A study of protein-DNA interactions using atomic force microscopy and DNA origami

UNIVERSITY OF CAMBRIDGE

Dandan Liu

Emmanuel College

Department of Pharmacology

This dissertation is submitted for the degree of Doctor of Philosophy

September 2018

A study of protein-DNA interactions using atomic force microscopy and DNA origami

By Dandan Liu

Summary

The development and application of genome editing tools has accelerated in recent years. However, their widespread application, especially in the medical field, is delayed for various issues with the safety of the tools being one of the top concerns. Much of the detail of how proteins find their target sites and how they cleave at the sites remains unclear. Despite much progress in lab environments, where the tolerance for imprecise cutting is relatively high, *in vivo* treatment remains difficult.

Most genome editing tools are derived from naturally occurring regulatory proteins. Better understanding of the mechanisms used by these natural proteins should facilitate the use of genome editing tools. In nature, gene regulation is usually sparked by a change in the cellular environment, such as viruses invading a bacterium. Restriction enzymes defend the host bacteria by recognising specific sites on the viral DNA and cutting invading viruses at those sites. We aim first to understand how these proteins translocate along the viral DNA molecules toward their recognition sites, and then to see how their accuracy of cleavage can be increased.

Protein translocation along DNA molecules has been studied for more than 40 years. Atomic force microscopy in fluid mode allows direct observation of protein/DNA interactions. This application is about twenty years old but most of the images taken, lack the spatial and temporal resolution for quantitative studies of protein translocation dynamics. Here we achieve second-level and nanometre-scale tracking of the translocation of EcoRV, a Type IIP restriction enzyme, using fast-scan atomic force microscopy (AFM) and DNA origami techniques. We find that EcoRV tends to jump toward its recognition site from afar and then switch to slow sliding mode when it is within about 20 base pairs of its recognition site.

Our methods demonstrate how BcgI, a type IIB restriction enzyme, brings together two recognition sites both *in cis* and *in trans* before cleavage, minimising mis-cleavage at a single recognition site. We show that the collision looping model is valid but not the sliding model. The two restriction enzymes were chosen as they represent different model systems of typical restriction enzymes.

Our methods will be useful in studies of other types of restriction enzymes and other proteins or protein complexes that interact with DNA. We expect that these methods will see broader applications in studies of protein-DNA interactions. We also hope that our studies will contribute to the safe application of the genome-editing tools in medical contexts.

Declaration

- This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.
- It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.
- It does not exceed the prescribed word limit for the relevant Degree Committee.

Acknowledgements

My first thanks go to my supervisor Robert Henderson. It is because of his trust I have never lost my faith in finding the way forward even in the most difficult moments of this PhD journey. Discussing with Dr.Henderson about my work is one of the things I have enjoyed the most during the past four years. Many suggestions Dr.Henderson has given me have led to important improvements of this work.

I am also very grateful for the help of all the members and visiting members in Robert Henderson and Mike Edwardson's lab. One special thanks should go to Ioanna Mela for introducing the lab techniques, including but not limited to Atomic Force Microscopy (AFM), to me, and for her support throughout all my PhD years. Another special thanks must go to Camilo Navarrete. His deep understating of the theories behind AFM imaging helped improve my image quality greatly.

Last but not least, I give my sincere thanks to Benjamin DeMeo. Without his support I would never have finished this thesis.

Abbreviations

A: Adenine

AFM: Atomic force microscope or Atomic force microscopy

AREs: Artificial restriction enzymes

bp: Base pairs

BPC: Biotechnology performance certified

C: Cytosine

CRISPR-Cas9: Clustered regularly interspaced short palindromic repeat associated enzyme 9

Cyro-EM: Cryo-electron microscopy

dsDNA: Double strand DNA

EM: Electron microscopy

FRAP: Fluorescence recovery after photobleaching

FRET: fluorescence resonance energy transfer

G: Guanine

GFP: green fluorescent protein

HDR: Homology-dependent repair

HS-AFM: High-speed AFM

mRNA: Messenger RNA

MSD: Mean square displacement

NHEJ: non-homologous end joining

NMR: nuclear magnetic resonance

R-M: restriction-modification

SPIP: Scanning Probe Image Processor

STM: scanning tunnelling microscope

ssDNA: single strand DNA

T: Thymine

TALENs: Transcription activator-like effector nucleases

ZFNs: Zinc-finger nucleases

Table of Contents

Chapter 1 Introduction and background

1.1 DNA

DNA is the fundamental carrier of genetic information in most organisms. Due to its complexity much about it still remains to be discovered even now, about 150 years after its discovery. In 1859, Charles Darwin's theory of evolution prompted scientists to wonder how characteristics are passed down from generation to generation (Darwin, 1859). Gregor Mendel's 1866 work on peas revealed there was an innate yet 'invisible factor' in providing visible traits in predictable ways (Mendel, 1866). That 'invisible factor' was actually isolated by Friedrich Miescher in 1869, soon after Mendel's paper, although Miescher did not know the function of this newly isolated material (Dahm, 2005). It was almost a century later, in 1944, that Oswald Avery identified DNA as the 'transforming principle' (Avery *et al.*, 1944). In 1952, Rosalind Franklin photographed crystallized DNA fibres, and in 1953, James Watson and Francis Crick discovered the double helix structure of DNA (Watson and Crick, 1953). Since then, much progress has been made on reading the genetic codes carried by DNA and interpreting and editing the information. In 1977 Frederick Sanger developed rapid DNA sequencing techniques (Sanger *et al.,* 1977). In 1990 the Human Genome Project began, and it was completed in 2003 with almost all the three billion base pairs of human genome sequenced with high quality. Now a basic gene sequencing service for a person can be performed within days, with interpretation of the sequence providing information on, for example, whether that person has high chance of developing breast cancer – diseases with relatively better-known genetic markers. In addition, gene editing has been made possible by a series of discoveries of bacterial proteins that cut DNA at specific sites, including thousands of restriction enzymes and

recently discovered Cas proteins. The precision of gene editing has been constantly improved with increasing understanding of the mechanisms employed by the proteins that interact with DNA.

The impressive scope of our knowledge about DNA means this section needs to be highly selective about what to discuss. As this thesis concerns mainly on protein-DNA interaction, we shall focus on DNA structure, sequence and topology.

DNA is a macromolecule made up of two strands, each of which is a polymer of deoxyribonucleotide monomers linked by phosphodiester bonds. The two strands twist around each other, resulting in an overall helical shape – the 'double helix' described by Watson and Crick (Watson and Crick, 1953) (Figure 1.1i). Each nucleotide base of one strand pairs with a nucleotide base of the other strand to link the two polymers together with the rule: adenine (A) pairs with thymine (T), and guanine (G) pairs with cytosine (C) (Figure 1.1ii). The pairing is achieved by hydrogen bonding, Van der Waals forces, and hydrophobic forces between the nucleotide bases.

The double helix structure produces regular grooves along a DNA chain. Each helical turn (comprising about ten base pairs) includes a 'major groove' and a 'minor groove' where the edges of the bases are exposed to the environment. Most proteins interact with DNA at the grooves, especially when the binding is sequence-specific.

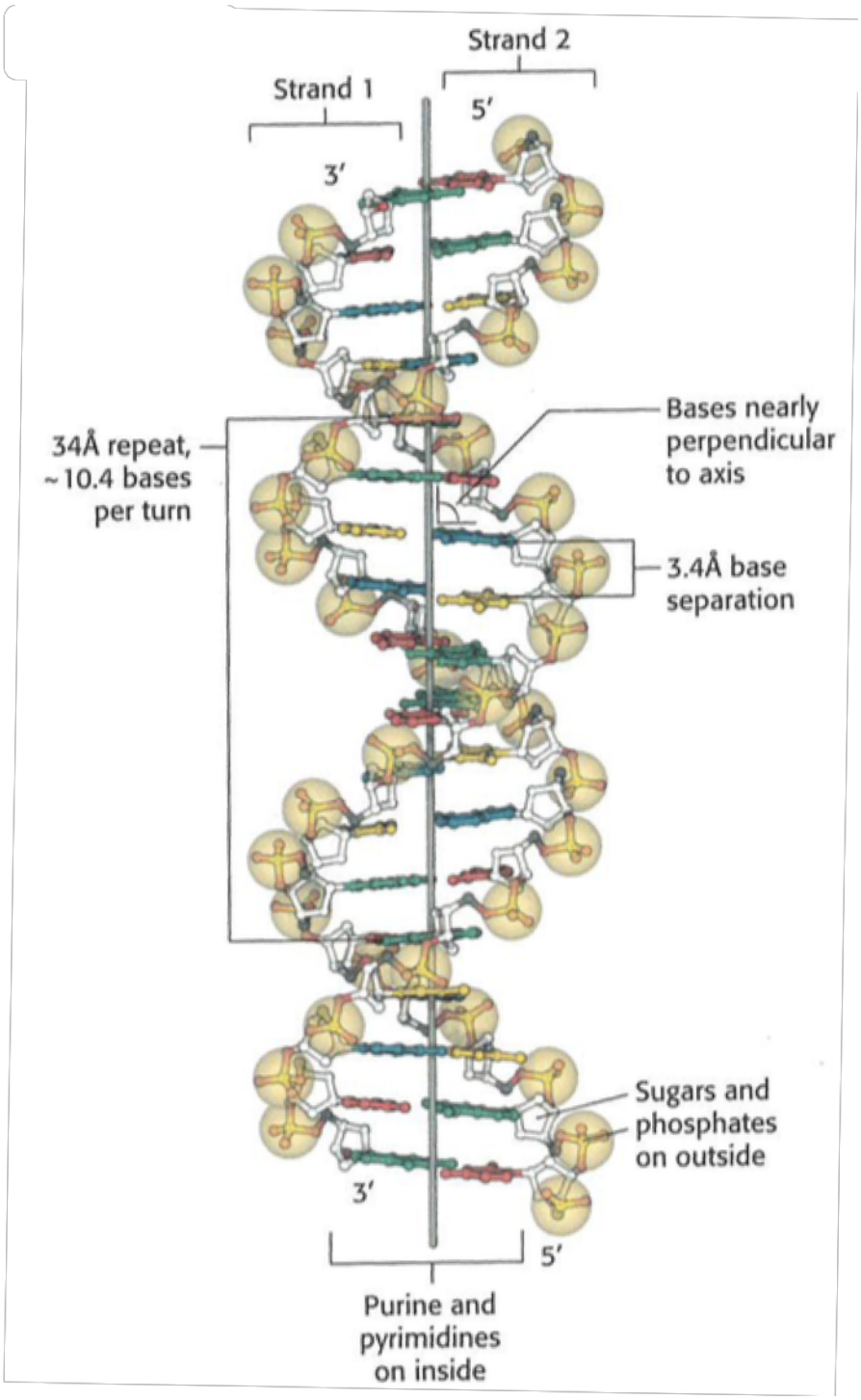


Figure 1.1i The Watson and Crick model of B form DNA structure. The double helix structure is formed by two single strands, each of which is a polymer of deoxyribonucleotide monomers linked by phosphodiester bonds. To stabilise the double helix structure, nucleotide base of one strand is paired with that of the other strand, leaving sugars and phosphates on the outside. The rule for the pairing is that purines adenine and guanine are paired with pyrimidines thymine and cytosine respectively. Each helical turn includes about 10.4 bases and spans 35.4 Å. The helix axis goes through the centre of the base pairs and the bases are nearly perpendicular to the axis. The figure is adapted from Berg *et al*., 2015.

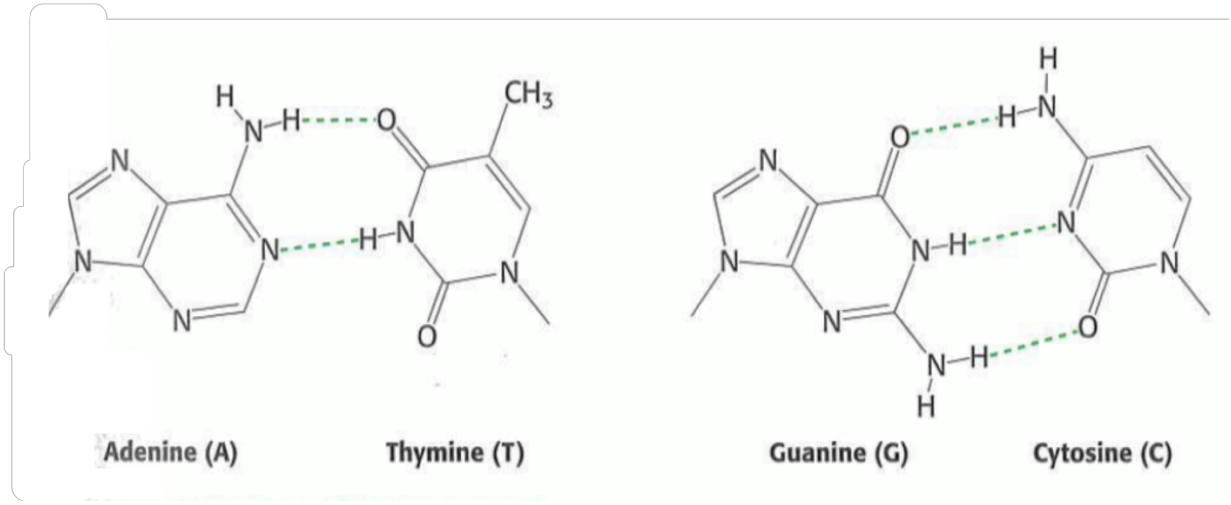


Figure 1.1ii Pairing of the nucleotide bases in the DNA structure. Adenine (A) pairs with thymine (T), and guanine (G) pairs with cytosine (C). The pairing is achieved by hydrogen bonding, Van der Waals forces, and hydrophobic forces between the nucleotide bases. The figure is adapted from Berg *et al*., 2015.

Although most DNA structural models (including the one mentioned above) are built based on the commonly encountered B-form DNA, it is worth noting that there are two other forms – A form and Z form. All three forms have a double helix structure and follow the same basic principles of base-pairing (i.e. A=T and G=C). The main difference between each form lies at the geometrical relationship between the bases and the helical axis and the direction of the helical turns (Figure 1.1iii). According to the eighth edition of Biochemistry (Berg *et al*., 2015), in B form, which is the fully hydrated DNA form, the helix axis goes through the centre of the base pairs and the bases tilt only about 1 degree towards the helix axis – almost perpendicular to the helix axis. The helix diameter is about 20 Å and pitch per turn of helix is 35.4 Å (Berg *et al*., 2015). The major groove is wide, and the minor groove is narrow. Dehydration can change a DNA strand's structure from B form to A form. During the change, the tilt of the bases towards the helix axis increases from 1° to 19° (Berg *et al*., 2015). The helix axis itself shifts towards the major groove, making the major groove deep and narrow and the minor groove shallow and wide. The overall structure is also widened to about 26 Å, an approximately 30 percent increase from the 20 Å-diameter of B form (Berg *et al*., 2015). Pitch per turn of helix is 25.3 Å, making A form more compact than B form (Berg *et al*., 2015).

Z form DNA is narrower and more stretched than B form DNA. Z form DNA has a helix diameter of about 18 Å, and the bases tilt about 9 degrees towards the helix axis (Berg *et al*., 2015). Each helical turn has a pitch of about 45.6 Å (Berg *et al*., 2015). In each turn, there is a flat major groove and a deep and narrow minor groove. Z form cannot be obtained from B form simply by dehydration, primarily because this process requires the change of the DNA from turning right-hand (in B-form) to turning left-hand (in Z form). *In*

vivo, this change usually happens during the unwinding of supercoiled B form during DNA replication, recombination and transcription (Berg *et al*., 2015).

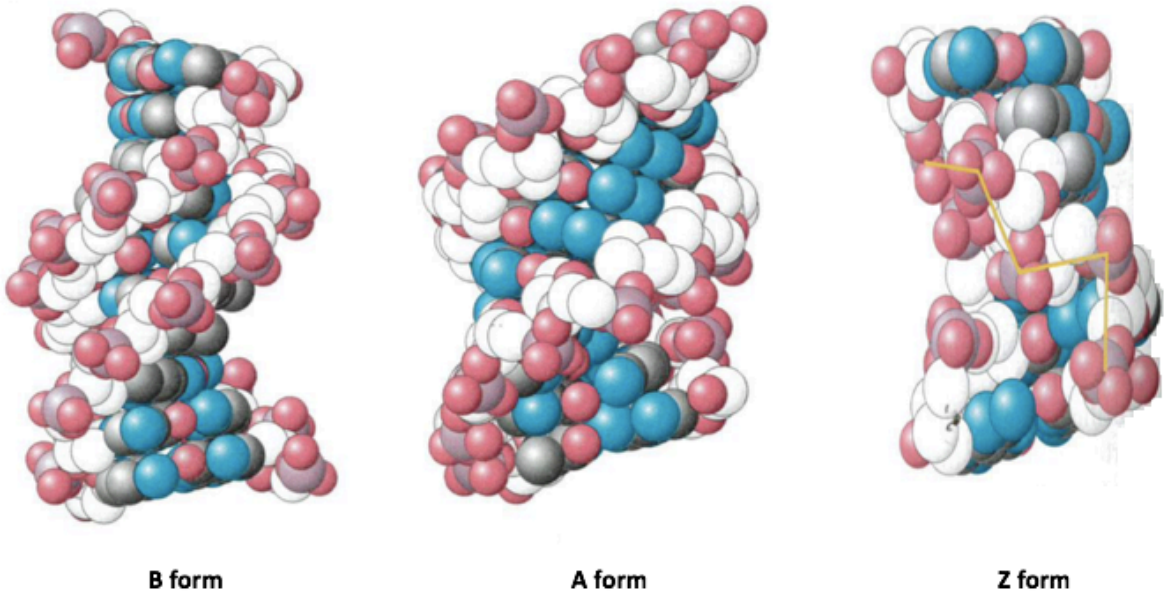


Figure 1.1iii DNA in three forms: the usual B form and two alternatives - A form and Z form. Base pairs per turn of the helices are 10.4, 11 and 12 in B form, A form and Z form respectively. Pitch per turn of the helices are 35.4 Å, 25.3 Å and 45.6 Å respectively. Helix diameter are about 20 Å, 26 Å and 18 Å respectively. Tilt of base pairs from perpendicular to helix axis are 1 degree, 19 degrees and 9 degrees respectively. B form and A form are right-handed, whereas Z form is left-handed. The figure is adapted from Berg *et al*., 2015.

While eukaryotic genomic DNA is linear, circular DNA and supercoiled DNA is commonplace in bacteria and archaea. In addition, many DNA molecules *in vivo* adopt compact secondary structures which store more energy than in their relaxed form (Berg *et al*., 2015). For example, both the DNA molecules wrapping around the histones in nucleosomes and the DNAs in supercoils have more energy stored in them. The energy is often used for driving DNA unwinding during DNA replication, recombination and transcription (Berg *et al*., 2015).

Changing the DNA topology causes a change in the available genetic information. For example, during the usually reversible transition of supercoiled B form DNA to Z form DNA in the process of transcription, each change of the right-hand turn of the B form to a left-hand turn of the Z form would cause the removal of two supercoils, and so the remaining B form DNA is more relaxed with better availability for potential interactions with proteins (Zhang *et al.,* 2006). Supercoiled DNA may also change topologically to facilitate long distance communication. For supercoiled DNA, fewer twists in the local secondary structure is compensated for by an immediate increase in writhe. This is good for long-distance communication, as a local change may affect a distant site by modulating the larger-scale property of writhe. It explains how co-regulatory proteins manage to interact with DNA sequences as far as thousands of base pairs away from each other during the initial stage of transcription (Mondal and Bhattacherjee, 2017). In addition, juxtaposition of distant sites usually facilitates the binding of proteins which need at least two DNA recognition sites to conduct proper interactions, such as site-specific recombination and DNA cleavage (Mondal and Bhattacherjee, 2017).

1.2 Proteins

In the "central dogma" of molecular biology, nuclear DNA is first transcribed to messenger RNA (mRNA) and then translated into proteins (Crick, 1970). Proteins are macromolecules that carry out critical functions in essentially all biological processes, including structural support, catalysis, and signal transmission (Berg *et al*., 2015).

A protein can have as many as four levels of structure. The *primary structure* is a sequence consisting of the 20 amino acids that are the basic components of all proteins (Berg *et al*., 2015). Certain amino acids near one another can form hydrogen bonds, resulting in a locally three-dimensional structure known as the *secondary structure* (Berg *et al*., 2015). The *secondary structure* brings some amino acids far away from each other in the sequence close to each other in the space, resulting in long-range interactions between amino acids that give the protein its *tertiary structure* (Berg *et al*., 2015). Some large proteins are composed of polypeptide chains, each with its own tertiary structure (Berg *et al*., 2015). These large proteins are said to have a *quaternary structure* formed by the polypeptide chains (Berg *et al*., 2015). The exact structure of a protein is almost impossible to predict solely from its amino acid sequence (Berg *et al*., 2015). The intricate structures of proteins allow them to interact with other biological macromolecules. These interactions alter the process of life.

