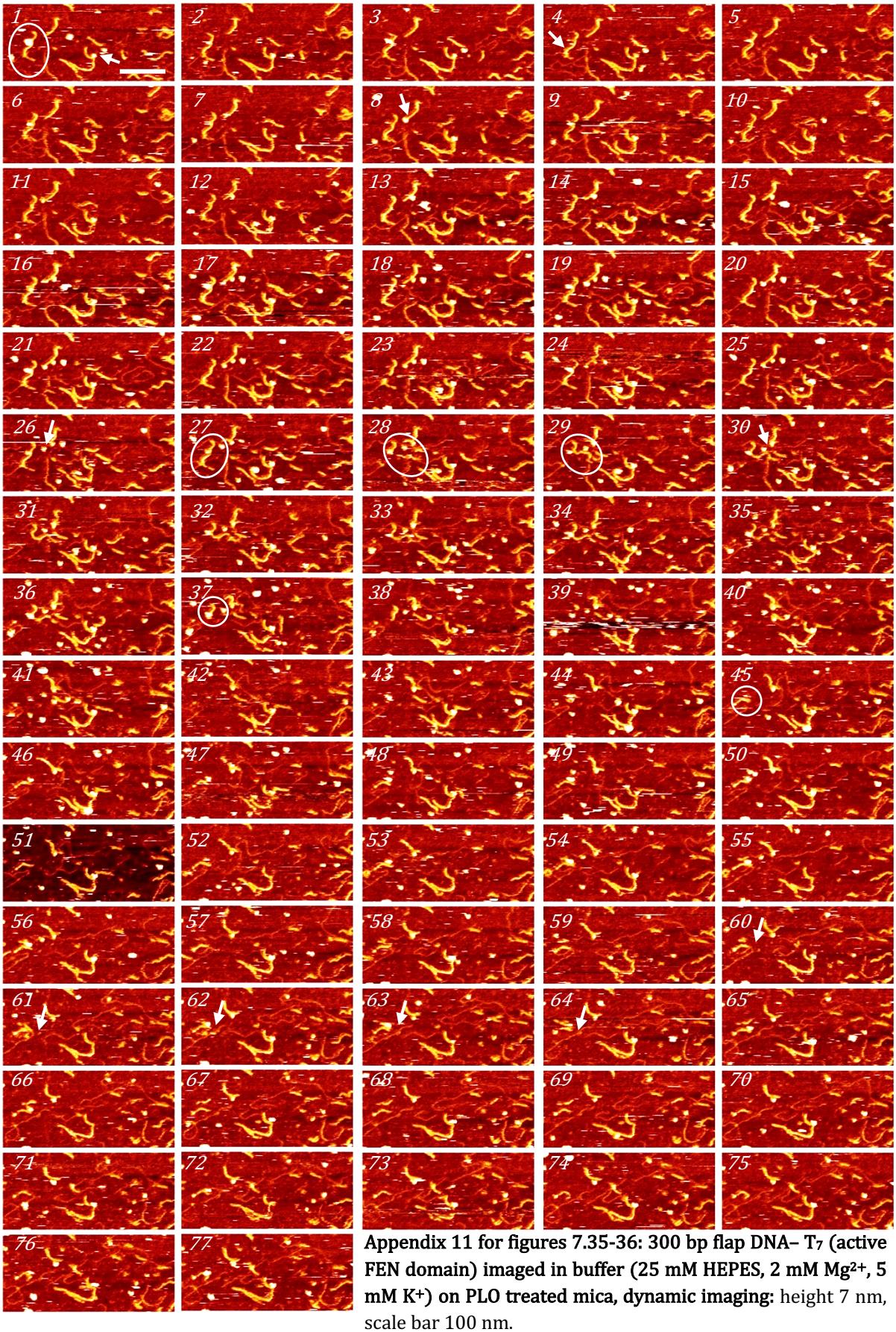
Figure 11: UV nuclease assay for *Thermus aquaticus* DNA Polymerase I: The positive control is the T_7 nuclease. The negative control is water. These results supplement the zymogram results, that there is no nuclease activity due to the inactive FEN domain.