The study of protein structure and function is central to the life sciences. Purification is the essential first step in discovering proteins' structure and function. Proteins can be separated from other molecules in the environment based on size, solubility, charge and binding affinity (Berg *et al*., 2015). There is an array of techniques for determining the structure of a purified protein. Mass spectrometry, X-ray crystallography and nuclear magnetic

resonance (NMR) spectroscopy have been the major contributors. Mass spectrometry is especially useful for detecting the amino acid sequence of proteins. The protein database has rapidly expanded over the past few years. Cryo-electron microscopy (cryo-EM), a recent Nobel-winning technology, has speeded the development of the structural studies (Henderson *et al.*, 1990, Frank, 2017, Dubochet, 2012)

Many protein functions are revealed by studying how they interact with other molecules. In these studies, the protein in question is often stained or tagged and the tags are then traced. Techniques in this vein include fluorescence resonance energy transfer (FRET), green fluorescent protein (GFP) and fluorescence microscopy. Other techniques that do not require tagging have been developed. Atomic force microscopy (AFM) is one of them. It has been widely used, including in studying protein-DNA interactions.

1.3 Protein-DNA interactions

1.3.1 Protein-DNA interactions - Overview

There are two types of protein-DNA interactions based on where the interaction happens – DNA-backbone interactions and base interactions. DNA-backbone interactions provide a scaffold for the proteins to translocate along the DNA, but they are not specific enough for sequence recognition (Luscombe *et al*., 2001). The sequence of nucleotides does not matter for non-specific interactions, as the interactions happen between the sugar-phosphate backbone of the DNA and the functional groups on the protein. On the other hand, specific protein-DNA interactions depend on the sequence of bases in the DNA and the orientation (e.g. twisting or writhing) of the bases. Base interaction offers specificity by requiring matching between the protein side chains and the base edges exposed in the major and minor grooves of the DNA structure (Luscombe *et al*., 2001). Specific or not, most protein-DNA interactions involve van der Waals contacts (Luscombe *et al.*, 2001). Specific interactions are mediated by forces from many interactions such as hydrogen bonds, salt bridges, and Van der Waals forces (Berg *et al*., 2015). Hydrogen bonds tend to be formed only when specificity is required (Luscombe *et al.,* 2001).

Specific interactions happen when the protein recognizes specific segments of the DNA, as in the case of restriction enzymes cutting only at their 'recognition sites'. Because the major grooves can accommodate larger structural motifs and the pattern of base pairs that are exposed in the major grooves is more specific and discriminatory, most of the protein-DNA interactions involve the major grooves (Berg *et al*., 2015).

Although researchers have striven to find a universal protein-DNA recognition code for specific interactions, no simple rules have been discovered. The consensus is that the specificity for protein-DNA interactions cannot be described by one-to-one correspondences between amino acids and bases because the rules governing interactions vary greatly between different protein families (Luscombe *et al.*, 2001). However, certain amino acids do prefer certain nucleotides. Bioinformatics studies have shown that arginine and lysine prefer guanine, and asparagine and glutamine prefer adenine (Luscombe *et al.,* 2001). Although the codes of proteins/DNA interaction remain elusive, some patterns for DNA recognition in the protein sequence and its secondary structure have been discerned. For example, if a protein has a prominent helix-turn-helix secondary structure, it is highly likely that the protein interacts with DNA. Cysteine is found quite commonly in proteins' DNA-binding domains, such as the Cys6 domain (i.e. Cys-X2-Cys-X6-Cys-X6-Cys-X2-Cys-X6-Cys, where X represents any amino acid) in zinc finger motifs (Luscombe *et al.*, 2001).

The specificity of protein-DNA interactions was first demonstrated in the 1950s by identifying the recognition sites of restriction enzymes on the DNA molecules (Luria and Human, 1952, Anderson and Felix, 1952, Bertani and Weigle, 1953, Luria, 1953). And restriction enzymes have been good models for studying protein-DNA interactions ever since. In fact, as gene therapy looms on the horizon, and the precision and efficiency of protein-DNA interactions becomes a paramount issue, the studies of restriction enzyme/DNA interactions might once again provide much-needed information to facilitate the new tools' development. With the help of new technologies developed in the past few decades, especially in single-molecule manipulation and visualisation, modern studies of restriction enzyme/DNA interactions go far beyond identifying the recognition sites.

1.3.2 Protein-DNA interactions - Restriction enzyme/DNA interactions

Restriction enzymes were first observed in the early 1950s (Luria and Human, 1952, Bertani and Weigle, 1953, Anderson and Felix, 1952, Luria, 1953). Since then, more than 3,600 restriction enzymes have been characterised. The original researchers noticed that after invading a new bacterium, a virus produced progeny that had a different bacterial host range than the parent virus (Luria and Human, 1952). Moreover, invading the original bacterial host would return the progeny virus to its parent's original form (Luria and Human, 1952). Such non-hereditary changes were termed 'host-controlled variation in bacterial viruses' (Bertani and Weigle, 1953). Now we know that these changes are caused by the endonuclease/restriction and methyltransferase/methylation activities of restriction enzymes (also called restriction-methylation enzymes or R-M enzymes). These enzymes protect the bacteria by cutting foreign invading DNA sequences at specific recognition sites. Because those recognition sequences are usually short and are likely to exist within the bacterial genome, methyl groups are added to those sequences in the bacterial genome to prevent its own DNA from being damaged.

Understanding of restriction enzymes has transformed biological and medical research. Arber and Linn demonstrated in the 1960s that such 'host-controlled variations in bacterial viruses' are caused by the restriction and methylation (R-M) enzymes, and they predicted that restriction enzymes would be used as a tool for the sequence-specific cleavage of DNA (Arber and Linn, 1969). The prediction became the reality after Type II restriction enzymes were discovered. The first ever discovered Type II restriction enzyme was HindII. It was discovered in 1970, in Hamilton Smith's laboratory (Kelly and Smith, 1970).

Since the early 1970s, restriction enzymes - especially Type II restriction enzymes with highly predictable cutting sites - have been essential in recombinant DNA technology (Jackson *et al*., 1972). Restriction enzymes have facilitated many remarkable biological studies and discoveries. They have been used to locate mutations, generate human linkage maps, and identify disease genes (e.g. Huntington's disease) and pathogenic bacterial strains (e.g. *Haemophilus influenzae*) (Gusella *et al*., 1983, Housman and Gusella, 1983, Kolata, 1983, Fraser *et al*., 1995). In addition, fingerprinting was also invented using restriction enzymes as the main tool (Gill *et al*., 1985).

Restriction enzymes are traditionally classified into four types: Type I – Type IV. This classification is based primarily on their behaviour in DNA recognition and cleavage – sequence specificity, cleavage position and cofactor requirements – rather than genetic or structural similarities. Type II restriction enzymes are the most well-studied among the four types.

Type II restriction enzymes

Over 3,500 Type II restriction enzymes, with about 350 unique DNA recognition sites, have been characterized; and thousands more 'putative' Type II restriction enzymes that are found via genome analysis of bacteria and archaea are yet to be characterized. (Loenen *et al.*, 2014b).

Type II restriction enzymes usually have only DNA restriction function. (DNA methylation for the same DNA sequence is usually conducted by another enzyme.) They cut the DNA near or within the recognition sequence at a fixed position. Because of this property, they are widely used as tools to cut DNA at specific sites.

Type II restriction enzymes are a large group (Table 1.3.2). In fact, under the umbrella of Type II restriction enzymes, there are many subgroups with distinctive properties – Type IIA, Type IIB, Type IIC, Type IIE, Type IIF, Type IIG, Type IIH, Type IIM, Type IIP, Type IIS and Type IIT (Loenen *et al*., 2014b). Because the subgroups are classified only based on the enzymes' behavioural features, a Type II restriction enzyme exhibiting multiple forms of behaviour can belong to multiple subgroups (Loenen *et al.*, 2014b). For example, BcgI is a member of Group IIA, IIB, IIC, and IIH, because BcgI recognises an asymmetric sequence, cleaves on both sides of the two recognition sites it binds, comprises a fused endonuclease-methyltransferase subunit and has a Type I-like DNA-specificity subunit respectively (Loenen *et al*., 2014b).

The five most-studied subgroups are Type IIP (P for 'palindromic'), Type IIB (B for 'both-sides cleavage at each one of the two recognition sites'), Type IIS (S for 'shifted

cleavage'), Type IIC (C for 'combined restriction and modification'), and Type IIT (T for 'two different catalytic sites').

Table 1.3.2. The five most-studied subgroups of Type II restriction enzymes.

Type II restriction enzymes	Type II restriction enzymes	Example(s)	General features
Typical	Type IIP	EcoRV	• Recognize a palindromic DNA sequence of 4-8 base pairs long • Cut at specific pointes within the sequence
Untypical (unexhausted)	Type IIB	SfiI, BcgI, BplI	• Bind and cleave at specific locations on both sides of the two recognition sites
	Type IIS	FokI	• Usually recognize asymmetric DNA sequences • Cleavage usually happens several bases away from the recognition site and at different points on both DNA strands
	Type IIC	BcgI	• Have both restriction and methylation functions • Cleavage usually happens several bases away from the recognition site and at different points on both DNA strands
	Type IIT	Bpu10I, BslI	• Use two different catalytic sites for cleavage, each of which is specific for one particular DNA strand

Type IIP restriction enzymes

The majority of restriction enzymes used as biological tools are Type IIP restriction enzymes. They usually recognize certain palindromic DNA sequences of 4-8 base pairs long, and they cut within the sequences. They are the simplest and smallest of all restriction enzymes (about 250-350 amino acids). Most of the Type IIP restriction enzymes form homodimers, with each monomer binding to and cutting one strand of the DNA, resulting in double-strand breaks. One example is EcoRV, which is the subject of the studies in Chapter Two.

Some Type IIP enzymes that recognize four base pairs are monomers. Therefore, they need to cleave one DNA strand before moving to and then cutting the other DNA strand. The movement is very fast, thus few 'nicked' intermediates (with only one strand being cleaved) have been observed. Some examples of such enzymes are MspI, HinP1I, BstNI and NciI.

Some more complex Type IIP enzymes, such as SfiI and NgoMIV, are also Type IIB enzymes, because they cleave on specific locations on both sides of the two recognition sites (Marshall and Halford, 2010). They do this usually by forming complex homotetramers - dimers of homodimers, which will be further examined in Chapter Three using BcgI, a Type IIB enzyme, as an example.

Type IIS restriction enzymes

Type IIS restriction enzymes have two separate protein domains for DNA recognition and DNA cleavage. They are typically formed by 400-600 amino acids. The separate domains mean the cleavage usually happens several bases away from the recognition site, thus Type IIS are known for this 'shifted cleavage' (Lundin *et al*., 2015). For example, the Type IIS enzyme FokI recognizes GGATG and cuts this ('top') strand 9 bases to the right and the complementary ('bottom') strand 13 bases to the same direction – i.e. GGATG 9/13 (In some uncommon cases FokI might cleave at 8/12 or 10/14, depending on how the domains are linked and the helical twist of the bound DNA in different conditions.) (Aggarwal and Wah., 1998). Type IIS restriction enzymes typically remain monomers and recognize asymmetric DNA sequences (Szybalski *et al*., 1991). Only dimerized Type IIS enzymes can conduct the cleavage, although the second enzyme does not always have to bind the DNA (Szybalski *et al*., 1991).

Type IIC (aka. Type IIG) restriction enzymes

Most of the better-known Type II restriction enzymes, such as Type IIP and Type IIS enzymes, have no methylation function. Type IIC restriction enzymes, however, have both restriction and methylation functions in one single enzyme (Pingoud and Jeltsch, 2001). Type IIC restriction enzymes do this by having three domains – one for cleavage, one for methylation, and another for sequence-recognition, which provides the same specificity for both of the other two domains in the way similar to the arrangement of Type I R-M enzymes (Pingoud and Jeltsch, 2001).

Separate recognition and cleavage domains result in the cleavage site being outside of the recognition site. The cleavage pattern is similar to that of Type IIS restriction enzymes, with the break on one DNA strand slightly farther away from the recognition site than the break on the other DNA strand from the other DNA recognition site (Pingoud and Jeltsch, 2001).

Type IIC restriction enzymes are usually 800-1200 amino acids long. They've been found to act as monomers, dimers and oligomers. Because of the usually large size of Type IIC enzymes, they are not widely used as tools in molecular biology. However, they are interesting subjects for biochemistry and enzymology. Type IIC enzymes, unlike other Type II restriction enzymes, are closely similar in amino acid sequences. This has propelled researchers to correlate the DNA sequence recognized with the amino acids at the active sites of the proteins. The correlation has led to better understanding of the amino acid to base pair 'recognition code' and might eventually become the key to improve specificity (Pingoud and Jeltsch, 2001).

Type IIT restriction enzymes

Type IIT enzymes use two different catalytic sites for cleavage, each of which is specific for one particular DNA strand (Pingoud and Jeltsch, 2001). If one catalytic site of a Type IIT enzyme is disrupted by mutation, the enzyme does not become inactive, but rather turns into a strand-specific ‘nicking’ enzyme for one DNA recognition site.

Many Type II restriction enzymes need two recognition sites to fully function, though they might not cut the second site (Loenen *et al*., 2014b). For example, BcgI needs two recognition sites and cuts them concertedly whereas EcoRII brings two recognition sites together only to cut one of them while using the other as an allosteric stimulator (Loenen *et al.*, 2014b). This necessity of two recognition sites will be discussed in detail in Chapter Three.

Because most DNA cleavage using Type II restriction enzymes results at staggered ends (also known as sticky ends) at the sites of the cleavage, different DNAs cut by the same Type II restriction enzyme can be annealed together by DNA ligases. This 'cut and paste' practice is the foundation of recombinant DNA technology. In 1970s bacterial plasmids were developed into vectors with specific Type II restriction enzyme recognition sites at fixed positions to facilitate DNA recombination and cloning (Cohen *et al*., 1973, Hershfield *et al.*, 1974). It accelerated the use of recombinant DNA technology. Over the past a few decades, many eukaryotic DNAs have been cloned with the help of Type II restriction enzymes, leading to the production of many pharmaceutically important proteins such as vaccines, hormones and insulin (Loenen *et al*., 2014b).

Type I restriction-modification (R-M) enzymes

Type I restriction-modification (R-M) enzymes are complex that have both restriction and modification functions and cut DNA at a variable distance from the DNA sequence they recognise (Loenen *et al*., 2014b). Cells need to change the specificity of their R-M enzymes periodically to fend off the evolving viruses and other mobile genetic elements (Loenen *et al*., 2014a). To ensure the cell's own DNA are not mistakenly cleaved, the specificity of the methylation and the restriction must change at the same time and in the same way (Loenen *et al.*, 2014a). Such changes are often exerted by Type I R-M enzymes, which have a common DNA sequence-recognition subunit for both methylation and restriction (Loenen *et al*., 2014a).

Type III R-M enzymes

The Type III restriction-and-modification enzymes are loosely defined. And they fall somewhere in between Type I and Type II restriction enzymes (Loenen *et al.,* 2014b). Generally, Type III restriction-and-modification enzymes recognize asymmetric sequences and cleave 25-27 nucleotides away from the recognition sites (Loenen *et al.*, 2014b). Type III restriction-and-modification enzymes often need two recognition sequences in opposite orientations within the same DNA molecule to initiate the cleavage (Raghavendra *et al.,* 2012). The distance between the two DNA recognition sites can be up to thousands of base pairs (Raghavendra *et al.,* 2012). There is some speculation about how they bring the two recognition sites together (Crampton *et al.,* 2007, Ramanathan *et al.,* 2009, Schwarz *et al.,* 2011).

Like many other proteins which need two DNA recognition sites to conduct cleavage, Type III restriction-and-modification enzymes can cleave a single recognition site, albeit with extremely low efficiency (Raghavendra and Rao, 2004).

Type IV restriction enzymes

Type IV restriction enzymes are modification-dependent. They recognize modified, typically methylated DNA. The understanding of Type IV restriction enzymes is still very limited (Machnicka *et al*., 2015).

1.4 Single-molecule observations for protein-DNA interactions

Before single-molecule studies, information on molecular-level activities was obtained by studying many molecules in bulk systems. Arguably, the reliability of the information obtained in this way is somewhat limited (Halford and Szczelkun, 2002).

Single-molecule techniques include fluorescence microscopy, optical tweezer, and atomic force microscopy. These approaches have allowed many discoveries on the underlying physical and statistical principles of the behaviour and mechanisms of biomolecules such as DNA, DNA-binding proteins and motor proteins.

Good single-molecule observations can provide evidence to support previous conclusions from theoretical studies as well as being a further method for novel discoveries. Bustamante *et al* have reviewed several single-molecule studies of DNA mechanics (Bustamante *et al*., 2000).

Not all single-molecule methods are suitable for studying protein-DNA interactions. For example, studies using hydrodynamic flow may flush away any dissociated protein from the DNA (Blainey *et al*., 2006). Studies using optical tweezer usually require the measurement of force changes and thus can be too cumbersome to be used for studying the protein translocation on DNA (Gemmen *et al*., 2006, Seidel *et al*., 2004, van den Broek *et al*., 2005).

Atomic force microscopy (AFM), especially high-speed AFM, provides further interesting possibilities for studying protein-DNA interactions.

1.5 Atomic force microscopy (AFM)

1.5.1 Principles of AFM imaging

The atomic force microscope (AFM) was invented in 1986 by Binnig, Quate, and Gerber (Binnig *et al.,* 1986). It has provided an alternative to study the sample surface topography and sample force as it does not require the samples to be coated before imaging - unlike electron microscopy or scanning tunnelling microscopy (Nelluri and Srilakshmi, 2015, Nguyen and Harbison, 2017). Since 1990, the use of AFM in both material and biological studies has steadily increased. Publications referencing AFM has increased significantly since 1990s (Last *et al*., 2010).

The ways of using AFM have been expanded and refined since its invention. The first AFM consisted of a diamond shard attached to a strip of gold foil and operated in contact mode (Binnig, *et al.,* 1986). Few biological samples were imaged under the contact mode because of the contact between the cantilever tip and the sample can cause much damage to soft samples. Non-contact mode was first introduced in 1987, with tip-sample distance of 30 -150 Å and image resolution of about 50 Å (Martin *et al.,* 1987). Because the cantilever under non-contact mode oscillated at a distance above the sample surface, imaging biological samples using AFM was possible. In 1993, tapping mode was introduced (Zhong *et al.,* 1993). In tapping mode, the cantilever gently and regularly taps the sample surface, resulting at better resolution compared with non-contact mode, and also causing little damage to soft biological samples. Tapping mode in fluid was introduced in 1994, which allowed the imaging of living biological molecules (Hansma *et al.,* 1994). AFM's use for studying biological samples has been shown in many

experiments and lauded in many reviews (For example, see Muller, 2008) Generations of small cantilevers were developed in the 1990s (Schaeffer *et al*., 1997). Smaller cantilevers usually mean higher resonant frequency, thus a higher scanning speed, and smaller spring constant, therefore the possibility of imaging further above the sample. Cantilever tips have been developed into a variety of designs for different purposes, including but not limited to sample topography scanning (Ikai *et al.*, 2018, Menozzi *et al*., 2005). The introduction of feedback loops to maintain the distance between the tip and the sample and the following improvement of the feedback loops have also been essential in improving the quality of AFM images (Mahmood and Moheimani, 2009).

1.5.2 The operation of an atomic force microscope

An atomic force microscope is usually made up of the following five components (Eaton and West, 2010) (Figure 1.5.2i):

1. A cantilever with a tip at an end (Figure 1.5.2ii)
2. A system to detect the cantilever's bending. This typically includes a laser beam generator and a laser beam detector (photodiode).
3. A feedback loop as a controller of the distance between the tip and the sample surface.
4. A movement system (i.e. piezo tubes) to enable the movement of the cantilever relative to the samples.
5. A graphical visualization system to present the data in the form of the sample topography.

A laser beam shines on the back of the cantilever, where it gets reflected and then caught by photodiode detector. The cantilever deflectors are pushed by the interatomic forces between the tip and the sample (Mahmood and Moheimani, 2009). Because the photodiode detector is position-sensitive, the information it obtains accurately reflects the level of the deflection of the cantilever, which is in turn determined by the height changes of the sample surface. Based on this relationship, the system translates the information obtained from the photodiode detector to surface topography information of the sample.