	Absorbance at 280 nm	Concentration (mg/mL)
i	0.063	0.63
ii	0.102	1.02
iii	0.194	1.94

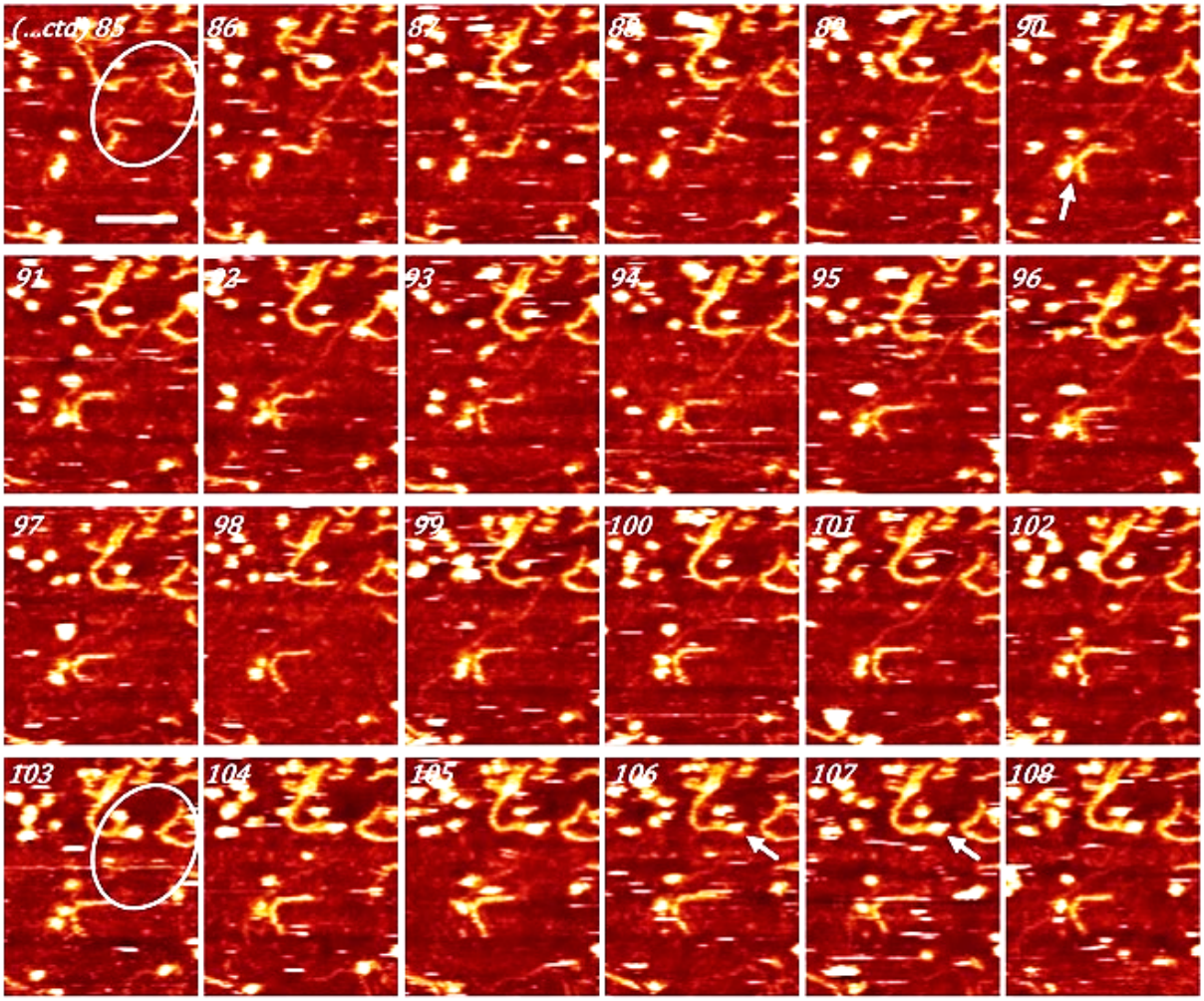
Table 4: Estimation of concentration of the protein samples using spectrophotometer.

Appendix 10 for Figure 7.19: 300 bp flap DNA–TaqPolI (inactive FEN domain) imaged in buffer (25 mM HEPES, 2 mM Ca^{2+}, 5 mM K^{+})on PLO treated mica, dynamic imaging: height 8 nm, scale bar 100 nm.