During the imaging, the signal of the positional change of the laser spot on the photodiode is passed on to the feedback loop, where it is processed and converted into a voltage to adjust the piezo's distance from the cantilever (Piontek and Roos, 2018).

The discussion of AFM operation can be found in detail in many papers (e.g. Morris *et al.*, 1999, Radmacher, 2007, West, 2007).

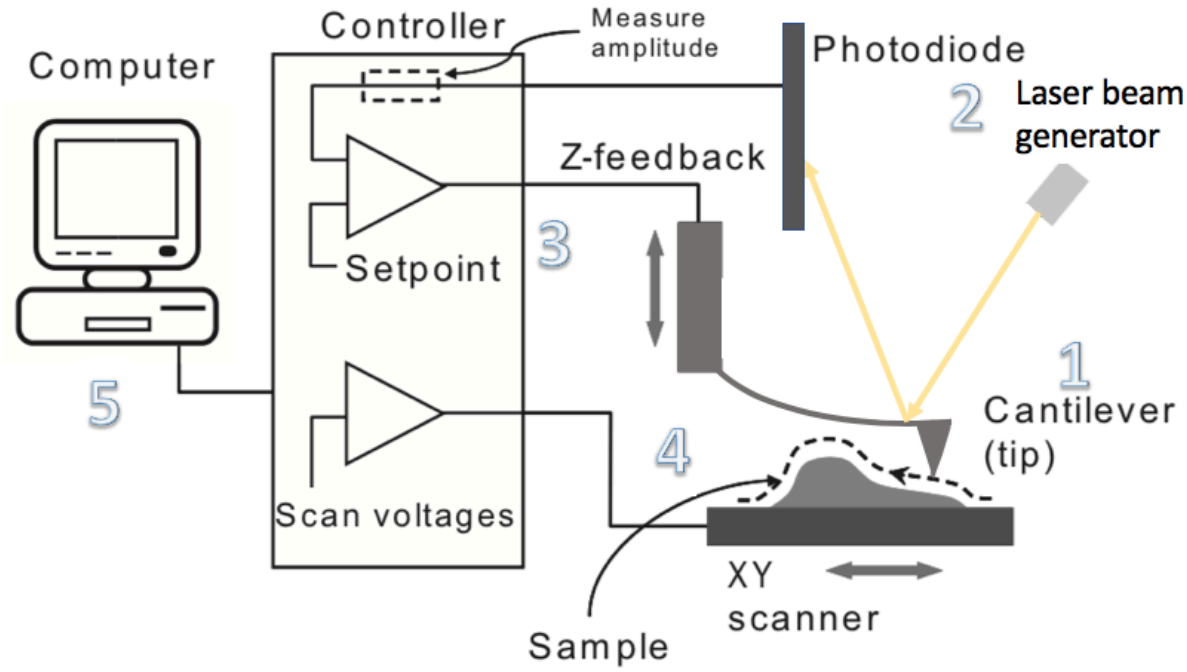


Figure 1.5.2i The components of an atomic force microscope and their functions during imaging. An atomic force microscope is usually made up of five components: a cantilever with a sharp tip at an end (marked "1" in the figure), a laser beam generator and a laser beam detector (aka photodiode) (marked "2" in the figure), a feedback loop (marked "3" in the figure), a piezo scanner (marked "4" in the figure) and a graphical visualization system to present the data (marked "5" in the figure). A laser beam shines on the back of the cantilever, where it gets reflected and then caught by photodiode. Because the level of the deflection of the cantilever is determined by the height changes of the sample surface, the information obtained from the photodiode is translated and presented by the visualization system as surface topography information of the sample. The figure is adapted from Last *et al*., 2010.

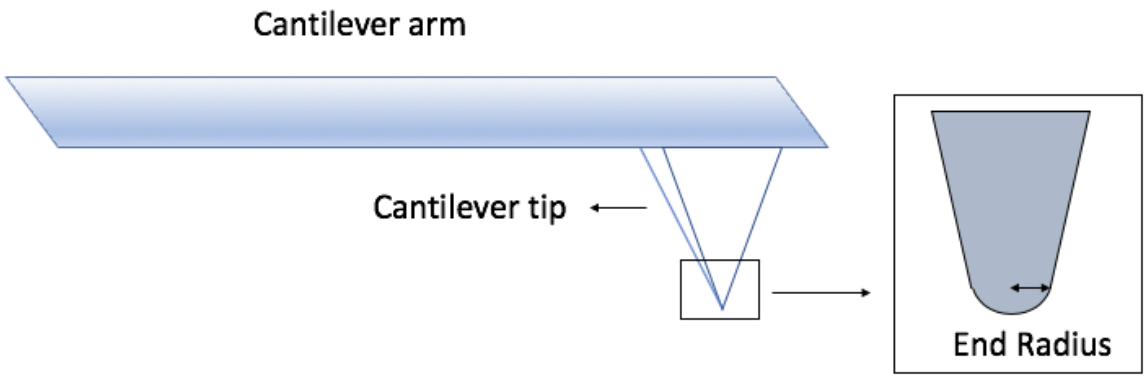


Figure 1.5.2ii A schematic sketch of the AFM cantilever. Cantilever is the part of the AFM that undertakes the task of scanning the sample. It has an arm of several up to about one hundred micrometers long. At one end of the cantilever arm there is a tip. The tip is usually pyramidal. The end radius of the tip is an important parameter in determining the sharpness of the tip.

1.5.3 Feedback loops

The cantilever moves over the sample in both the x-y plane (in a raster pattern) and along the z axis (Figure 1.5.3i). The cantilever's movement in the z axis is in most cases facilitated by feedback loops to maintain an optimal and constant distance between the tip and the sample. In doing so, in tapping mode, the feedback loop monitors the distance between the AFM cantilever tip and the sample and compares this distance with a desired distance (aka amplitude setpoint) specified by the user. It then minimises this difference so that the quality of the scanning is remained.

The feedback loop is based on the force between the cantilever tip and the sample surface. Forces applied to the samples are largely determined by the distance between the tip and the sample (Figure 1.5.3ii). The force equals to the incremental change in potential energy with respect to a change in separation and can be fitted into the force-distance curve which can be acquired by displaying the deflection of the cantilever (Piontek and Roos, 2018). The force, caused mainly by the spinning of the electrons of the tip molecules as well as the sample molecules, can be attractive ($F<0$) or repulsive ($F>0$) depending on the distance between the tip and the sample. The tip first experiences an increasing attractive force as it approaches the surface from afar. The increase in the attractive force becomes sharp when the gradient of the attractive force equals the cantilever spring constant (Piontek and Roos, 2018). When the tip is several angstroms away from the sample, the attractive force reaches its peak and starts to sharply decrease, eventually becoming a quickly growing repulsive force to push the tip away (Piontek and Roos, 2018). For a good image, it is necessary to apply the right amount of force on the tip. If too little force is applied, the topography of the sample cannot be accurately and sharply detected. If too much force is

applied, the adhesive interaction between the tip and the sample might cause irreversible deformation of the samples (Ikai *et al*., 2018).

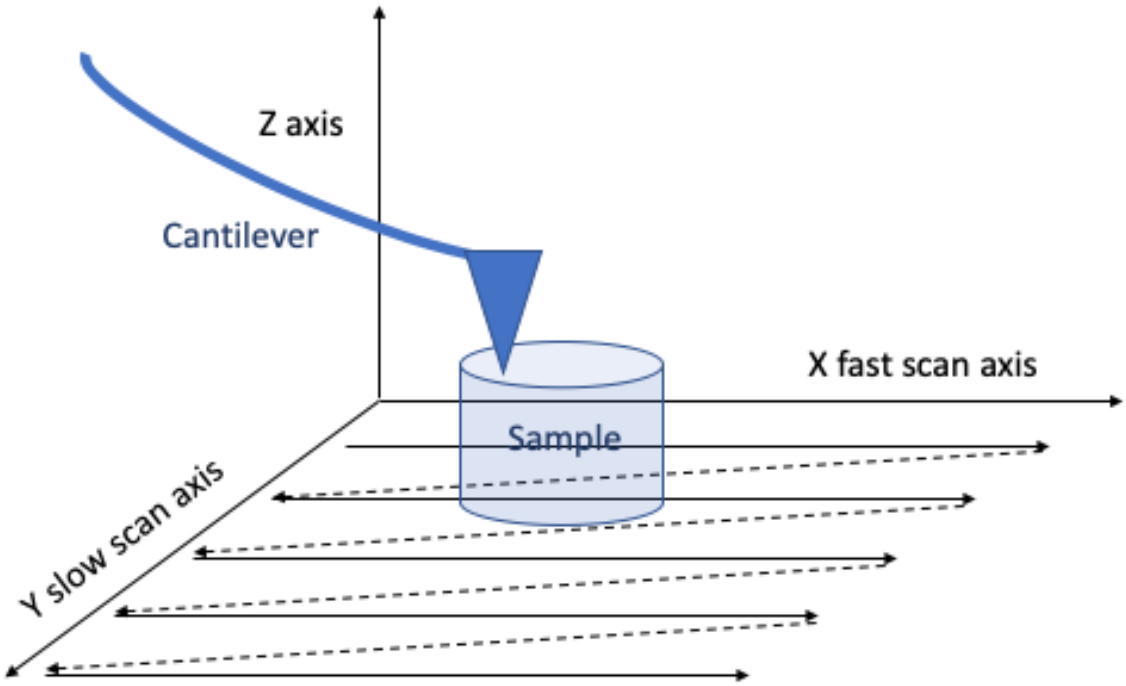


Figure 1.5.3i **The cantilever moves in the x-y plane in a raster pattern and in the z axis.** The cantilever scans in the x-y plane in a raster pattern. The cantilever moves along the x axis in 'trace' lines from left to right and 're-trace' lines in the opposite direction. Because those lines are slightly tilted downward, the cantilever also moves slowly along the y-axis. In the z axis the cantilever follows the height change of the sample.

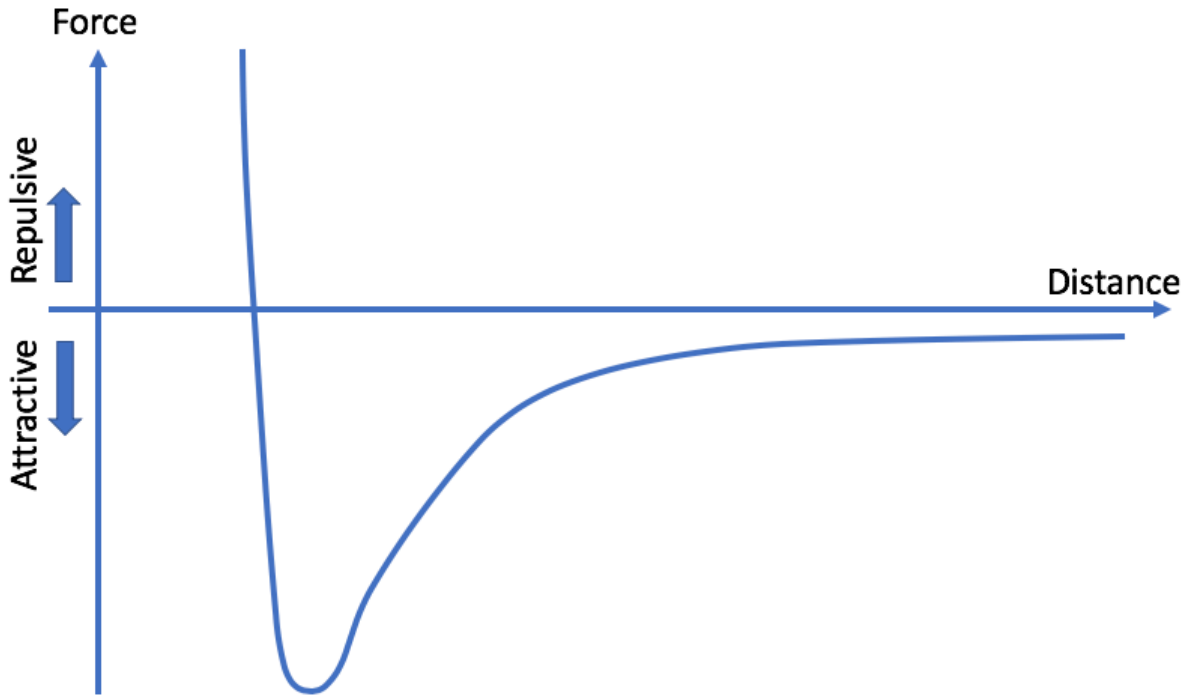


Figure 1.5.3ii **The force between the AFM cantilever tip and the sample during imaging.** The tip experiences an increasing attractive force as it approaches the surface from afar. The increase in the attractive force becomes sharp when the gradient of the attractive force equals the cantilever spring constant. The attractive force soon decreases and eventually becomes repulsive force when the tip is within a distance of only several angstroms away from the sample. The repulsive force grows quickly and pushes the tip away.

A series of problems tend to affect the scanning stability of AFM. They include Z piezo creep, thermal drift, hysteresis nonlinearity and scan-induced vibration (Mahmood and Moheimani, 2009). Z piezo creep happens when the scanner moves along the full offset distance without a consistent speed, caused often by the sudden change of the applied voltage to a piezoelectric actuator, resulting at elongated or stretched features in the direction of the offset in the image. Piezo creep has been shown to cause severe loss in precision when positioning is required over a long time-course (Fett and Thun, 1998). It can also significantly distort the images (Robinson, 1996). Thermal drift is the reduced stability of the scanner due to temperature change. Thermal drift usually occurs when the temperature of the operating environment changes. This is mostly due to the thermal expansion and contraction of the mechanical components (Mahmood and Moheimani, 2009). For example, an increase of one degree centigrade can cause 50 nm drift (Mokaberi and Requicha, 2006). Hysteresis nonlinearity is the mismatch between the input signal and output signal due to the dependence of the state of scanning on its history. Scan-induced vibration is often due to the highly resonant nature of the scanner (Mahmood and Moheimani, 2009). To achieve the raster pattern movement in the x-y plane, a triangular signal is applied to the x-axis (Mahmood and Moheimani, 2009). When the scanning speed is high, it is possible that the triangular signal excites the resonance of the scanner (Mahmood and Moheimani, 2009). As a result, the scanning on the x-axis follows a distorted triangular waveform (Mahmood and Moheimani, 2009).

Because AFM usually works at the micrometre or nanometre scale, a high degree of operating accuracy is required (Fan *et al*., 2012). The effects of the thermal and mechanical drifts of the cantilever (usually towards the sample) are non-negligible. These technical errors may reduce image quality or shorten imaging time (Fan *et al.*, 2012). Because of these issues, a lot of imaging by AFM requires accurate and highly-sensitive tip-surface distance control (Martin *et al.*, 1987, Cleveland *et al.*, 1995, Willemsen *et al.*, 2000, Ashby and Lieber, 2004). Closed-loop scanners are often used for fast and precise scanning to increase the sensitivity of the feedbacks.

The most commonly applied type of closed-loop scanner has the proportional-integral (PI) controller. During imaging, the PI controller continuously calculates the difference between a desired setpoint and the real-time variables, and it applies proportional and integral corrections to the imaging system to maintain the distance between the tip and the sample. The PI controller is effective in reducing the effect of hysteresis, creep and thermal drift (See Section 1.5.3 second paragraph for more explanation) (Mahmood and Moheimani, 2009). It is relatively weak in compensating issues such as scanner vibration, which is a result of virous internal and external disturbances, and cross coupling between axes, which is that an imaging error on one axis affects the quality of imaging on another axis (Mahmood and Moheimani, 2009).

Improving feedback loop functioning has been the subject of much research. Such attempts include optical tracking of the probe, controlling the thermal conditions and predicting the drift to compensate for motion (Tranvouez *et al*., 2007, King *et al.,* 2009, Abe *et al.*, 2007).

A recent publication suggested a new way of building the feedback loop (Fan *et al*., 2012). Fan and his group noticed that the change of the distance between the cantilever tip and the sample surface leads to changes in the cantilever damping and mass loading - The quality factor (Q) decreases with decreased tip-surface distance (Fan *et al*., 2012). They demonstrated that by maintaining a constant Q they could maintain the tip position (of soft cantilevers) within 40 nm of a setpoint in air over the course of many hours and within 3 nm in water with 95% reliability (Fan *et al*., 2012). Improved accuracy of feedback-loop is especially beneficial when studies require sensitive interactions between the tip and a sample surface over a long time-course.

1.5.4 Operational modes

AFM imaging is usually conducted in three modes – contact mode, non-contact mode and tapping mode.

The first AFM operated in contact mode (Binnig, *et al.,* 1986). In contact mode, the tip is in constant contact with the sample surface. The constant contact usually means a high resolution. Because contact mode requires the tip to be in constant contact with the sample surface, contact mode has to operate in the repulsive force regime (Piontek and Roos, 2018) (Figure 1.5.4). That means a force needs to be continuously applied to the surface to react to the strong repulsive force. The repulsive force the tip experiences is kept constant during the scan by a feedback loop (Mahmood and Moheimani, 2009). The contact mode, where the repulsive force between the tip and the sample is considerable, is normally used for hard and non-biological samples as it can easily damage soft biological samples (Mahmood and Moheimani, 2009). For example, a constant lateral force arises while scanning in the contact mode, which can drag and slide features or damage and remove features, resulting in distorted images (Piontek and Roos, 2018).

Non-contact mode was introduced in 1987 (Martin *et al.,* 1987), using a feedback loop to control the tip and sample distance by tracking the AFM cantilever vibration change caused by the attractive force between the tip and the sample (See Section 1.5.3 for more information on AFM feedback loops). In the non-contact mode, the cantilever oscillates at a small distance of 30 Å to 150 Å above the sample. The non-contact mode solves the problem of sample damaging at the expense of image resolution.

The tapping mode was introduced in 1993 (Zhong *et al.,* 1993). This largely reduced sample damage compared with contact mode. In the tapping mode, the tip touches the surface periodically during scanning and the majority of the lateral movement occurs when there is no contact between the tip and the surface (Piontek and Roos, 2018). The lateral forces, therefore, are largely reduced, resulting in better sample preservation (Piontek and Roos, 2018). For example, even with a force as large as 40 - 90 pN, with a force-acting time of 100 ns and oscillation at 1 MHz, the impulse would be only 10 to the power of -17 N per tap (Ando, 2017). As a result, since the introduction of the tapping mode, more soft biological material has been studied by AFM (Last *et al*., 2010).

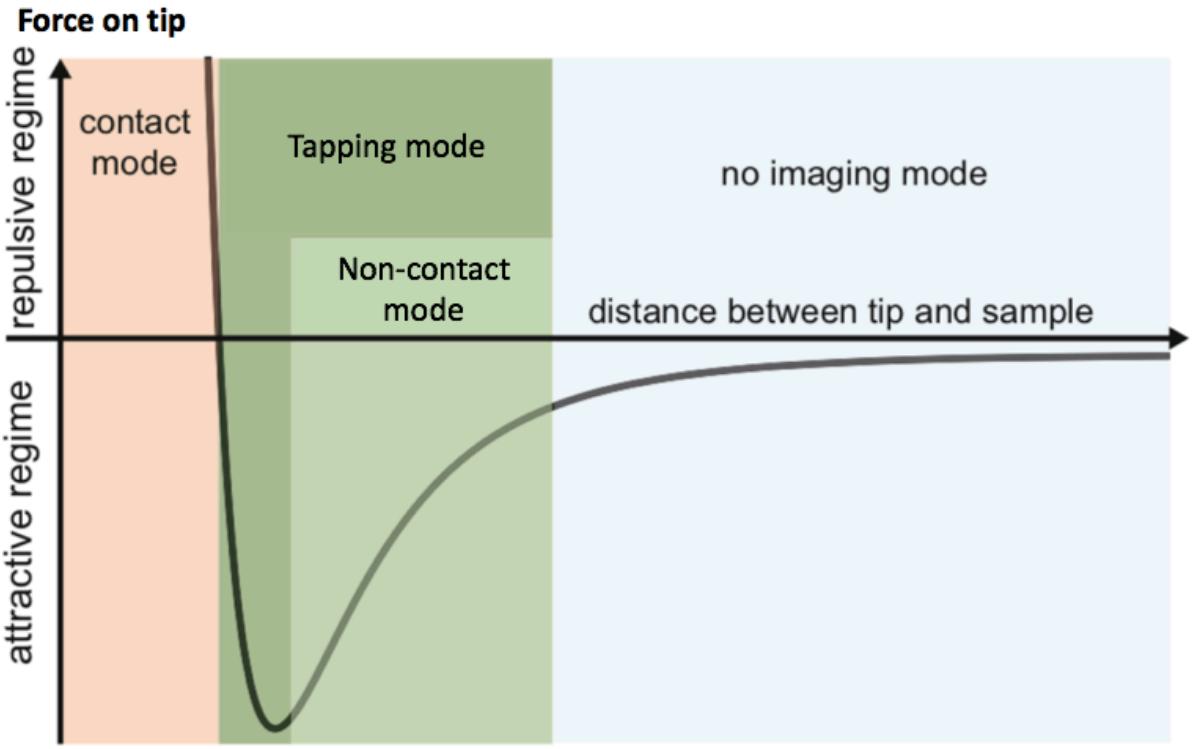


Figure 1.5.4 Three AFM modes operate in different force regimes. When operated under non-contact mode, the force between the AFM tip and the sample is in the attractive regime. Whereas under the contact mode, the force is in the repulsive regime. When operated under tapping mode, the force enters the repulsive regime only when the tip taps on the sample. The figure is adapted from Piontek and Roos, 2018.