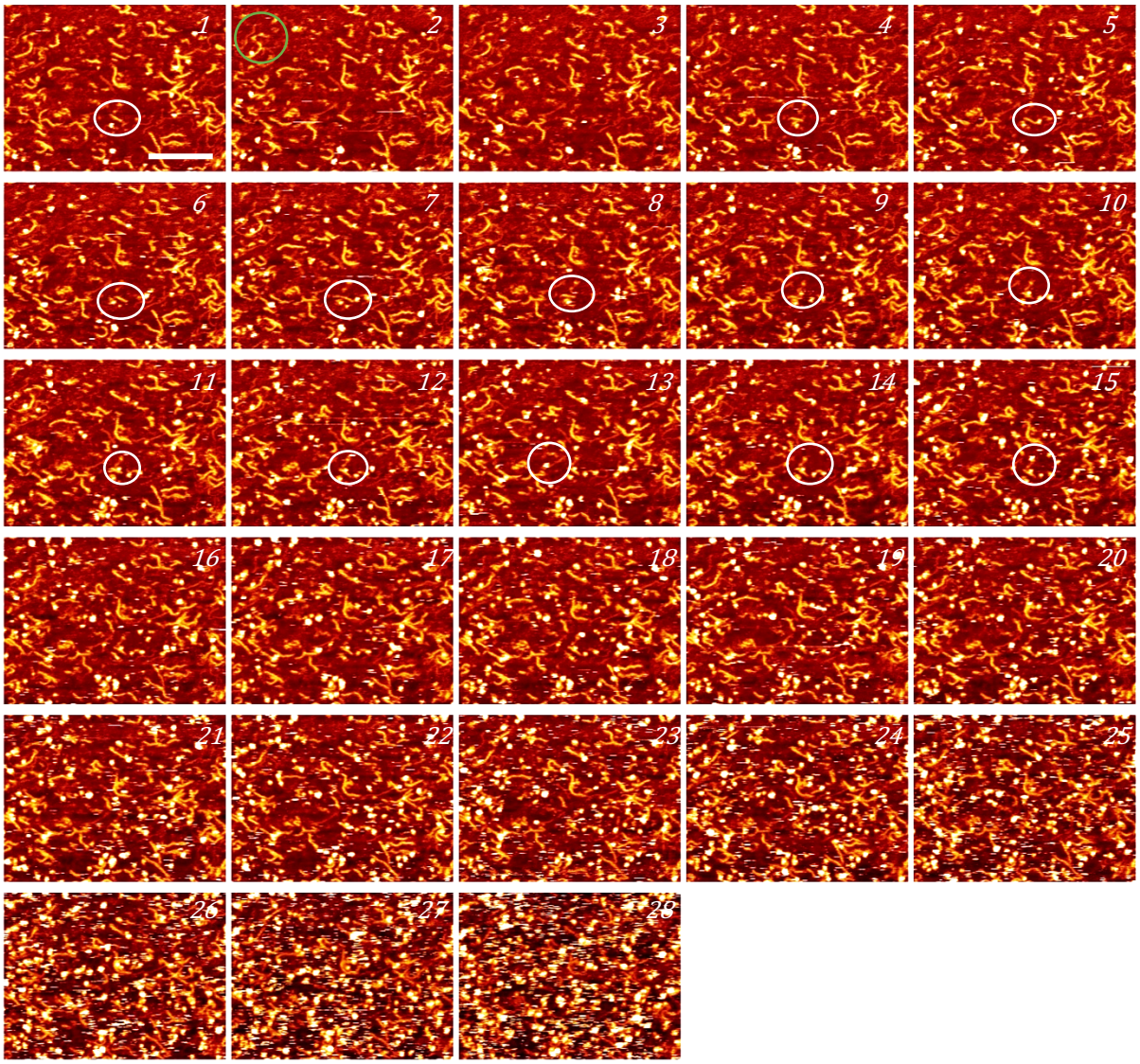




Appendix 11 for figures 7.35-36: 300 bp flap DNA– T_7 (active FEN domain) imaged in buffer (25 mM HEPES, 2 mM Mg^{2+}, 5 mM K^{+}) on PLO treated mica, dynamic imaging: height 7 nm, scale bar 100 nm.



Appendix 12 for figure 7.37-38: (last frames) 300 bp flap DNA– T_7 (active FEN domain) imaged in buffer (25 mM HEPES, 2 mM Mg^{2+}, 5 mM K^{+}) immobilized on PLO treated mica, dynamic imaging: height 7 nm, scale bar 50 nm.