1.5.5 Resolutions

There are two kinds of resolution - highest obtainable resolution and the smallest resolvable change. For each kind, there are both lateral (X,Y) resolution and vertical (Z) resolutions.

The highest obtainable lateral (X, Y) resolution of the AFM images is physical and is primarily determined by the tip's sharpness. The radius of the curvature of a tip is about how 'sharp' the tip is. The minute size of the sample often means the resulting image is essentially a convolution of the tip and the sample (Piontek and Roos, 2018). The smaller the tip's end radius is compared with the sample, the more features of the sample are represented in the images (Figure 1.5.5). It means that for good practice, the dimensions of the tip must be small enough for the sample, except when they are used for purposes other than visualising the topography of the sample.

The Smallest resolvable changes are in practice digital and are determined by the data points (i.e. pixels) recorded in the images. One cannot resolve features smaller than the pixel size of an image. For example, if the pixel per line is set to be 512, chosen from the normal availability of 1024, 512, 256 and 128, for a 50 μm x 50 μm image the pixel size is then 98 nm (50 μm/512 = 0.098 μm, or 98 nm). It would mean that only features that are bigger than 98 nm can be resolved in the image. If the targeted feature is, say, 10 nm, one way to meet the needs would be to reduce the image size to about 5 μm x 5 μm, resulting at a pixel size of smaller than 10 nm (5 μm/512 < 0.010 μm, or 10 nm).

The situation on the Z axis is only slightly different. The highest obtainable vertical (Z) resolution is not determined by the tip size but by the vertical scanner movement and the overall system noise. The number of data points (i.e. pixels) in the Z axis is determined by the conversion of 16 bits over the full vertical range of the scanner. The larger the Z range of the scanner is, the better the image resolution in the Z axis should be.

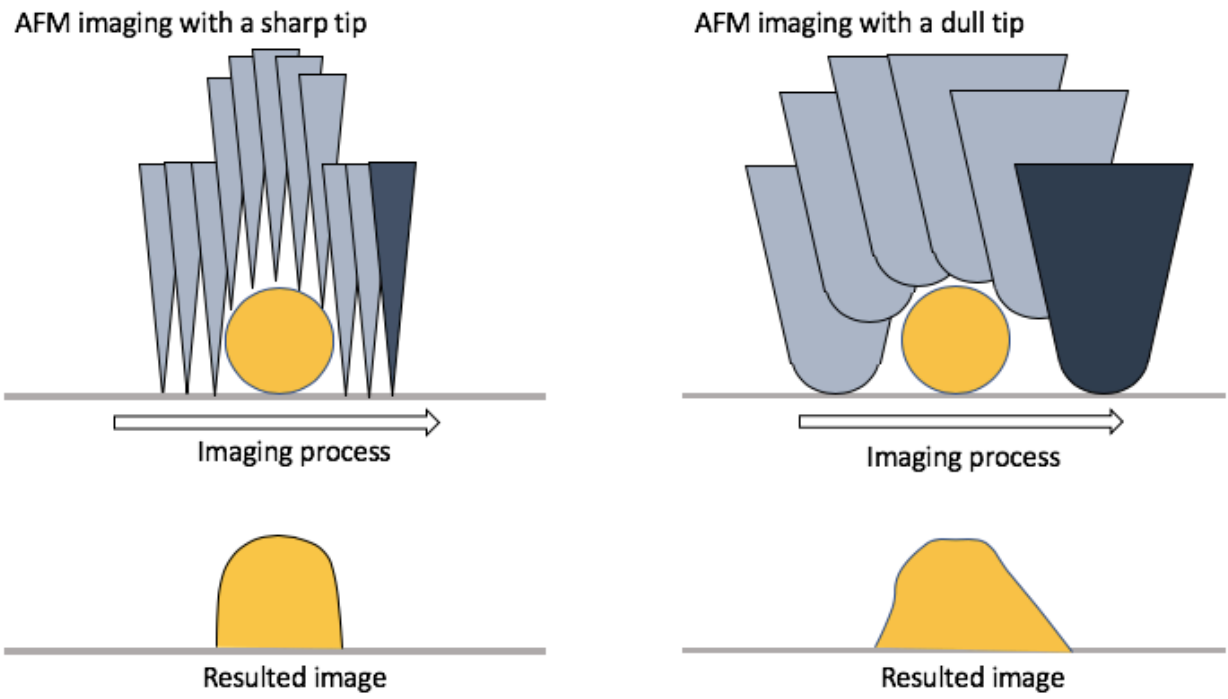


Figure 1.5.5 **The highest obtainable lateral (X, Y) resolution of the AFM images is primarily determined by the tip's sharpness.** The smaller the tip's end radius is compared to the sample, the more features of the sample are represented in the images. In contrast, imaging with a dull tip (relative to the sample) can give very limited reliable information about the sample's topography.

1.5.6 Cantilever characteristics

The cantilevers are usually made of silicon or silicon nitride (Piontek and Roos, 2018). The dimensions of the cantilevers are standardised for their application to different AFM instruments (Piontek and Roos, 2018).

Successful application of the AFM requires the right choice of both cantilever stiffness and geometry of the tip (Last *et al*., 2010). The most common geometry of a tip is pyramidal with an end radius of 20 to 30 nm (Last *et al*., 2010) (See Figure 1.5.2ii).

Recent years have seen various cases of tip modifications (Figure 1.5.6). The chemically modified tips can form covalent bonds with the sample surface (Saito *et al*., 2015). Bond forming facilitates measuring the adhesive force between the tip and the sample surface. In addition to the chemically modified tips, a variety of physically modified tips have been created for specific uses. A blunt/dull tip or a tipless probe is more suitable for evenly distributing an applied force over a wide area of the sample surface (Ikai *et al*., 2018). More sharpened tips can be used for penetrating cell membranes (Obataya *et al*., 2005a, Obataya *et al*., 2005b)

Some alternate designs of tips are used to bring the interactions between the tip and sample surface beyond basic contact. For example, a tip with a half circle shape places beads on the sample, scoops up the beads one by one, measures adhesion force between beads and the cell surface (Watanabe-Nakayama *et al*., 2010). In a more elaborate example, a microfluidic cantilever can inject liquid into the cell at controlled flow rate (Guillaume-Gentil *et al*., 2014).

Such modifications of the tips have a wide range of applications, including the manipulation of a single DNA molecule for its sequence determination without amplification, extraction of genomic DNA, and studies on the possible mechanical connection between the mitochondria and the cell membrane (Ikai *et al*., 2018, Long *et al*., 2011, Korayem *et al.*, 2014, Xu and Ikai, 1998, Silberberg *et al*., 2008, Shibata *et al*., 2015).

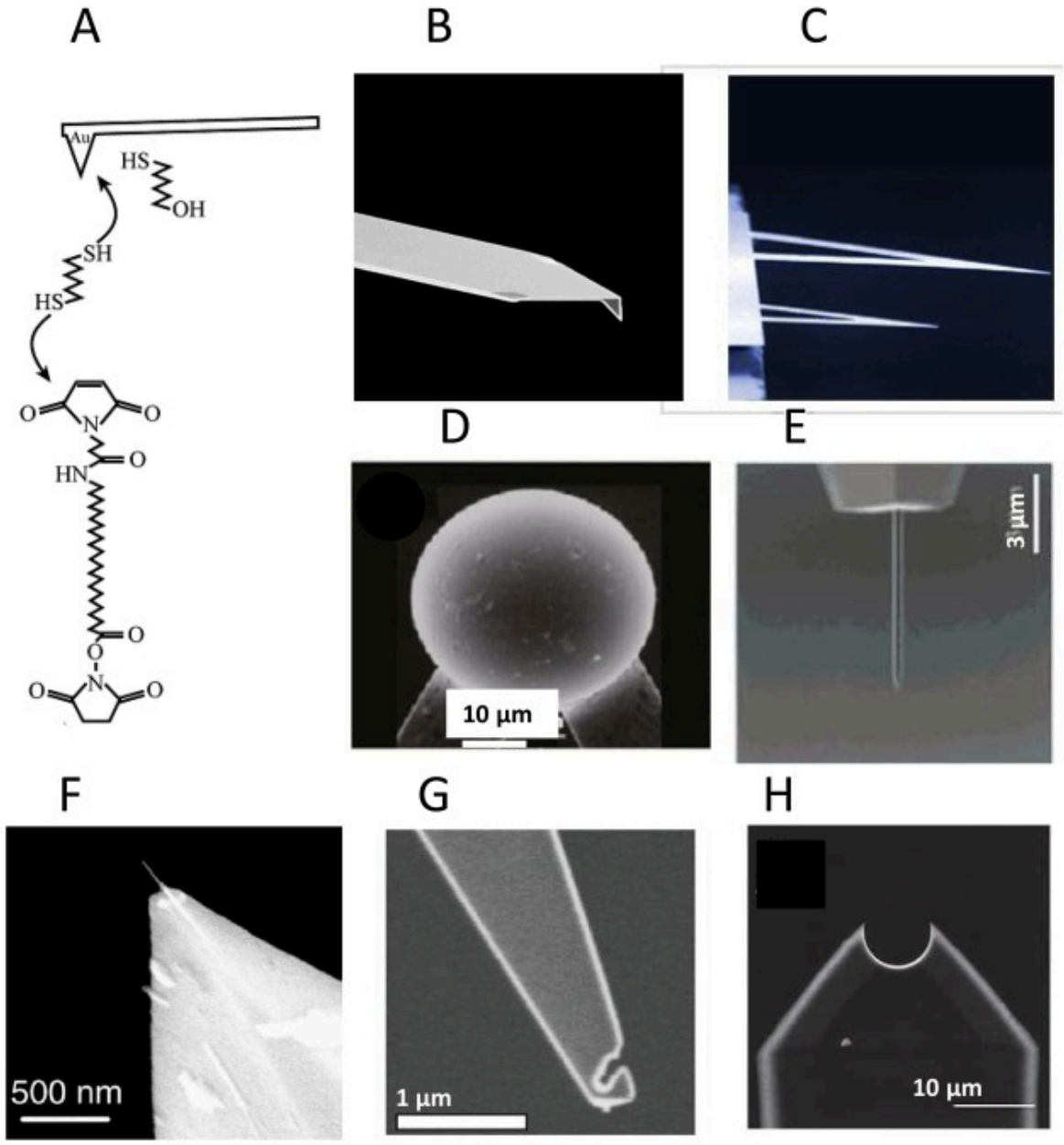


Figure 1.5.6 Several types of AFM cantilevers. (A) chemical modification of an AFM tip with MAL-PEG-NHS, (B) a rectangular cantilever with a pyramidal tip, (C) triangular tip-less cantilever, (D) cantilever with a glued colloidal bead, (E) a sharpened tip using focused ion beam technology, (F) carbon nanotube tip, (G) cantilever with a hooking capacity and (H) scooping cantilever. The figure is adapted from Ikai *et al*,. 2018.

1.5.7 Sample preparation

A widely used material for AFM sample substrate is ruby muscovite mica. It is non-conductive, and it reveals atomically flat surface when cleaved along its crystal layers (Muller *et al*., 1997). Sample solution is deposited onto the flat and smooth mica surface as soon as one new surface is revealed and left to be incubated with the mica usually for a few minutes. Following the incubation, the excess of the samples is washed off with mica deposition buffer. The mica surface is subsequently dried with nitrogen gas if the sample is for AFM dry imaging.

Because mica surface is negatively charged, an appropriate deposition buffer, which usually contains Mg^{2+}, plays a big role in the attachment of a negatively charged sample (e.g. DNA) to the mica surface. The deposition buffer is also often used to dilute the sample. However, not all positively charged ions are good components for a deposition buffer. For example, many monovalent cations such as K^+, Na^+ and Li^+ have been demonstrated to inhibit the attachment of proteins onto the mica surface (Czajkowsky and Shao, 2003). Some other divalent cations, such as nickel, cobalt and zinc ions, and transitional metal cations although proved successful as components for deposition buffer, can create too strong attachment between the sample molecules and mica surface, which might undermine the molecules' natural conformation and reaction ability (Hansma and Laney, 1996, Kirat *et al.,* 2005, Rouzina and Bloomfield, 1996, Pastre *et al.,* 2003). When strong attachment of sample molecules to mica surface is preferred and the molecules' natural conformation and reaction ability are of less importance, the newly cleaved mica surface can be coated with poly-L-lysine before the sample deposition.

1.5.8 AFM vs other forms of microscopy

The desire to see the world at the micro- to nanoscale has driven the invention of many tools. The landmark inventions include light microscope in the late 16th century, the electron microscope (EM) in 1931, scanning tunnelling microscope (STM) in 1981, and atomic force microscope (AFM) in 1986. They differ primarily in the design of their probes.

Light microscopy uses photons to probe the sample. The resolution of the light microscopy is limited to about 1 μm, primarily by the diffraction limit of photon (Berg *et al.,* 2015). Likewise, electron microscopy uses electrons as the probe; and the resolution of electron microscope is mostly limited by the diffraction limit of electron to about 10 nm (Berg *et al.,* 2015). EM has better resolution thanks to the much shorter wavelength, thus smaller diffraction limit of the electrons (Last *et al.,* 2010). But neither of the two techniques allow atomic resolution. Nor does EM work on living biological samples as the imaging has to be done on samples with conductive coating and in a vacuum environment.

The probe used by STM is much different from that used by light microscopy or EM. In STM, a metal tip mounted at the end of a cantilever is used to scan the sample surface with a constant tunnel current (Piontek and Roos, 2018). This probing system is extreme sensitive to alteration of distance and so it allows atomic resolution to be achieved (Piontek and Roos, 2018). However, this current-based probing system of STM requires both the probe and the sample to be conductive, thus limiting STM's application in biological studies (Piontek and Roos, 2018).

AFM, like STM, can achieve atomic resolution, too. Its lateral resolution is in theory unlimited, but in practice it is usually in the range of a few nanometres. The practical barrier for even better resolution in the x-y plane is the tip size and shape (Last *et al*., 2010). The z-axis resolution of the AFM is about one angstrom, limited primarily by electronic and thermal noise in the system (Last *et al*., 2010).

AFM has another major advantage besides resolution compared with EM in the study of biological samples: It can study the samples in their natural states. AFM exploits the natural bending of the cantilever pushed by the interacting force between the tip and the sample surface. It does not require the sample to be labelled, fixed or coated (Last *et al*., 2010).

And because AFM allows the visualization of a non-labelled molecule of interest, it allows direct studies of the subject (Ando, 2017). In contrast, fluorescent microscopy is constrained by the limitations of the size of the fluorescent tags because the fluorescent tags are in fact the subject of the imaging, not the samples.

1.5.9 High-speed AFM (or Fast-scan AFM)

Arguably, the most straightforward way of understanding how proteins function is to watch them with a high resolution in real time. High-speed atomic force microscopy is a strong candidate for such a task. The invention of High-speed AFM (HS-AFM, or Fast-scan AFM) has enabled imaging at the time intervals of smaller than a second, down from one image every several minutes for a traditional AFM (Ando, 2017).

For a traditional AFM, taking one image often takes several minutes, meaning the temporal resolution is often too low to study most molecular interactions in biological system (Piontek and Roos, 2018). Many refinements of AFM have been performed in order to improve the temporal resolution, mainly through improving the feedback systems and increasing the resonance frequency of the cantilevers (Ando, 2012, Piontek and Roos, 2018).

The advent of HS-AFM has expanded the application of AFM in biological studies to the observation of real-time dynamic processes (Piontek and Roos, 2018). The imaging rate of faster than 0.1 seconds per image has been shown in the studies of Ando's groups (Ando *et al*., 2008, Ando, 2012). Speeds of 10 - 16 frames per second have been reached without disturbing the functions of the proteins (Ando, 2017).

The first HS-AFM approach in biological studies probably dates back to 2001, where myosin V conformational changes were studied with a speed of 80 ms per frame (Kodera *et al.*, 2010). Examples of HS-AFM's applications to biological studies are presented in several reviews (e.g. Eghiaian *et al.,* 2014, Kodera *et al*., 2010, Nievergelt *et al*., 2015).

In summary, the high resolution and fast speed of AFM allow the nanometre scale of features of a biological sample to be observed in its native (usually liquid) environment, and in real time. This provides insights that can hardly be obtained by other techniques.

Although Ando HS-AFM has advantage over Bruker AFM on scanning speed, Bruker Fast-scan AFM can reach higher resolution and larger image scale (Ando *et al*, 2001). Therefore, in our experiment where high resolution is nexessry for tracing the proteins' locations on the DNA strands, Fast-scan Bruker AFM was used.

1.6 DNA origami

DNA origami is a form of nanotechnology allowing scientists to make 2D and 3D custom-shaped structures using DNA molecules based on the pairing principles (i.e. A=T, G=C). DNA origami has drawn much attention since the concept was first raised 36 years ago by Seeman (Seeman, 1982). It has seen rapid development in recent years since Rothemund published a relatively simple method of making structures using DNA origami (Rothemund, 2006). A number of 2D and 3D DNA structures have been made using this method, including some complicated curved and twisted 3D structures (Figure 1.6i). Software has been developed for the designing of DNA origami structures and even for the prediction of the structural properties such as flexibility of complex structures (Douglas *et al*., 2009). These developments have made DNA origami increasingly practical for building DNA structures at the nanoscale. Indeed, DNA origami has become one of the essential techniques for a series of applications at the nanoscale (Zadegan and Norton, 2012).

DNA origami requires a long single-stranded DNA (7.3 kilobase genome of the M13 bacteriophage is the most commonly used), and a large number of short DNA strands (usually synthesized oligonucleotides of 20-50 nucleotides in length). The long DNA strand is what the designed 2D or 3D structure is built upon. The short strands serve to build and stabilise the designed structure by bringing together the sequences on the long DNA strand that they are complementary to – one short strand of DNA oligonucleotides is usually synthesized to be complementary to two or more distant short sequences of the long single-stranded DNA (Figure 1.6ii). Neighbouring helices are held together by crossed-over strands. The design is usually made using software caDNAno, which was

initially developed in William Shih's lab at the Dana Farber Cancer Institute and released under the MIT license (Ke *et al*., 2009a, Dietz *et al*., 2009, Andersen *et al*., 2008).

In making the structure, a mixture of the long DNA strand and the short DNA strands in an origami folding buffer is set to go through cycles of thermal bath (which is usually rapid heating followed by slow cooling).