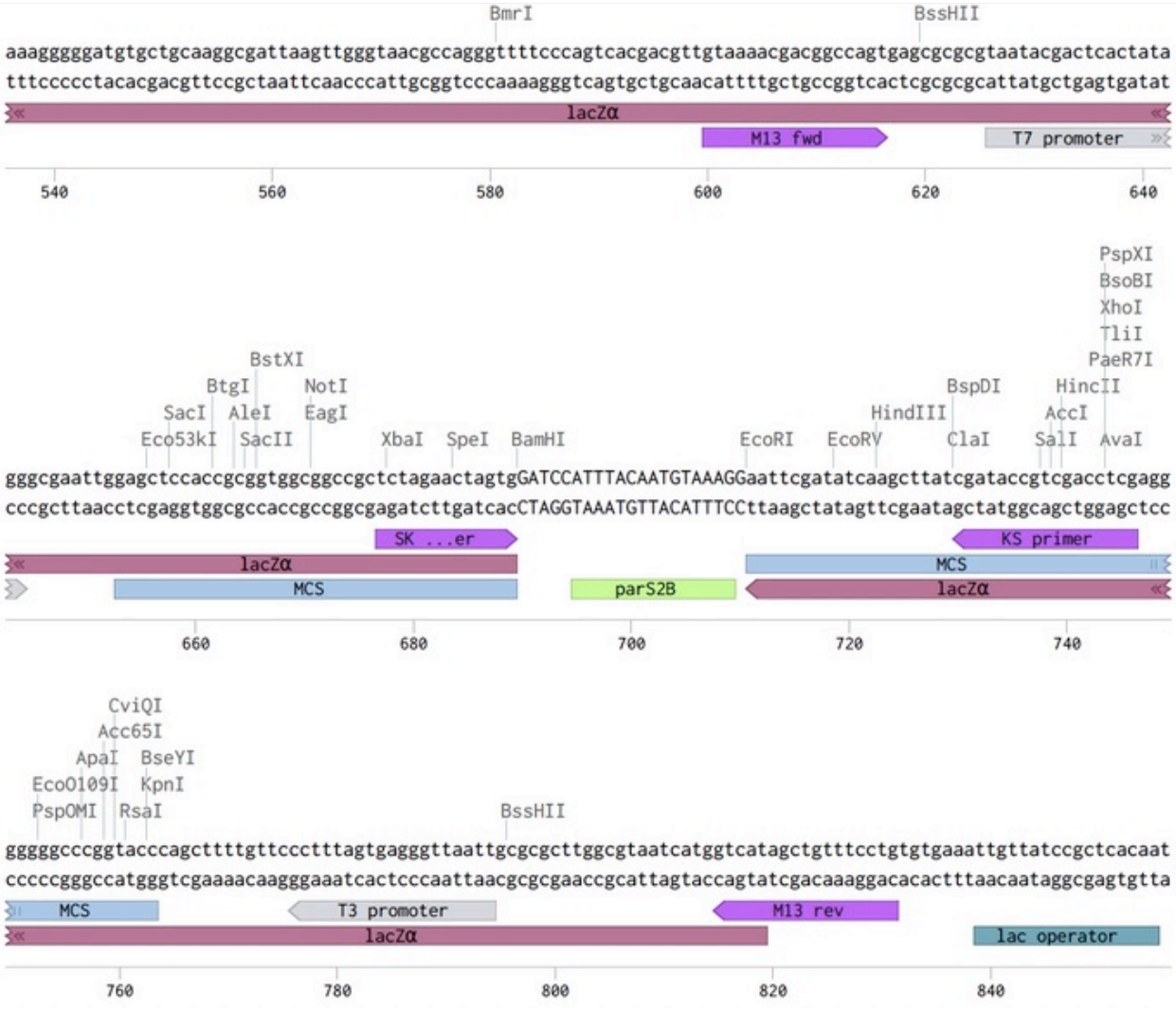


Appendix 13 for figure 7.45: 300 bp flap DNA– T_7 (active FEN) imaged in buffer (25 mM HEPES, 2 mM Mg^{2+}, 5 mM K^{+}) immobilized on PLO treated mica, dynamic imaging: height 7 nm, scale bar 100 nm.