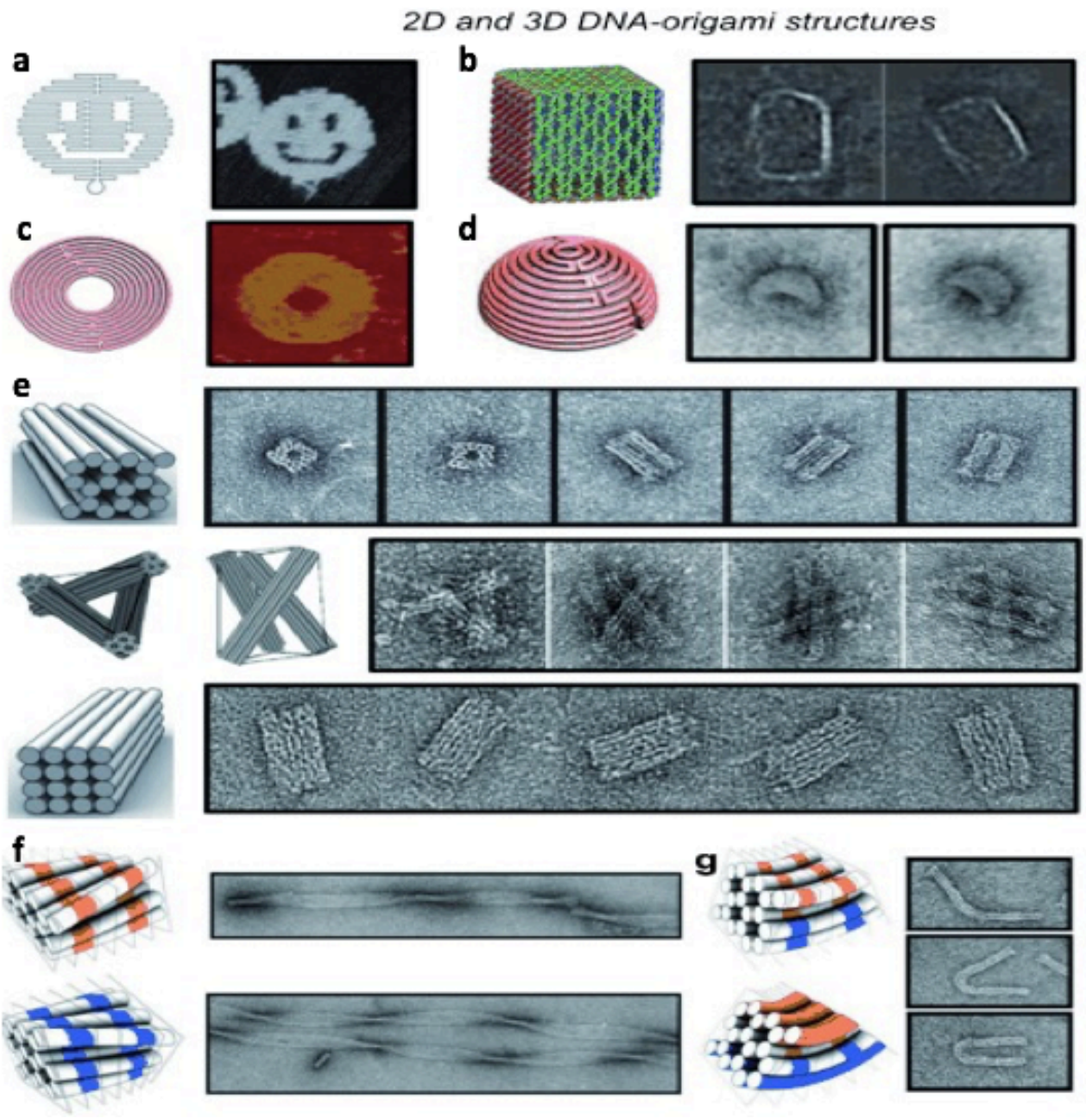


Figure 1.6i Examples of 2D and 3D DNA origami structures. From top to bottom and left to right they are: (a) 'disk with three holes' (Rothemund, 2006), (b) 'a DNA origami box' (Anderson *et al.,* 2009), (c) 'a curved 2D DNA nanostructure' (Han *et al.,* 2011), (d) 'A DNA nanostructure with complex curvature' (Han *et al.,* 2011), (e) 'three-dimensional DNA origami shapes' (Douglas *et al.,* 2009), (f) 'global left-handed/right-handed DNA origami twisting' (Dietz *et al.,* 2009) and (g) 'tuneable global bending of the DNA origami bundles' (Dietz *et al.,* 2009). The figure is adapted from Sacca and Niemeyer, 2012.

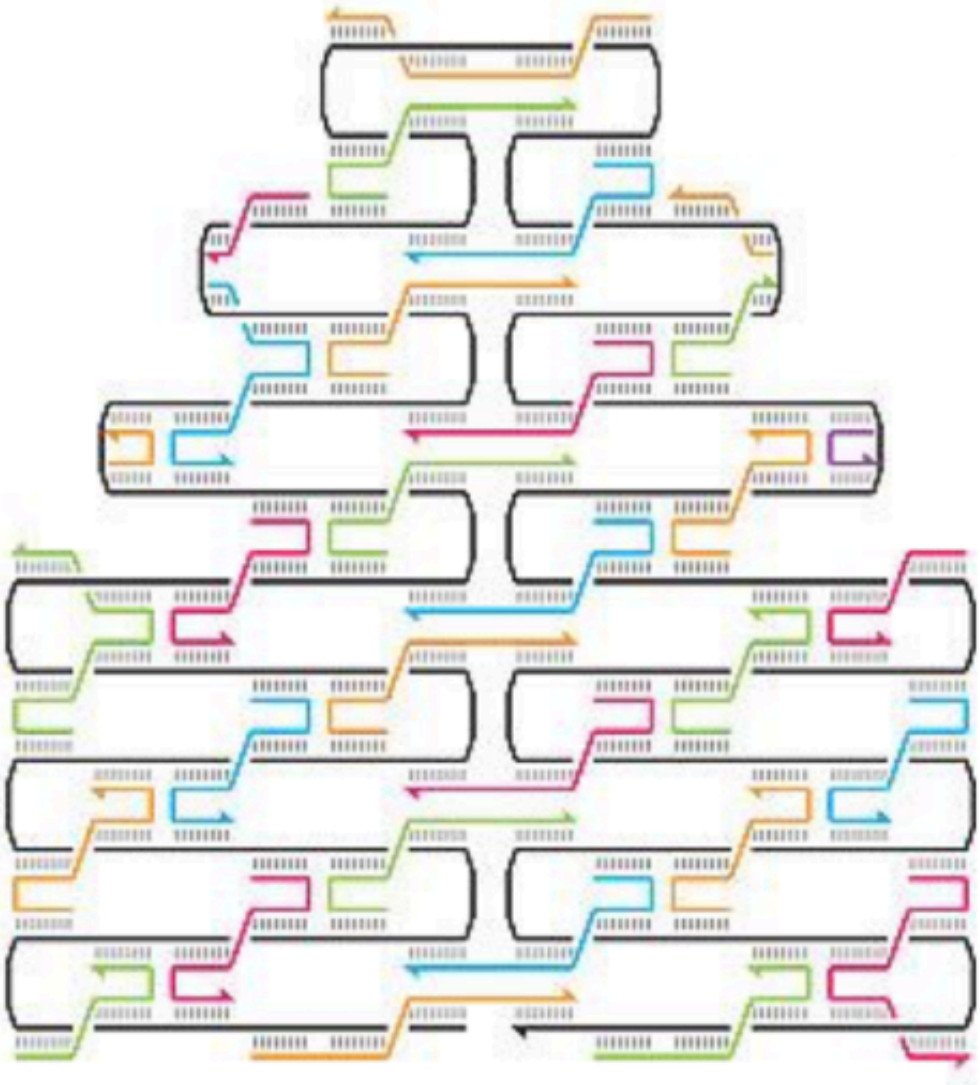


Figure 1.6ii The Rothemund method for making DNA origami structures. This DNA origami structure requires a long single-stranded DNA (black in the figure) and a many short strands of synthesized DNA oligonucleotides (coloured in the figure). The short strands serve to build and stabilise the designed structure by bringing together the sequences on the long DNA strand that they are complementary to. They also partially pair with each other. Most of the short strands of DNA oligonucleotides in this design span across three helices. The figure is adapted from Rothemund 2006.

1.7 Applications of AFM and DNA origami (respectively and jointly)

1.7.1 Applications of AFM

Since its invention in 1986, AFM has been used to study the topography of molecules at a molecular level (Binnig *et al*., 1986). More biological molecules such as proteins and DNA have been studied using AFM since the introduction of the tapping mode in 1993 (Zhong *et al*., 1993). The scope of subjects for AFM studies is still expanding. Some of its potential has been realised by imaging viruses, cells, membranes and proteins at nanometre resolution (Piontek and Roos, 2018).

More studies of biological interactions have been made possible by the introduction of fast-scan AFM. Interactions that have been successfully observed by AFM include bacteriorhodopsin responding to light, myosin V walking on actin filaments, and intrinsically disordered proteins undergoing order/disorder transitions (See this review: Ando, 2017).

Because AFM works through scanning on or over the sample surface, a newer function has been developed for it to apply force to the samples to trigger or accelerate change (Ikai *et al*., 2018). This technique involves breaking the sample with a modified cantilever tip (Ikai *et al*., 2018, Piontek and Roos, 2018). This is particularly useful for measuring local elastic modulus of a sample surface, modulus variations across a sample surface, and ligand-receptor interactions (Last *et al*., 2010). This function is extremely helpful in understanding how certain biophysical cues like substrate modules influence cell behaviour (Last *et al.*, 2010). Substrate modulus changes have been shown to affect

cytoskeletal organization, cellular orientation and alignment, proliferation and differentiation (Last *et al*., 2010). Because the hardness of all biological samples lies in a wide range of ~10 kPa up to 10 GPa according to Young's modulus, applied force should have a variety of applications (Fung 1993, Ikai *et al*., 2018). Many of the new developments involve direct mechanical manipulation of proteins, DNA/RNA, lipid-bilayers, and cells (Ikai *et al*., 2018). This facilitates better understanding of the biological samples' physical nature and in turn prompts the development of nano-medical applications.

A further expansion of AFM's abilities should be achieved by combining AFM with other techniques such as fluorescence microscopy, super-resolution optical microscopy and optical tweezers (Ando, 2017). For example, fast-scan AFM and fluorescence microscopy have been combined to study dynamic interactions in living cells (Hards *et al*., 2005, Watanabe *et al*., 2013).

1.7.2 Applications of DNA origami

After a short period of exploration by making a number of arbitrary 2D structures since Rothemund introduced the new way of making DNA origami structures in 2006 (Rothemund, 2006), researchers came to focus on DNA origami's research applications. Soon DNA origami structures were used as scaffolds to facilitate the folding of other macromolecules. For example, DNA origami scaffold was used to template protein assembly (Kuzyk *et al*., 2009, Kuzuya *et al*., 2009). DNA origami's ability to precisely orient and pattern proteins might have promising applications on constructing novel biomaterials for applications such as tissue engineering (Nangreave *et al*., 2010). DNA origami has also been used in helping arrange materials such as carbon nanotubes, silicon nanowires, or quantum dots into complex patterns (Lin *et al*., 2006, Maune *et al.*, 2010). These have built the foundation for DNA origami to be used for the synthesis of electronic circuits (Nangreave *et al*., 2010). Around 2009, the first 3D DNA origami structures, such as a monolith, a square nut, a railed bridge and a slotted cross, were made and software that could help design the 3D structures was introduced (Douglas *et al.*, 2009). Before long, bundled tubes were built with excellent control of their curve and twist (Dietz *et al.*, 2009). A controllable DNA box was also constructed, with huge potential for targeted drug delivery (Andersen *et al*., 2009). Other container structures made by DNA origami include a closed tetrahedron and a box made by selective closing of a preformed open motif (Kuzuya and Komiyama, 2009, Ke *et al*., 2009b). DNA origami structures can also be made into nano-switches with light-sensitive metal attached on (Zadegan *et al*., 2012). The technique may someday help build DNA capsules for targeted drug delivery that targets cancer cells, and some progress in this has been achieved (Jiang *et al*., 2012). There has also been success in building self-assembling and self-destructing DNA origami devices

that can remain in circulation for hours before penetrating diseased cells and releasing lethal drug payloads (Perrault and Shih, 2014). The future of the applications of DNA origami in medical, engineering and many other fields are promising. DNA origami structures have also been widely used as scaffolds to facilitate the studies of interactions between other molecules. On acting as scaffold to help studies on other interactions, 3D DNA origami multilayer shapes have huge potential because they can orient molecules such as proteins in three dimensions by allowing several cavities to be carved out of the multilayer shape (Nangreave *et al*., 2010). In the long term, as DNA origami grows more complex and more DNA strands will be needed for building one structure, the limiting factor of the scope of DNA origami's applications might be synthesizing DNA strands with high yield but low cost (Shih and Lin, 2010).

1.7.3 Using DNA origami with atomic force microscopy catalyses new biological discoveries

The highly organised and relatively inactive DNA origami scaffolds allow precise placement of molecules into the nanostructures. AFM, especially Fast-scan AFM in tapping mode, has been widely used to study living biological molecules. Naturally, then, methods using both DNA origami scaffolds and AFM have been developed to study a variety of molecular interactions. DNA origami has been made into tiles with two ligands attached to different part of the tiles to study the effect of the distance between the ligands on their interactions (Rinker *et al*., 2008). Single-molecule chemical reactions have been performed on a DNA origami and the cleavage and bond forming reactions have been observed using AFM (Voigt *et al*., 2010). One important DNA origami structure that has been used in many studies of protein behaviours is a frame – a rectangle with a large cavity (Endo *et al*., 2010b). This frame structure has been proved effective in controlling the tensions and rotations of the DNA strand(s) attached to it and has been used in a number of studies about DNA structural changes such as formation of G-quadruplex and B-Z form transition (Endo and Sugiyama, 2014, Sannohe *et al*., 2010, Xu *et al*., 2008, Rajendran *et al*., 2013c, Rajendran *et al*., 2013a). It has also been used in many studies on protein-DNA interactions. For example, DNA methylation using EcoRI, a Type I restriction enzyme, was studied using a frame-shaped DNA origami structure and AFM (Endo *et al*., 2010b). DNA repair by enzymes such as 8-oxoguanine glycosylase (hOgg1) and T4 pyrimidine dimer glycosylase was studied with a similar framed-shaped DNA origami structure and AFM (Endo *et al*., 2010a). DNA recombination was also visualised at the single-molecule level by attaching two dsDNA strands with the loxP sequence, which is recognisable to Cre recombinase, to a DNA scaffold (Suzuki *et al*., 2014). Transcription, including sliding

of T7 RNA polymerase and RNA synthesis, has also been visualized using DNA origami technique and AFM (Endo *et al*., 2012). HIV-1 nucleocapsid proteins' effects on G-quadruplex formation, have also been studied using DNA origami frames and AFM (Rajendran *et al*., 2013b).

The studies in Chapter Two and Three use the frame-shaped DNA origami structure as the scaffold and AFM as the visualisation tool.

1.8 Future directions of our studies

The development of recombinant DNA technology, which is based on enzymes that use DNA as their substrate, has transferred biology from an exclusively analytical science to a synthetic one, with new combinations of unrelated genes produced in labs all over the world. Researchers are making rapid progress using genome editing tools in areas such as biological research, medical application, and improvement of food organisms (Carroll, 2017).

The study of genome editing has come a long way. In the early days, people had no control over where mutations would happen and what effects they would produce, but they used radiation and chemical treatment to enhance the mutagenic process (Muller, 1927, Auerbach *et al*., 1947). Later, a degree of control on what the mutation would be like was obtained using transposon insertions, but the monogenetic locations were still random (Bukhari and Froshauer, 1978, Foster, 1977). The control of the site of the mutation stayed unobtainable until the discovery of the specificity of the restriction enzymes on DNA cleavage around 1970s (Kelly and Smith, 1970). Nowadays, various genome editing techniques allow directed genetic manipulations in almost all types of cells and organisms (Gaj *et al*., 2013). The key to high-efficiency genome editing is to make precise targeted DNA double-strand breaks. The three main modern techniques that can make double-strand breaks at any desired target are zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and CRISPR-Cas (Carroll, 2017). They are all originally from natural proteins: ZFNs are hybrids between a bacterial protein's DNA cleavage domain and zinc finger sequence recognition domains from eukaryotic transcription factors; TALENs is a hybrid of the same DNA cleavage domain and DNA

recognition modules from bacteria; and CRISPR-Cas is from bacteria (Carroll, 2017). Over the past decade or so, a rapid development on recombinant DNA technology has been made primarily based on using new proteins, either natural or artificial, to work on precisely-determined sequences on a DNA. The modification of genomic DNA has become highly specific. Recently, restriction enzymes have been modified to create artificial nucleases such as Zinc-finger nucleases and TAL-effector nucleases, whose recognition sites are designable (Loenen *et al*., 2014). The DNA cleavage domain of FokI is often used as a component of engineered nucleases. The other main component of such engineered nucleases is usually a transcription factor whose specificity can be artificially altered.

Much more can be done if we know more about how exactly those proteins work. Restriction enzymes are where the recombinant DNA technology is originally developed from. And there are still much more to know about them to refine the techniques.

1.9 Goals of this thesis

Gene modification *in vivo* is usually initiated by a change in the cellular environment and is facilitated by the interactions between DNA and DNA-binding proteins, such as when restriction enzymes defend the host bacteria by cutting invading viruses at specific sites. To understand the dynamics of protein-DNA interaction, it is necessary to understand how proteins translocate along DNA molecules.

Protein translocation along DNA molecules has been studied for more than 40 years. Over the years, protein translocation has been found to be far more complex than simple sliding along DNA chains, which was generally believed to be the only way of protein translocation along DNA before 1970s.

The general progression is as follows: First, qualitative studies proved the existence of two other modes of translocation besides one dimensional sliding - three-dimensional (3D) movement or hopping/jumping, and intersegmental transfer. Unlike simple 1D sliding, in 3D movement the protein does not always stay in contact with the DNA. Instead, the protein achieves the relocation by many cycles of dissociation from one site followed by re-association at another site (See this review: Halford and Marko, 2004). Intersegmental transfer is considerably more complex. It is characterised by an intermediate loop or loops formed when a protein complex concurrently binds two or more distant sites on one DNA. The protein might then cut at only one of the sites, or at all of the sites bound in the complex.

Many proteins have been found to employ both 1D and 3D movements in one round of search for their recognition sites. Today, the consensus is that most of restriction enzymes that only bind a single recognition site employ combined 1D and 3D movements to translocate along DNA. The discovery that proteins employ both 1D and 3D movements raised interest in the switch between 1D and 3D movements, their respective speeds, and the effect of the variable distance between the protein and the recognition site during the translocation on the adopting of one way over the other, among other issues. To answer these questions, more quantitative studies are necessary.

Ideally, such studies would provide fast and precise tracking of the proteins during their translocation along DNA. The nanometre-scale and second-level tracking of proteins on a thin chain of DNA had been rare prior to this decade (See the review Ando, 2012).

Intersegmental transfer has been studied in detail only recently. Gradually researchers have found that, although most of restriction enzymes used as lab-tools recognise only one DNA site and cut it afterwards, the majority of DNA-interacting proteins (including about 75% of all the restriction enzymes) are multimeric proteins that interact with multiple DNA sites (Kong *et al.*, 2000, Roberts *et al.*, 2003). In intersegmental transfer, restriction enzymes bring together multiple recognition sites before cutting at least one of them. Such restriction enzymes usually have low cutting efficiency when only one recognition site is available. The recent focus on these enzymes arises both from developments in techniques such as microscopy, and from the newfound importance of highly precise gene editing.

Genome editing tools are increasingly applied in the medical field (Tebas *et al.,* 2014, Menger *et al.,* 2016, Cyranoski, 2016, Kaiser, 2016, Carroll, 2017). However, despite

much progress in a lab environment, where the tolerance for imprecise cutting is relatively high, *in vivo* treatment remains difficult.

Safe application in medical contexts hinges on refining the tools, especially in regards to their precision. Therefore, the study of restriction enzymes is important for both better and more widespread use of the restriction enzymes, and for the development of more complex genome editing tools and their applications.

The experiments in Chapter Two and Chapter Three use Type II restriction enzymes as subjects. They are good models for studying how proteins recognise and cleave specific DNA sites.

In Chapter Two, we study in detail how proteins find their recognition DNA sites. EcoRV is the example protein used in this study because its translocation mechanism is representative for Group IIP restriction enzymes, which is the most common type in recombinant DNA technology (See Section 1.3.2). We investigate how EcoRV translocates along DNA, and measure the speed of translocation. In addition, we determine how the distance from the recognition site affects both the pattern and speed of translocation. In addressing these questions, we develop novel quantitative techniques on top of using combined atomic force microscopy and DNA origami.

In Chapter Three, we use BcgI, an untypical Type II restriction enzyme, to study how it brings together two DNA recognition sites and what complex it forms to conduct concerted cleavage at four places. The work is focused on understanding the structure of the complex and the mechanism BcgI uses to form the complex. We shall discuss why certain proteins

need multiples DNA recognition sites, how the proteins form complexes, and how the cleavages are conducted.

The studies were conducted using a combination of atomic force microscopy and DNA origami, an effective new method for single-molecule observations. Our studies demonstrate the power and versatility of these techniques.

The methods used in the experiments for EcoRV and BcgI reported in this thesis should hopefully see broader applications on studying protein-DNA interactions. They employ very powerful tools for understanding the mechanics of the protein-DNA interactions. In addition, we have shown that the combined HS-AFM and DNA origami technique can be useful in quantitative studies, as they provide high resolutions in both space and time. The experiments in this report also have demonstrated how to use the methods quantitatively by importing the AFM images to data analysis software such as ImageJ, which is an image processing software maintained by developers in Laboratory for Optical and Computational Instrumentation at the University of Wisconsin-Madison, and R, which is a programming language used widely for statistical computing and graphics and originally developed by Ross Ihaka and Robert Gentleman among others. These methods should serve to encourage better quantitative studies of protein-DNA interactions using AFM and other imaging tools.

Chapter 2 A study of 1D and 3D translocation along DNA with EcoRV as the example

2.1 Introduction

2.1.1 Aim of the present work

The aim of the work is to study the mechanisms of protein translocation on DNA using atomic force microscopy combined with DNA origami. Although the combination of these techniques was only introduced recently, it has proved effective in investigating distance-dependent multivalent binding effects (Rinker *et al*., 2008), bond cleavage and forming in chemical reactions (Voigt *et al*., 2010), regulation of DNA methylation (Endo *et al*., 2010a), DNA base-excision repair (Endo *et al*., 2010b), movement of T7 RNA polymerase (Endo *et al*., 2012), formation and distruption of G-quadruplex (Endo and Sugiyama, 2014), and many other topics (Sannohe *et al*., 2010, Xu *et al*., 2008, Rajendran *et al*., 2013c, Rajendran *et al*., 2013a, Suzuki *et al*., 2014).

The approach was used by us to study in detail how proteins locate their DNA recognition sites. As a model protein we used EcoRV, a Type IIP restriction enzyme, which is frequently used as a tool in recombinant DNA experiments. This study not only provides details regarding EcoRV's translocation, but also introduces new quantitative methods for combining atomic force microscopy and DNA origami technique.

2.1.2 Protein translocation along DNA molecules

The different modes of protein translocation

Naturally, gene regulation is usually initiated by a change in the cellular environment and is facilitated by the interactions between DNA and DNA-binding proteins, such as when restriction enzymes defend bacteria against viral infection by cutting the DNA of invading viruses at specific sites. To understand the dynamics of gene editing, it is necessary to understand the beginning of the protein-DNA interaction - protein translocation along DNA molecules in target site recognition.

Protein translocation along DNA molecules has been studied for more than 40 years. In 1970, Riggs et. al. reported that the rate that *Escherichia coli Lac* repressor protein finds its target on DNA was about 1000 times faster than simple diffusion collision would allow (Riggs et al., 1970). This rapidity of proteins finding their target sites on DNA molecules suggested more efficient searching methods than random diffusion - *facilitated diffusion* (von Hippel and Berg, 1989).

Facilitated diffusion consists of three modes of motion: one-dimensional (1D) movement or sliding, three-dimensional (3D) movement or hopping/jumping, and intersegmental transfer (Halford and Marko, 2004). In 1D movement, the protein remains in contact with the DNA chain while diffusing from the initial and usually non-specific contact site to its target site (Halford and Szczelkun, 2002) (Figure 2.1.2). In 3D movement, the protein achieves the relocation by many cycles of dissociation from one site followed by re-

association at another site (Halford and Marko, 2004) (Figure 2.1.2). The proteins may re-associate at a site near or far from the dissociation site by hopping or jumping (Halford and Marko, 2004) (Figure 2.1.2). Intersegmental transfer is characterised by an intermediate loop formation when a protein complex concurrently binds two or more distant sites on one DNA chain. Initially, the intersegmental transfer was thought to be used by only a few proteins such as *Lac* repressor and SfiI endonuclease which have two DNA binding surfaces (Halford *et al*., 2000). But recently, more proteins have been found to form complexes and use this mode to bring multiple recognition sites together for synaptic cleavage (See Chapter 3).

Both 1D and 3D movements take place and, with the development of experimental techniques, such as atomic force microscopy, fluorescent microscopy and DNA tweezer, more proteins are considered to incorporate both 1D and 3D movements. A series of recent studies have contributed to the forming of this new consensus (Elf *et al*., 2007, Tafvizi *et al.*, 2011, Hammar *et al.*, 2012, Halford and Marko, 2004, Gowers *et al.,* 2005). Most of these studies are indirect and qualitative. However, there are some exceptions. One bioinformatics study reports that the average distance on DNA covered by a protein in a single sliding event is 14 base pairs with standard deviation of around 6 base pairs (Mondal and Bhattacherjee, 2017).

We visualised the EcoRV translocation on DNA at the nanometre scale in real time and drew quantitative conclusions about the nature of the translocation. It is important to note that 'hopping' and 'jumping' are often used interchangeably in this report, and long jumps (possibly 100bp to 1000bp) are not particular concerns of this study because the length of

the DNA duplex with EcoRV recognition sites used our experiments is limited to below 200bp by the size of the DNA origami frame, to which the duplex is attached.

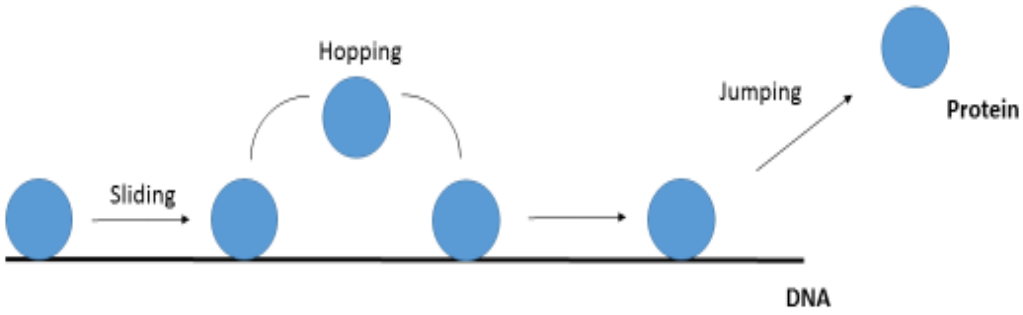


Figure 2.1.2 Schematic sketch of sliding, hopping and jumping. In sliding, the protein remains in contact with the DNA chain while diffusing from the initial and usually non-specific contact site to its target site. Alternatively, the protein achieves the relocation by many cycles of dissociation from one site followed by re-association at another site. The proteins may re-associate at a site close to (hopping) or far from (jumping) the dissociation site.

The study of protein translocation on DNA - Indirect observation

Protein translocation on DNA has attracted researchers' attention for decades. Besides the importance of studying gene regulation, it is probably also because that 3D translocation as a main method of protein translocation can seem counterintuitive and therefore was not accepted easily. Early studies suggested that 1D sliding was the only mode of motion of protein translocation. In 1976 Berg and his group used the *Lac* repressor to study protein translocation and they concluded sliding was the sole mode of motion employed by the protein (Berg and Blomberg, 1976). This was questioned later when the *Lac* repressor was found to have two DNA binding surfaces and to use segmental transfer as the primary mode of translocation along DNA molecules (Friedman *et al*., 1995, Matthews *et al.*, 1982). Subsequent studies revealed the existence of hopping and jumping, and some researchers suggested it to be the dominant mode of protein translocation for certain enzymes (Halford, 2001, Stanford *et al*., 2000).

Among those early studies, one method termed 'processivity' study was one of the most attempted (Terry *et al.*, 1985, Jeltsch *et al.*, 1994, Bennett *et al.,*1995). Despite the many derivatives of the processivity study, essentially a processivity study would involve many DNA chains each designed to have two recognition sites with various lengths of non-specific sequences in between (Stanford *et al*., 2000, Terry *et al*., 1985). The DNA products' lengths after the cleavage by the restriction enzyme would then be revealed by gel electrophoresis (Stanford *et al*., 2000, Terry *et al*., 1985). They would then be used to tell the probability of the two sites on one DNA chain being visited sequentially by one restriction enzyme (Stanford *et al*., 2000, Terry *et al*., 1985). If sliding is the only

translocation mode adopted by the enzyme, the cleavage incidences on the second recognition site should see a reduction that is in proportional to n^2 where n is the number of base pairs between the two recognition sites (Stanford *et al*., 2000). But such a relationship was not found in the processivity studies. In an experiment with EcoRV, for example, when the length in between two recognition sites was increased from 54 bp to 200 bp, 387 bp and 764 bp, cleavage incidence reduced by factors of 1.5-, 2- and 3-fold respectively, in contrast to predicted values of 15-, 50- and 200-fold based on the relationship (Stanford *et al.*, 2000). The processivity studies provided strong evidence for the existence of hopping and jumping. They also suggested that the distance covered by one sliding motion should be smaller than 50 base pairs, and hopping for less than approximately 50 base pairs would not be preferred because it was a waste of energy (Stanford *et al*., 2000).

Processivity studies are simple, but indirect and labour-intensive. Other indirect methods have yielded similar conclusions: both 1D and 3D movements, rather than 1D movement alone, are widely employed by restriction enzymes for translocation on DNA. For example, a study using DNA catenanes (two interlinked DNA rings) demonstrated the existence of 3D translocation by showing that the enzymes translocated to the second ring, which would be impossible if the only mode of translocation was sliding (Gowers and Halford, 2003). Studies using fluorescence recovery after photobleaching (FRAP) reached similar conclusions when some proteins were found translocating between the DNA binding sites via the free solution (Houtsmuller *et al*., 1999, Lever *et al*., 2000, Misteli *et al*., 2000).

The study of protein translocation - Direct observation

Indirect observations have provided evidence for the existence of 3D movement and suggested that both 1D and 3D movements are widely employed by restriction enzymes for translocation. However, many more sceptical researchers urged for evidence from direct observation (Coppey *et al*., 2004, Hu *et al*., 2006, Klenin *et al*., 2006, Lomholt *et al*., 2005, Slutsky and Mirny, 2004). Moreover, more detailed issues, like the patterns and speeds of 1D and 3D translocation, must be addressed through direct observation.

Adequate techniques for direct observation have become available. In 2008 Bonnet's group used fluorescence microscopy to directly observe EcoRV translocation on DNA. In their experiment, jumps larger than 200 nm (about 600 bp) were observed with a precision of 30 nm (Bonnet *et al*., 2008).

Direct observation improves on indirect methods, but the precision is often far from ideal. The ideal resolution for direct observation should be comparable to the size of the restriction site (i.e. 6 bp, or about 2 nm, for EcoRV and most Group IIP restriction enzymes). However, fluorescent particles, such as the ones used in Bonnet's experiment in 2008, are normally 10-50 nm in diameter and so imaging with fluorescent particles can hardly yield a resolution finer than 10 nm (Halford and Marko, 2004). (In the past few years, several small (<5 nm in diameter) fluorescent tags have been introduced (Plamont *et al.*, 2016).) This explains why Bonnet's group observed only long jumps, and could not distinguish short hops from sliding. They admitted that hops were difficult to distinguish from sliding due to the resolution issues and speculated that small jumps shorter than 200

nm were 10 times more likely to happen than big jumps (Bonnet *et al*., 2008). Moreover, the DNA could not be seen, so the contact between the enzyme and the DNA was supported by calculation using mean square displacement (MSD), a statistical method. The MSD results were positive about the contact of EcoRV and the DNA duplex before and after EcoRV jumping (Bonnet *et al*., 2008). However, in another direct observation of the translocation of EcoRV on DNA, using EcoRV labelled with quantum dots, DNA manipulated by double optical tweezers and MSD for trajectory analysis, sliding was suggested as the predominant way (Biebricher *et al*., 2009). Therefore, a way that provides high enough resolution for direct observation of the EcoRV and DNA contact is necessary.

Atomic force microscopy provides a way to address these issues. The study discussed in this chapter, about the interactions between EcoRV and DNA, was conducted using fast-scan atomic force microscopy with the help of DNA origami. Using this method, a resolution of close to 1 nm was achieved. Whether the protein was on or off the DNA could be easily discerned from the images. The high resolution and fast speed allowed a large amount of quantitative information to be collected from measuring the protein locations on the images. This in turn allowed detailed location and travel distance analyses on EcoRV-DNA interactions. In addition to providing more insight on protein translocation, including the most frequent travel distances and how the closeness to a recognition site affect the modes of EcoRV's translocation, this work sets a template for future quantitative studies of similar subjects.

2.1.3 EcoRV

One of the better studied Type II restriction enzymes (See Section 1.3.2) is EcoRV, which has a recognition site of 6 base pairs, GAT|ATC. The cleavage is achieved in a blunt-ended fashion at the location marked by '|', with both DNA strands terminating at the same base pair (Winkler *et al*., 1993). Mg^{2+} is required for cleavage.

The gene coding for EcoRV has been sequenced and, like that of many other Type II restriction enzymes, it has a short consensus sequence, Pro-Asp-Xaa10-20-(Asp/Glu)-Xaa-Lys, which forms the active site for the phosphodiester hydrolysis reaction (Aderson, 1993, Venclovas *et al*., 1994). Crystal structures of three states of EcoRV are available. They are the structures of a free EcoRV enzyme, EcoRV bound to non-specific DNA sequence, and EcoRV bound to its recognition site GATATC (with Ca^{2+} instead of Mg^{2+} in the buffer to prevent cleavage) (Jeltsch and Pingoud, 2001, Pingoud and Jeltsch, 2001, Vipond and Halford, 1995, Winkler *et al*., 1993). Based on the crystal structures, a three-step mechanism of EcoRV reaction was generated and generally accepted (Zahran *et al*., 2010):

Step 1- EcoRV binds the DNA molecule at probably a non-specific site and starts searching for its recognition site.

Step 2- EcoRV arrives at its recognition site and recognises the outer base pairs -- GAxxTC.

Step 3- The flexibility of the central TA base pairs and their recognition by hydrophobic contacts allows for the 50-degree bending of the DNA which is followed by the cleavage at in between the T and A.

Unlike many other Type II restriction enzymes that contact all the base pairs of the recognition sites, EcoRV only directly contacts the first and last 2 base pairs ("GA" and "TC") of its 6 bp recognition site (Winkler *et al*., 1993). This 'indirect readout' of the central base pair "TA" probably based on the difference on the energetic cost of partially unstacking the "TA" – a kink of 50 degrees was found at the central base pair of "TA" in the EcoRV-recognition DNA sequence crystal structure (Horton and Perona, 1998, Martin *et al*., 1999).

The experimental design has been refined based on these features of EcoRV.

2.2 Materials and methods

2.2.1 The design of the DNA origami frame structure

The DNA origami frame design is made using CadNano, a software initially developed in William Shih's laboratory at Dana Farber Cancer Institute to simplify and enhance the process of designing DNA origami nanostructures and has since been updated several times (CadNano 2014). The DNA origami frame is built using M13mp 18ss DNA as template (250 μg/ml from New England Biolabs) and short DNA oligomers as stabilisers (from Integrated DNA technologies) to shape the template into the designed frame structure. The frame surrounds an empty space of 190 bp long and approximately 30 bp wide. Two 16 bp overhangs intrude from opposite sides of the frame into the empty space. The design is demonstrated in Figure 2.2.1. The M13mp 18ss DNA is shown in the figure in dark blue and the DNA oligomers in multiple other colours.

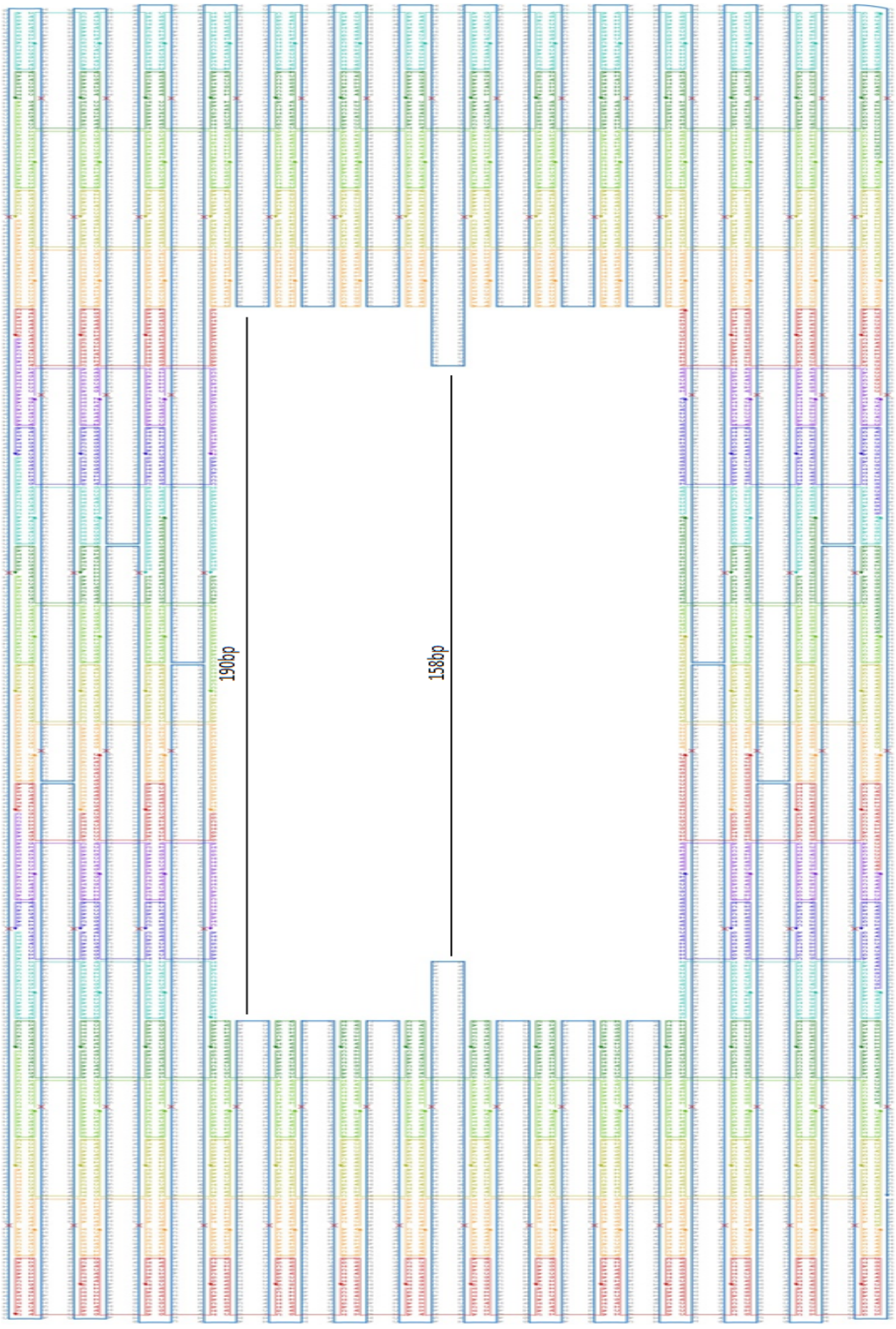


Figure 2.2.1 The DNA origami frame design. The frame surrounds an empty space of 190 bp long and approximately 30 bp wide. Two 16 bp overhangs intrude from opposite sides of the frame into the empty space. The DNA origami frame is built using M13mp 18ss DNA as template. The M13mp 18ss DNA is shown in the figure in dark blue and the DNA oligomers are in multiple other colours.

2.2.2 The DNA duplex that is designed to attach to the DNA origami frame

A DNA duplex is created by two complementary single strands, each of which has a 16 bp sequence (as shown in bold in the sequence below) complementary to one of the two overhangs from the frame structure (See Section 2.2.1 for the frame structure). The result is a duplex of 175 bp attached to the middle of the frame.

The DNA duplex has two EcoRV recognition sites. The first EcoRV recognition site GATATC is 34 base pairs away from the left overhang and the second GATATC site is 48 base pairs from the right overhang. The distance between the two restriction sites is 81 base pairs. The direct observation prefers a smooth process without permanent modification of the chemical structure of the DNA molecule. One can achieve the smoothness by excluding target sites of EcoRV from the DNA molecule. But some criticised 'diffusion to capture' is not the same as 'diffusion' itself, and it is necessary to include the recognition sites in the DNA sequence to study the translocation of the protein to its target site (Halford, 2009).

The duplex is designed to be slightly longer than the distance between the two overhangs – which is 158 bp - so the duplex provides enough flexibility for the EcoRV to search on the DNA without imposed tension by the artificial DNA structure. On the other hand, the length of the duplex has to be designed to be comparable to the distance between the overhangs for two reasons. Firstly, it prevents cleavage of the restriction sites, for the duplex is not long enough for the formation of the 50-degree kink required for EcoRV cleavage. As mentioned in the Introduction, a sharp kink of 50^{o} has been found at the

central base pairs TA in the EcoRV-recognition DNA crystallographic complex (Winkler *et al*., 1993). This kink is not seen in other state of reaction, suggesting the 50^{o} is necessary for cleavage, but not for site recognition (Zahran *et al*., 2010). In this way, the direct observation of 'diffusion to capture' is made possible and the results would not be biased by enzyme cleavage kinetics, which often affected previous biochemical studies (Halford and Marko, 2004). This experimental design provides a solution for the challenge of involving the target sites of the enzyme in the DNA duplex, yet maintaining a smooth process for direct observation without permanent modification of the DNA. Secondly, the duplex's comparable length to the length of the middle space of the DNA frame structure facilitates the measuring of EcoRV's locations on the duplex with the reference to one inter side of the DNA frame structure during the AFM image analyses.