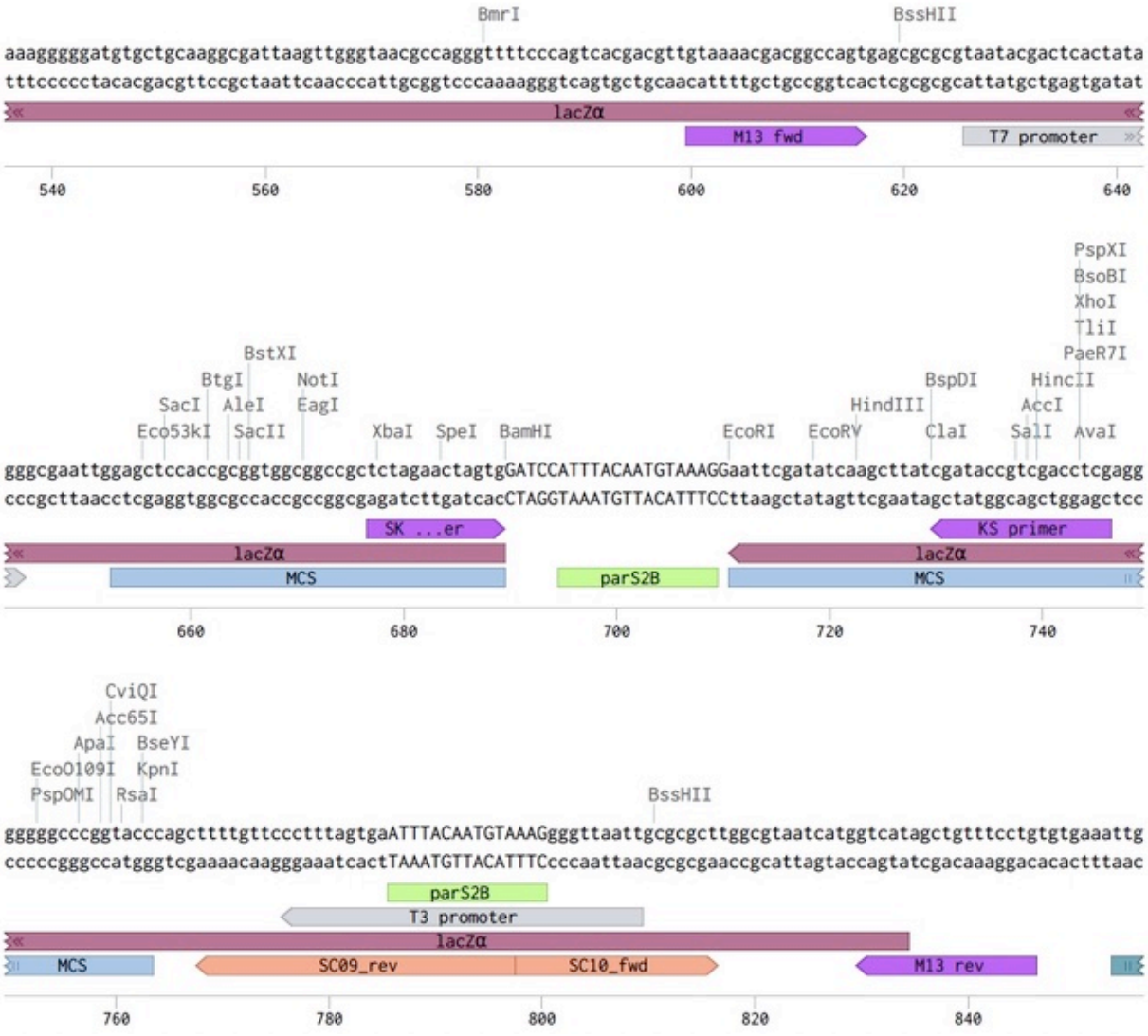
Appendix 14: Movies for dynamic imaging of FEN-flap/overhang DNA interactions from chapter 7 images: Reference for the Supplementary Material

Movie A. For image 7.13.
Movie B. For image 7.14.
Movie C. For image 7.15.
Movie D. For image 7.16.
Movie E. For image 7.17.
Movie F. For image 7.18.
Movie G. For image 7.19 and appendix 10.
Movie H. For images 7.20-21.
Movie I. For image 7.24.
Movie J. For image 7.25.
Movie K. For images 7.26-27.
Movie L. For images 7.29-30.
Movie M. For images 7.32-33.
Movie N. For images 7.35-36 and appendix 11.
Movie O. For images 7.37-38 and appendix 12.
Movie P. For images 7.39-40.
Movie Q. For images 7.42-43.
Movie R. For image 7.45 and appendix 13.
Movie S. For images 7.46-48.

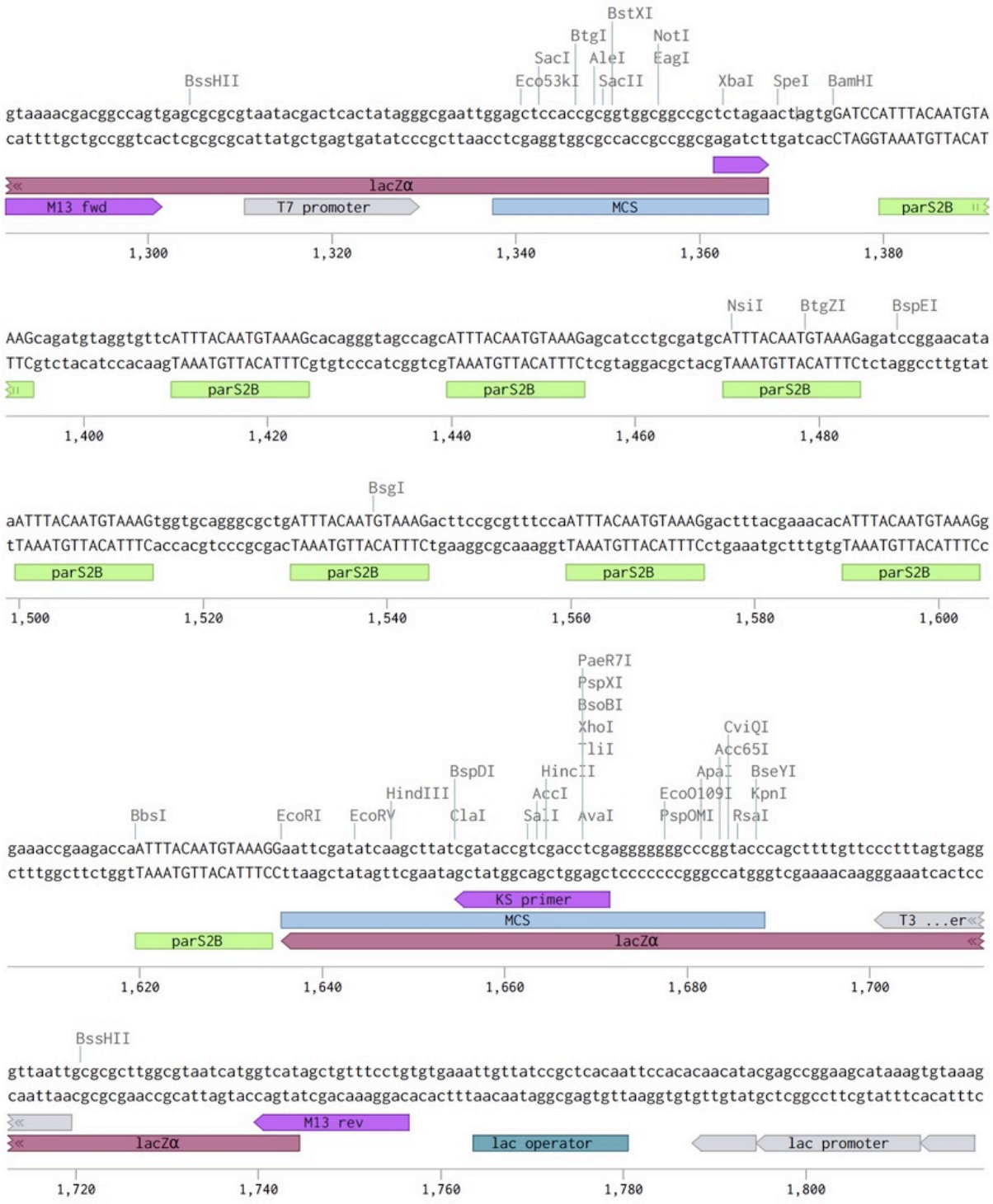
Appendix 15: DNA sequence for 1 *parS* sites containing plasmid DNA used as a template for short DNA with M13 primers used for PCR.