Single DNA strands that form the duplex are shown below. The sequences in bold are complementary to the two overhangs intruding into the middle space from the DNA origami frame structure. The underlined sequences are the EcoRV restriction sites GATATC.

Forward strand

GGT GTG TGG AAG TTT GTT TTT AAC GGA TCC GTG CCC ATT TCA GTT AAC GAG ATA TCT GCA CGT TAC CAG GGG ATG TTA GGG ATC TTA CGC AGT GTA CCG AGA GAG AGA AAC GGA TCC GTG CCC ATT TCA GTT AAC GAG ATA TCT GCA CGT TAC CAG GGG ATG TTA GGG ATC TTA CGC AGT GTA CCG TTT TT (191)

Reverse strand

CAA ATT CTT ACC AGT GAA AAA CGG TAC ACT GCG TAA GAT CCC TAA CAT CCC CTG GTA ACG TGC AGA TAT CTC GTT AAC TGA AAT GGG CAC GGA TCC GTT TCT CTC TCT CGG TAC ACT GCG TAA GAT CCC TAA CAT CCC CTG GTA ACG TGC AGA TAT CTC GTT AAC TGA AAT GGG CAC GGA TCC GTT AAA AA (191)

2.2.3 Experimental procedure

2.2.3.1 The dry imaging procedure

The EcoRV stock was diluted from 75 nM to 0.075 nM using lab-made origami folding buffer (20mM Tris-HCl, 10mM $MgCl_2$, 1mM EDTA, PH7.6, 1x) to prepare dry samples for imaging in air. 40µl of this diluted solution was pipetted onto a newly revealed ruby mica surface. The loaded mica disc was left at room temperature for 2 minutes before the surface was washed by biotechnology performance certified (BPC) water (from Sigma-Aldrich). The loaded mica was then dried using nitrogen gas and stored in a silica gel container overnight for further drying. Then the samples were taken out of the silica gel container to be observed under AFM in dry imaging tapping mode. The volume of the particles in the images were then measured and compared with the volume of the particles in the images taken under AFM fluid imaging tapping mode. The model of all the cantilevers used for the dry imaging was Fastscan-A from Bruker, a manufacturer of scientific instrument based in Massachusetts, USA. Fastscan-A is a triangular Silicon Nitride cantilever with 1400 kHz resonant frequency and 17 N/m force constant. Its Silicon tip has 5 nm tip radius.

2.2.3.2 The Fluid imaging procedure

All the DNA materials were diluted using lab-made origami folding buffer (20 mM Tris-HCl, 10mM $MgCl_2$, 1mM EDTA, PH7.6, 1x), which is also used as the mica deposition buffer for the imaging. To form the DNA origami frame, 2 µl of the M13mp 18ss DNA

(10 nM), 2 µl of 10x origami folding buffer, 5 µl of the DNA oligomers (200 nM) and 11 µl of biotechnology performance certified (BPC) water (from Sigma-Aldrich) were mixed and put to undergo a thermal bathing from 80 °C to 25°C, descending 1°C every minute. The formed frames were stored at 4°C.

To form the duplex, 2 µl of the forward strand (100 µM), 2µl of the reverse strand (100 µM) and 16 µl of 1x origami folding buffer were mixed and put to undergo a thermal bathing of 90°C for 5 minutes which then descended at 1°C per minute to 25°C. The formed duplexes were put to undergo electrophoresis in a 2% agarose gel and Tris-acetate-EDTA (TAE) buffer at 110 V. The gel was then put in a plastic holder with 100 ml of deionised water and 6 µl of ethidium bromide and was shaken gently for 30 minutes and then in another 100 ml of deionised water without ethidium bromide for another 30 minutes. The gel was then illuminated under UV light and the target band of about 200 bp was cut off from the gel pad. The target DNA duplexes of about 200 bp were then extracted from the band using QIAquick gel extraction kit from Qiagen.

The obtained duplexes (about 60 µl) were then mixed with the formed frames (20 µl) and put to undergo a thermal bathing of 40°C to 15°C, descending 0.5°C every minute. The resulting products were put to undergo electrophoresis in a 0.5% agarose gel and 1x TAE buffer with 10 mM $MgCl_2$. The dyeing and illumination steps were the same as described above. The band of the size of the target DNA products (duplex attached to the frame) was then cut off from the gel pad and further purified using a Freeze and Squeeze purification kit from Bio-Rad. The 'Freeze and Squeeze' purification method was developed based originally on Thuring's theory (Thuring *et al*, 1975). It requires freezing of the sample for

5 minutes at -20°C and then spinning for 3 minutes at a rate of 13,000 rpm. The spinning rate was lowered to 6,000 rpm for this experiment after several trials. Sephacryl S-400 purification from GE Healthcare with elution limit of 271 bp was also used to purify the mixture.

For AFM observation, each time, 40 µl of the frame + duplex complex solution was attached on a mica disc and subsequently put under the AFM. Depending on the number of the frame and duplex complexes in one AFM image, the frame and duplex solution could be concentrated using Amicon Ultra centrifugal filters from Merck Millipore before being pipetted onto the mica disc.

Fluid imaging and data collection

The images were generated using a Bruker Dimension Fast-scan AFM from Bruker, a manufacturer of scientific instruments based in Massachusetts, USA, in the tapping in fluid mode. The model of all the cantilevers used for fluid imaging was Fastscan-DX from Bruker. Those cantilevers have silicon tips and have a force constant of 0.25 N/m and resonant frequency of 110 kHz. The tip radius is 5nm.

During the imaging process, 1 ml of 0.075 nM EcoRV (diluted using 1x origami folding buffer) was introduced into the system through a 1 ml syringe attached to the AFM scanner. The continuous image capturing was initiated right after a good capture area was discerned at a 500 nm x 500 nm scale. A good capture area should include:

1) an intact DNA frame

2) a clearly visible DNA duplex firmly attached to the DNA frame, and

3) an active EcoRV moving along the DNA duplex.

The imaging of a particular capture area was terminated when the DNA frame was worn out after repetitive scanning, when the DNA duplex broke off from the DNA frame, or when the EcoRV went out of the frame. The locations of the EcoRV at each image-capturing time point was recorded in time series images.

The above steps were repeated until the whole mica surface was searched through by the AFM, resulting in groups of images of EcoRV's location on the DNA duplex during translocation.

The above procedures were repeated until enough data was collected for statistical analyses.

2.2.4 Data analyses of the time series images

The processing of the time series images was conducted using Nanoscope software, Bruker's data processing software, and ImageJ, an image processing software maintained by developers in the Laboratory for Optical and Computational Instrumentation at the University of Wisconsin-Madison. Microsoft Excel was used to store the EcoRV location data. R, a programming language used widely for statistical computing and graphics and originally developed by Ross Ihaka and Robert Gentleman among others, was used to analyse the data and draw the plots.

The raw AFM images were flattened using Nanoscope. The flattened images were imported into ImageJ and aligned in stacks using the "Align slices in stacks" function. For each of the stacked images the orientation of the DNA origami frame was adjusted using the "Rotate" function to ensure superimposition. The locations of the EcoRV on each image was tracked manually under the "MTrackJ" function. The locations were measured as the distance between the EcoRV and one side of the DNA origami frame. The location information was then stored using Excel and further analysed and visualised with R. The R scripts for each R plots are included in the Appendix for verifying and regenerating the results.

2.2.5 The difficulties of capturing the AFM time series images of this delicate system

As mentioned in Section 2.2.3, a good area for image capturing usually includes one intact DNA frame in a good rectangular shape, a clearly visible DNA duplex firmly attached to the DNA frame, and an active EcoRV movement along the DNA duplex. The image capturing of one particular area was terminated whenever the DNA frame was disfigured, the DNA duplex detached from the DNA frame or the EcoRV no longer remained in the frame (At the beginning of the imaging, about 60% of the frames had the duplex attached).

At the rate of two seconds per capture, producing enough high-quality images for statistical analyses required continuous imaging of the same area for several minutes. A clearly visible DNA duplex firmly attached to the DNA frame for the whole duration of repetitive scanning was rare. As a result, a large number of attempts were made.

2.3 Results and discussions

2.3.1 Volume analysis

For the AFM fluid images, each sample was made by constructing a DNA frame, attaching the designed DNA duplex to the frame and introducing EcoRV to the system for imaging. Purification results of the DNA duplex (Figure 2.3.1i), and AFM fluid images of the DNA frame alone and with the duplex are shown below (Figure 2.3.1ii).

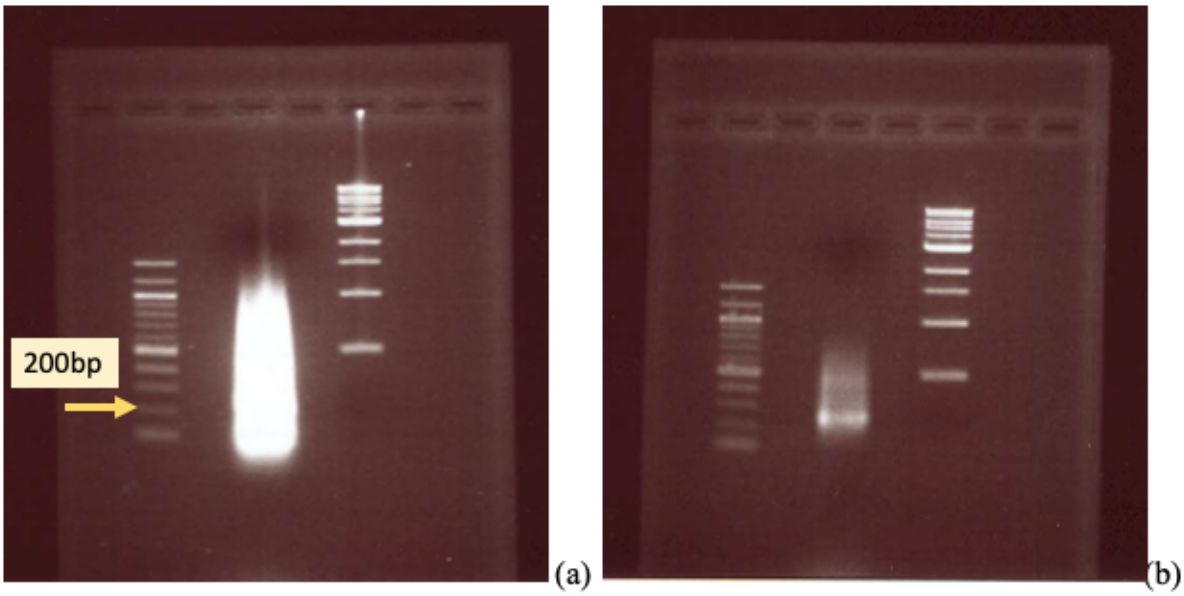


Figure 2.3.1i Purification results of the DNA string with two EcoRV recognition sites. From (a), a gel segment containing the DNA products of approximately the size of the designated duplex (about 191bp) was extracted, redissolved and run on another agrose gel (b) where clear and narrow DNA bands of approximately the size of the duplex was observed. The duplex was then attached to the DNA frame.

Figure 2.3.1ii AFM fluid images of the DNA frame alone and with the DNA duplex attached to it.

After EcoRV was introduced to the system, bright spots were observed on (and occasionally by the side of) the DNA duplex in the continuously captured AFM real-time images. To confirm that these were EcoRV, we performed a volume analysis.

The reference volume of EcoRV was calculated based on the following equation:

$$Vc = \left(\frac{Mo}{No}\right)(V1 + dV2)$$

Where M_0 is the molecular weight of the protein (i.e. EcoRV in this case), N_0 is Avogadro's number. Here, V1 and V2 are the partial specific volume of particle (0.74 cm^3/g) and water (1 cm^3/g) respectively; and d is the level of hydration (normally 0.4 g water/g protein) (Schneider et al, 1998). According to data provided by New England Biolabs, which is the supplier of the EcoRV used in this experiment, the molecular weight of EcoRV is 28.5 kDa. Therefore, the reference volume of EcoRV is 54 nm^3.

The sample preparation process of the dry images (See Section 2.2.3.1) means that only the bright spots are shown in the dry images (Figure 2.3.1iii, see the appendix for more images like this). The volume of the spots in the dry images was measured using the Scanning Probe Image Processor (SPIP) software, a software developed by Image Metrology for processing and analysing microscopic images. The mean across 2897 of the bright spots is 53 nm^3, with a standard deviation of 19 nm^3. The histogram of volume measurements can be found in the Appendix.

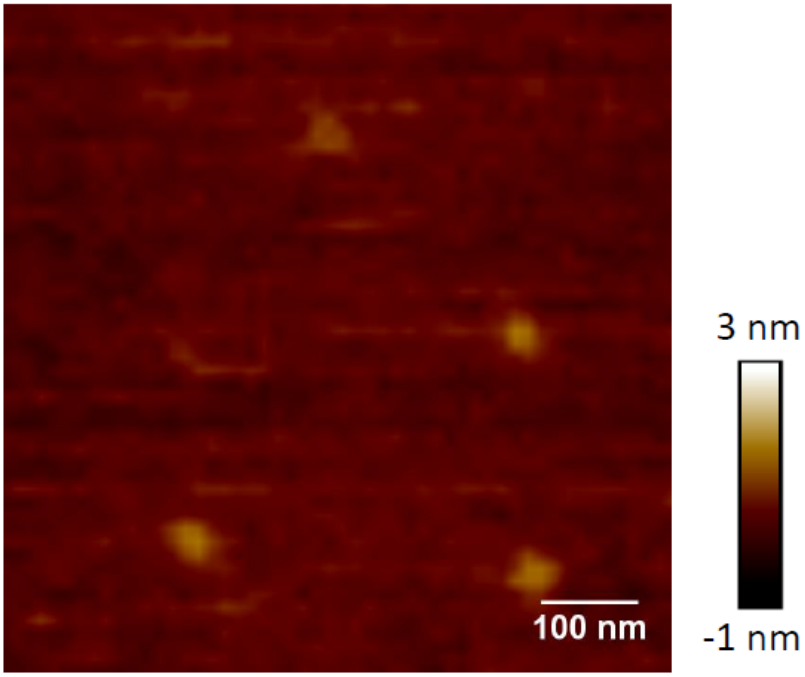


Figure 2.3.1iii A sample AFM dry image (i.e. taken with the sample in air. Subjects shown as bright spots are easily discernible from the clear background. The sample was prepared by diluting the EcoRV stock from 75 nM to a concentration of 0.075 nM using origami folding buffer (20 mM Tris-HCl, 10mM $MgCl_2$, 1mM EDTA, PH7.6, 1x), which itself has no discernible particles under AFM. The scan rate is 18.6 Hz. The resolution is 512 x 512 pixels. The amplitude setpoint is 320 mV. The drive amplitude is 40 mV.

In the fluid images, many of the bright spots are on the DNA duplexes (Figure 2.3.1iv, see the appendix for more images like this, and Figure 2.3.1v). Consequently, their outlines could not be automatically detected by SPIP. In these cases, SPIP was used to measure the basal radius (r) and height (h) of the proteins manually. After obtaining those variables, the volume of the molecule was calculated as a segment of a sphere using the following equation (See Neaves *et al.,* 2009 for a similar calculation of particle volume in AFM images):

$$Vm = (\frac{h}{6})(3r^2 + h^2)$$

108 volume results were obtained. After removing two outliers, the measurements showed a mean volume of 56 nm^3, with a standard deviation of 11 nm^3. Figure 2.3.1vi shows a histogram of the volume results. The R script of this volume analysis can be found in the Appendix. The outliers are stated in the data summary and shown in the box-plot in the R script.

An independent Samples T-Test Assuming Different Variances (which is included in the Appendix) was used to compare the mean of the particle's volume in the fluid images with the mean volume in dry images. And the 95 percent confidence interval of the mean of the particle's volume in the fluid images is 54.0 nm^3 to 58.2 nm^3. The calculated reference volume falls into this interval. Statistical tests verified that the volume mean of the bright spots from the fluid images is statistically similar both to the volume mean of the bright spots in the dry images and to the reference EcoRV volume value calculated based on its molecular weight. Together, these results strongly suggest that the objects observed and tracked in the fluid images are EcoRV.

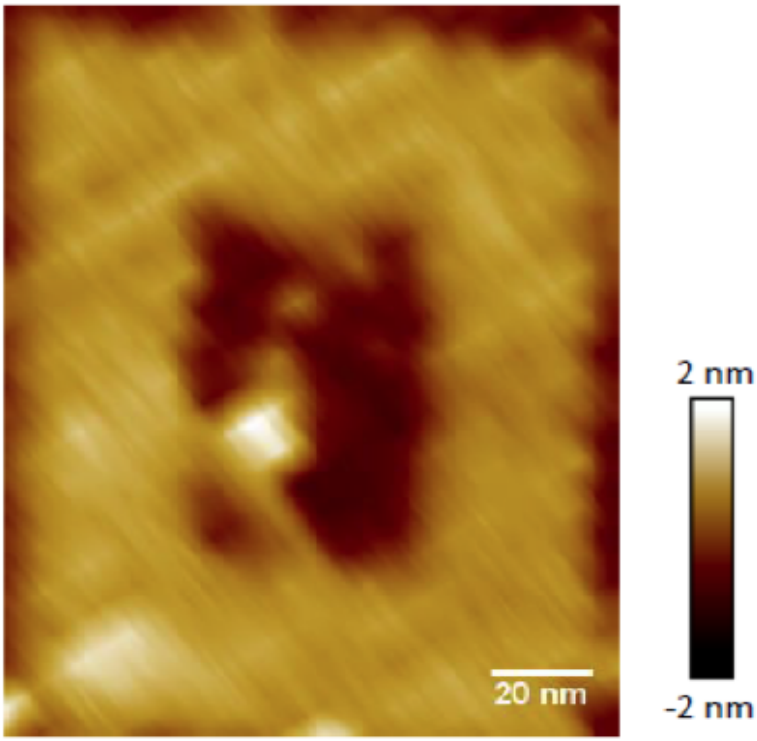


Figure 2.3.1iv. **A sample AFM fluid image (i.e. taken with the sample in buffer).** It shows that a DNA strand is attached to the middle of the DNA origami frame and a bright spot is on the DNA strand. The volume of the bright spot was measured to identify it, and the location of the bright spot on the DNA strand was also measured for further analyses. The scan rate is 27.9 Hz. The resolution is 512 x 64 pixels. The amplitude setpoint is 210 mV. The drive amplitude is 820 mV.

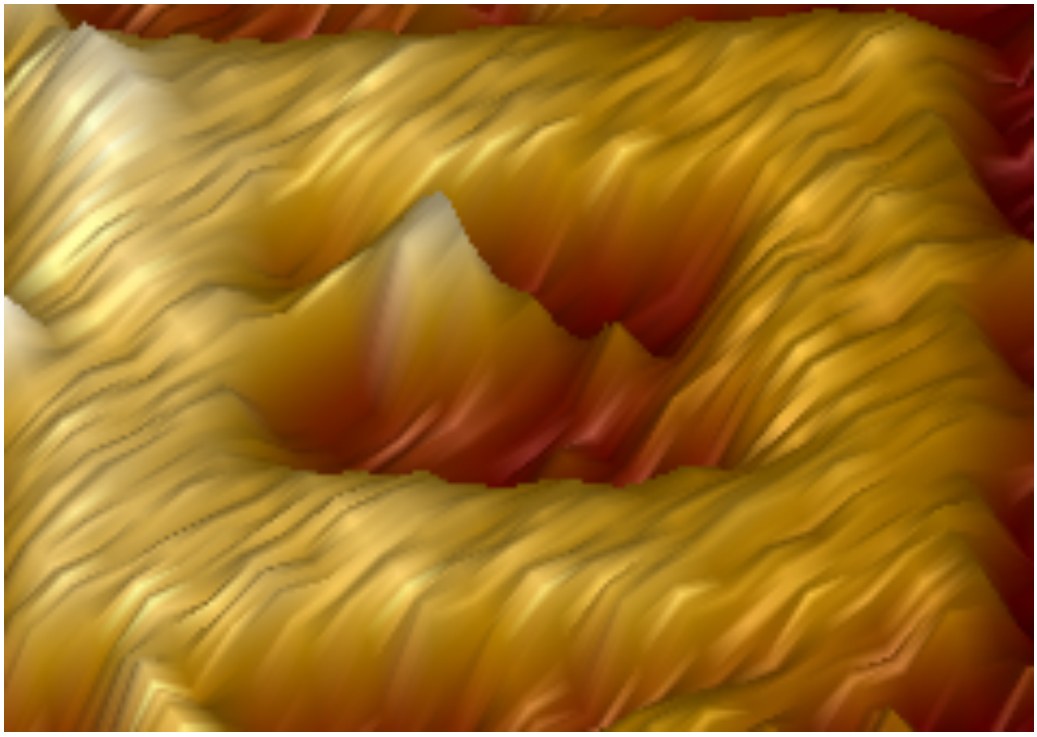


Figure 2.3.1v **The 3D image of Figure 2.3.1iv.** The 3D image of Figure 2.3.1iv (with 90-degree rotation to the right) shows clearly the attachment of the DNA strand to the middle of the DNA frame, and that something of significant size is on the DNA strand. The 3D image is processed using Nanoscope software.