Appendix 16: DNA sequence for 2 *parS* sites containing plasmid DNA used as a template for short DNA with M13 primers used for PCR.



Appendix 17: DNA sequence for 9 *parS* sites containing plasmid DNA used as a template for short DNA with M13 primers used for PCR.



Appendix 18: Comparison of lengths for *parS* DNA imaged in air with and without the presence of ParB protein. The reaction buffer did not contain CTP. t-Test for Two-Sample Assuming Unequal Variances.

0 *parS DNA*	**DNA in air**	**DNA-protein in air**
Mean	60.86637931	63.1130952
Variance	71.12249572	153.312544
Observations	464	336
Hypothesized Mean Difference	0	
df	553	
t Stat	-2.877642766	
P(T<=t) one-tail	0.002080689	
t Critical one-tail	1.647613736	
P(T<=t) two-tail	0.004161378	
t Critical two-tail	1.964263051	

1 *parS DNA*	**DNA in air**	**DNA-protein in air**
Mean	70.93109244	46.6649123
Variance	41.57941884	145.485058
Observations	595	570
Hypothesized Mean Difference	0	
df	861	
t Stat	42.5579496	
P(T<=t) one-tail	2.9562E-214	
t Critical one-tail	1.646625311	
P(T<=t) two-tail	5.9123E-214	
t Critical two-tail	1.962723047	

2 *parS DNA*	**DNA in air**	**DNA-protein in air**
Mean	63.76666667	53.5618557
Variance	97.18135003	124.308825
Observations	480	388
Hypothesized Mean Difference	0	
df	779	
t Stat	14.11297602	
P(T<=t) one-tail	9.65705E-41	
t Critical one-tail	1.646812027	
P(T<=t) two-tail	1.93141E-40	
t Critical two-tail	1.963013919	

9 *parS DNA*	DNA in air	DNA-protein in air
Mean	127.4385113	102.771084
Variance	198.9046409	855.378838
Observations	618	249
Hypothesized Mean Difference	0	
df	296	
t Stat	12.72614831	
P(T<=t) one-tail	3.44672E-30	
t Critical one-tail	1.650017743	
P(T<=t) two-tail	6.89345E-30	
t Critical two-tail	1.968010728	

Appendix 19: Comparison of lengths for *parS* DNA imaged in air with and without the presence of ParB protein. The reaction buffer contained CTP. t-Test for Two-Sample Assuming Unequal Variances.

0 *parS DNA*	DNA in air	DNA-protein in air
Mean	71.38297872	54.56
Variance	56.71970398	143.0689286
Observations	47	225
Hypothesized Mean Difference	0	
df	101	
t Stat	12.39309578	
P(T<=t) one-tail	2.67933E-22	
t Critical one-tail	1.66008063	
P(T<=t) two-tail	5.35866E-22	
t Critical two-tail	1.983731003	

1 *parS DNA*	DNA in air	DNA-protein in air
Mean	72.58870968	54.66595745
Variance	185.33353	74.55129066
Observations	124	470
Hypothesized Mean Difference	0	
df	150	
t Stat	13.93914356	
P(T<=t) one-tail	3.7221E-29	
t Critical one-tail	1.6550755	
P(T<=t) two-tail	7.4442E-29	
t Critical two-tail	1.975905331	

2 *parS DNA*	**DNA in air**	**DNA-protein in air**
Mean	64.24313725	58.93782383
Variance	32.54695075	63.72599421
Observations	255	386
Hypothesized Mean Difference	0	
df	635	
t Stat	9.805701152	
P(T<=t) one-tail	1.56225E-21	
t Critical one-tail	1.647256789	
P(T<=t) two-tail	3.12449E-21	
t Critical two-tail	1.963706855	

9 *parS DNA*	**DNA in air**	**DNA-protein in air**
Mean	141.028169	137.2808989
Variance	190.1268605	172.4315628
Observations	142	89
Hypothesized Mean Difference	0	
df	194	
t Stat	2.070234475	
P(T<=t) one-tail	0.019877352	
t Critical one-tail	1.652745977	
P(T<=t) two-tail	0.039754704	
t Critical two-tail	1.972267533	