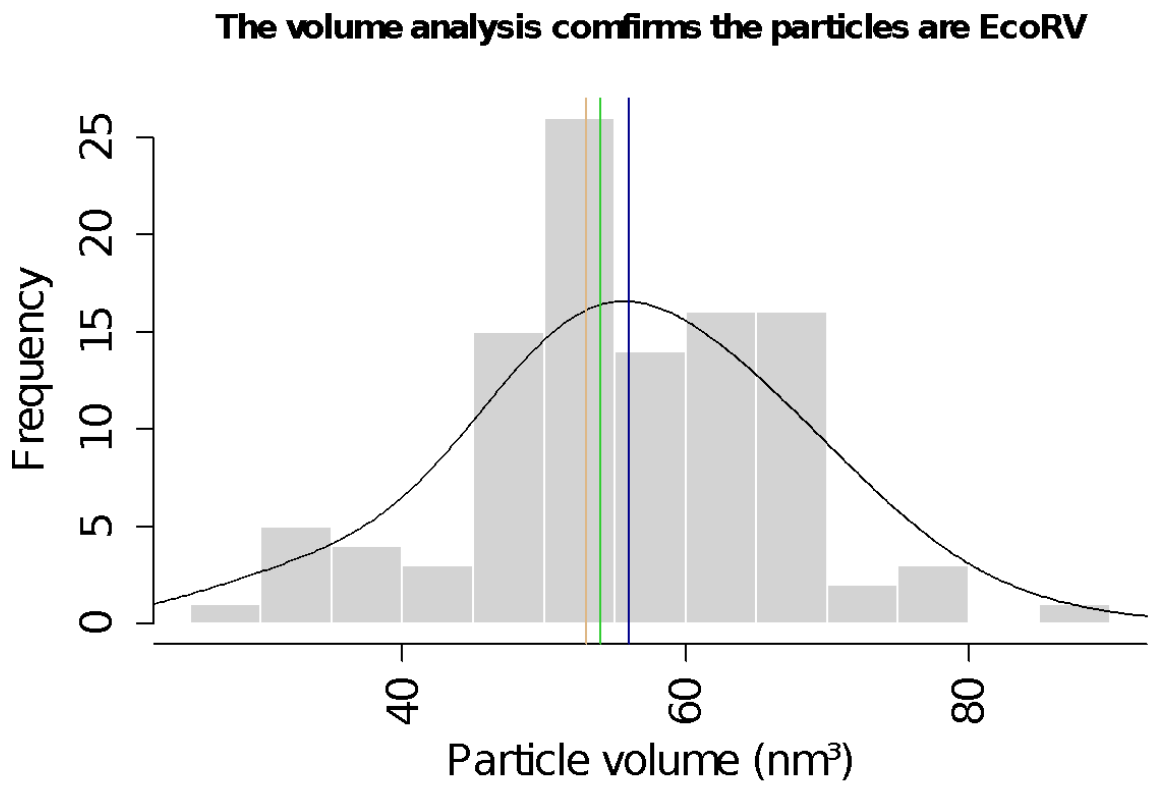


Figure 2.3.1vi A histogram of the 106 volume results of the bright spots on the DNA duplex in the AFM fluid images. The mean is 56 nm³, which is indicated using a dark blue line in the image. It is both close to the expected volume of 54 nm³ (indicted using a green line), which is calculated based on EcoRV's molecular weight, and also close to 53 nm³ (indicated using a brown line), which is the mean of the volume measurements of the bright spots in the dry images. Statistical analyses confirm that the objects on the DNA duplexes in the fluid images are EcoRV. The R script for this figure can be found in the Appendix.

2.3.2 Time series images

The movement of the particle on the DNA duplexes were captured with images taken every two seconds. Further methodological details of the image capturing and processing can be found in Section 2.2.3.2 and 2.2.4. For each image, the distance between the EcoRV and a side of the DNA frame was measured. 987 measurements from 54 different EcoRV molecules were obtained. The data was used for distribution analysis and travel distance analysis. The continuous movement of one and the same EcoRV (1 of the 54 diference EcoRVs) is shown below. More montage figures like this can be found in the Appendix.

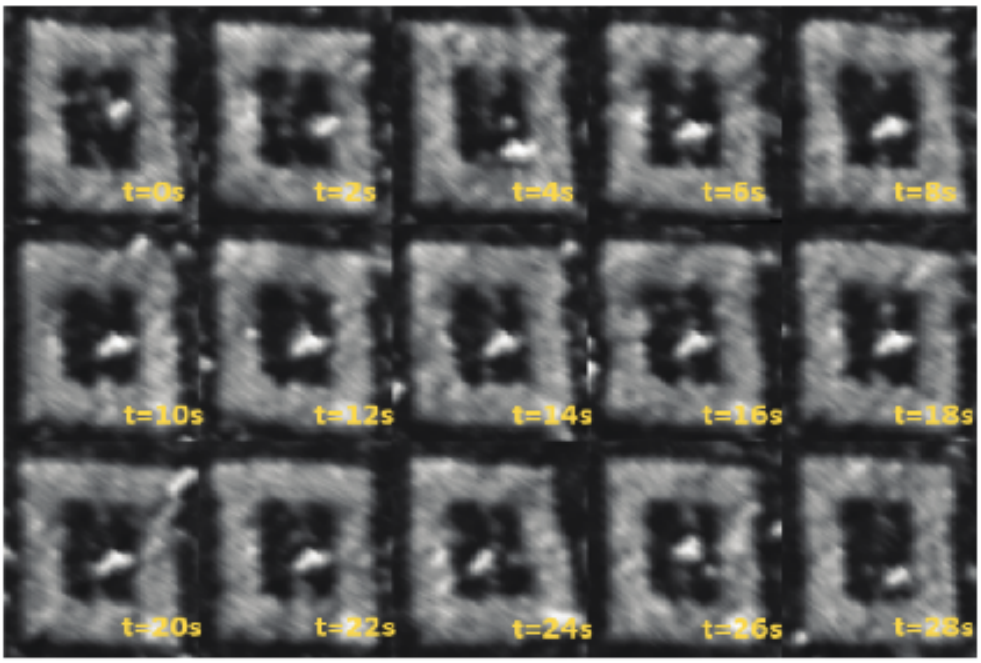


Figure 2.3.2 A set of time series images taken with time interval of 2s shows the motion of one and the same EcoRV on the DNA duplex. The montage figure was made using ImageJ (See Section 2.2.4 for details regarding the image processing).

2.3.3 Distribution analysis

All 987 EcoRV location measurements are presented in this histogram. From it we can see that most of the EcoRV are found at or near its recognition sites. Travel distance analysis reveals more insight as to how this localization occurs. As a control experiment, Type II restriction enzyme SwaI was introduced into the same frame and duplex (which has no recognition sites of SwaI's) system. No similar patterns were discerned.

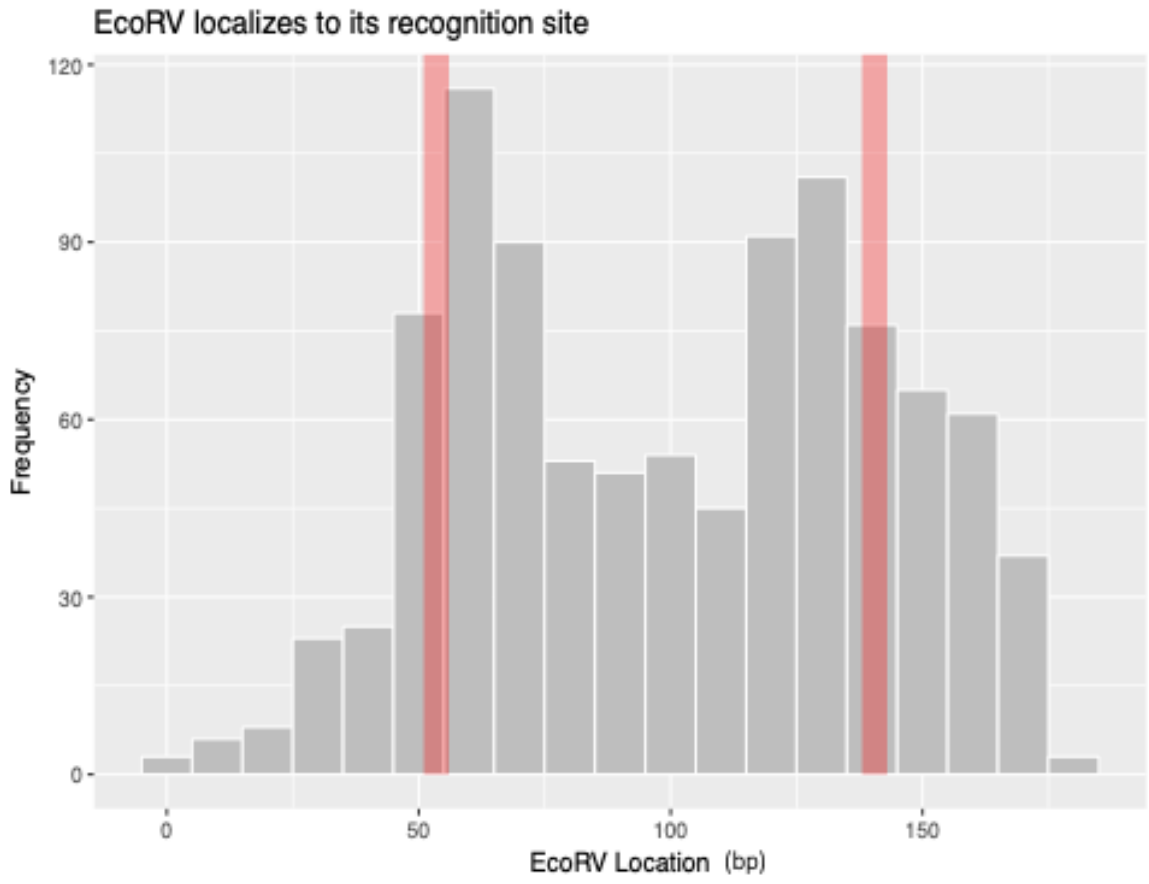


Figure 2.3.3 A histogram of all 987 measurements of the EcoRV locations on the DNA duplex. It shows the EcoRV localizes to its recognition sites at 51-56 base pairs and 138-143 base pairs, which are marked by pink bars in the graph. The R script for this figure can be found in the Appendix.

2.3.4 Travel distance analysis

From each of the 54 sets of the location measurements taken at regular time intervals, we derived data about travel distance per second.

Previous analyses have suggested EcoRV molecules move in a combined style of sliding and hopping on the DNA (Elf *et al*., 2007, Wang *et al.,* 2006, Tafvizi *et al.*, 2011, Hammar *et al.*, 2012, Halford and Marko, 2004, Gowers *et al.,* 2005, Bennet *et al.*, 2008). Our data support this conclusion. Most of the bar-plots of the travel distance of the 54 sets (Figure 2.3.4i) show sudden jumps – suggesting the existence of hopping or jumping. On the other hand, excluding these jumps, the movement appears fairly regular, suggesting that continuous sliding plays a role. Note that the starting point of measuring for each set is different depending mainly on where the EcoRV lands on the duplex.

In a dataset with both sliding and hopping/jumping, we expected a bi-modal or a tri-modal histogram with peaks corresponding to sliding and to hopping/jumping, which, however, was not observe in the histogram of all 932 travel distance values (Figure 2.3.4ii). Instead, there appears to be a continuum of jump sizes, with larger jumps less frequent. Slides and short hops of around 20 bp occur much more frequently than long hops or jumps, and travel distances larger than 40 bp are rarely observed. The largest travel distance value in our experiment is 85 bp, which was recorded once (Note that the experiment was not designed to capture jumps longer than the DNA duplex (i.e. 191bp)). Our results do not agree with the previous prediction that hopping for less than approximately 50 base pairs would not be preferred because it was a waste of energy (Stanford *et al*., 2000).

We next plotted the starting point of travelling (Figure 2.3.4iii), to see how the DNA sequences affect EcoRV's travel distance. Figure 2.3.4iii shows that EcoRV tends to move further when it is far from its recognition site, and moves only by small amounts when it is closer to its recognition site (within about 20 bp).

To test whether such mode switch is because of the attraction of the recognition site to EcoRV, a plot of EcoRV's travel distance with direction was made (Figure 2.3.4iv). This new plot takes into consideration not only the absolute number of each travel distance, but also their direction. The resulting boxes show clearly that most of the hops are toward the nearest recognition site: EcoRV on the left of a recognition site tends to travel from left to right and EcoRV on the right of a recognition site tends to travel in the opposite way.

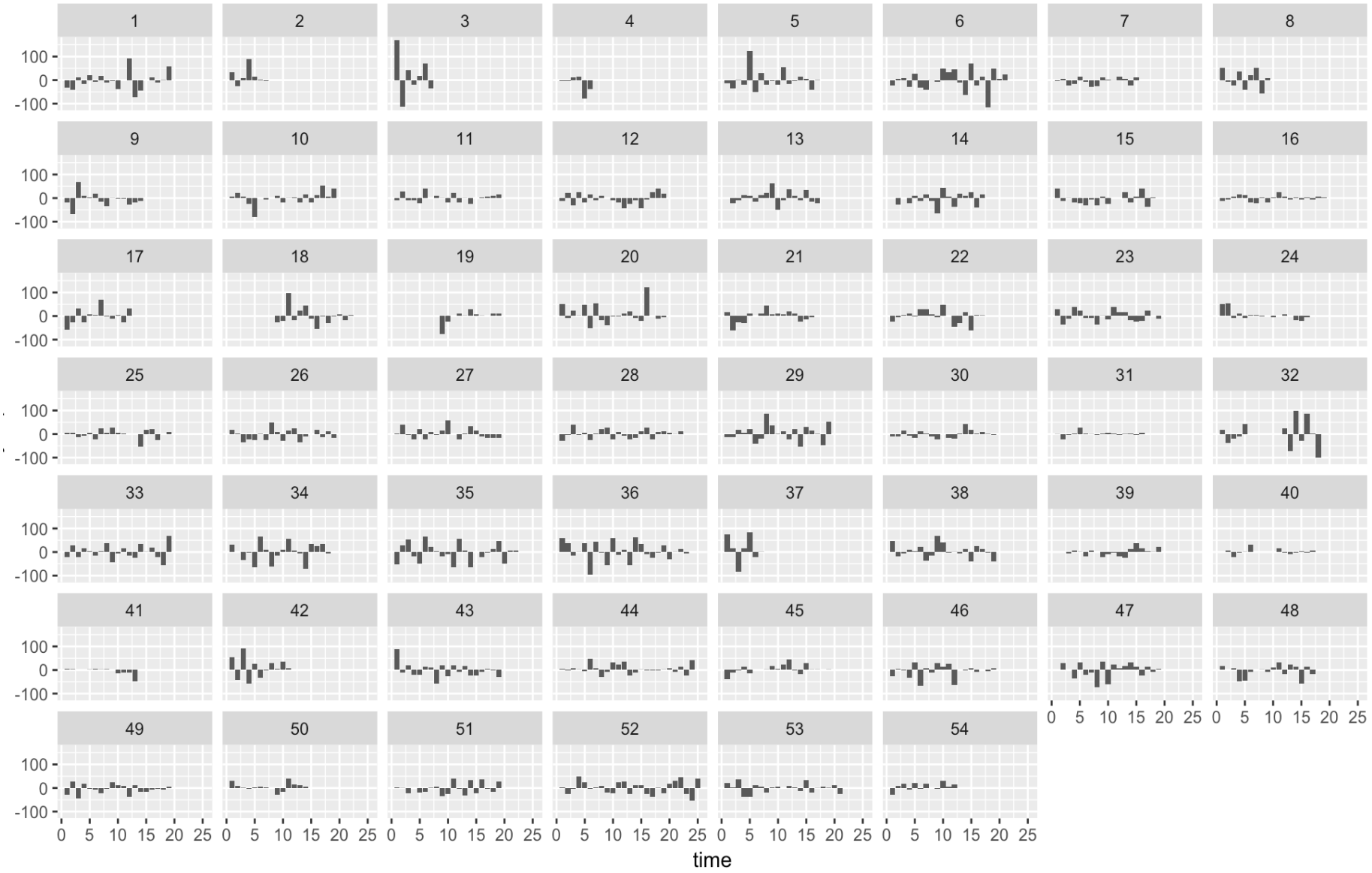


Figure 2.3.4i Bar plots for each of the 54 sets of travel distance values. The spikes in most facets suggest hops are employed by EcoRV for translocating on the DNA duplex. The potisitve and negative values along the y axis suggest the different direction of travelling The R script for this figure can be found in the Appendix.

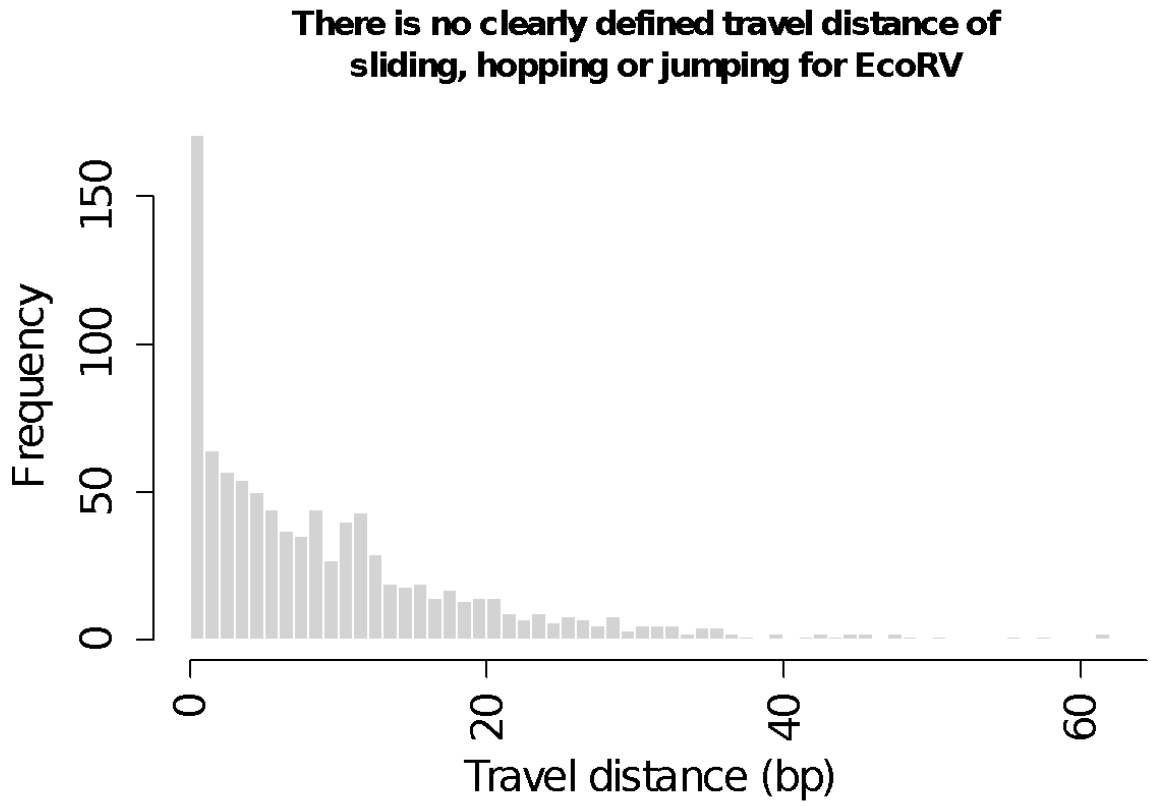


Figure 2.3.4ii Histogram of all 932 travel distance values. It suggests that slides and short hops under 20 bp and slightly above 20 bp are more common than long hops or jumps, and travel distance larger than 40 bp is much rarer. The R script for this figure can be found in the Appendix.

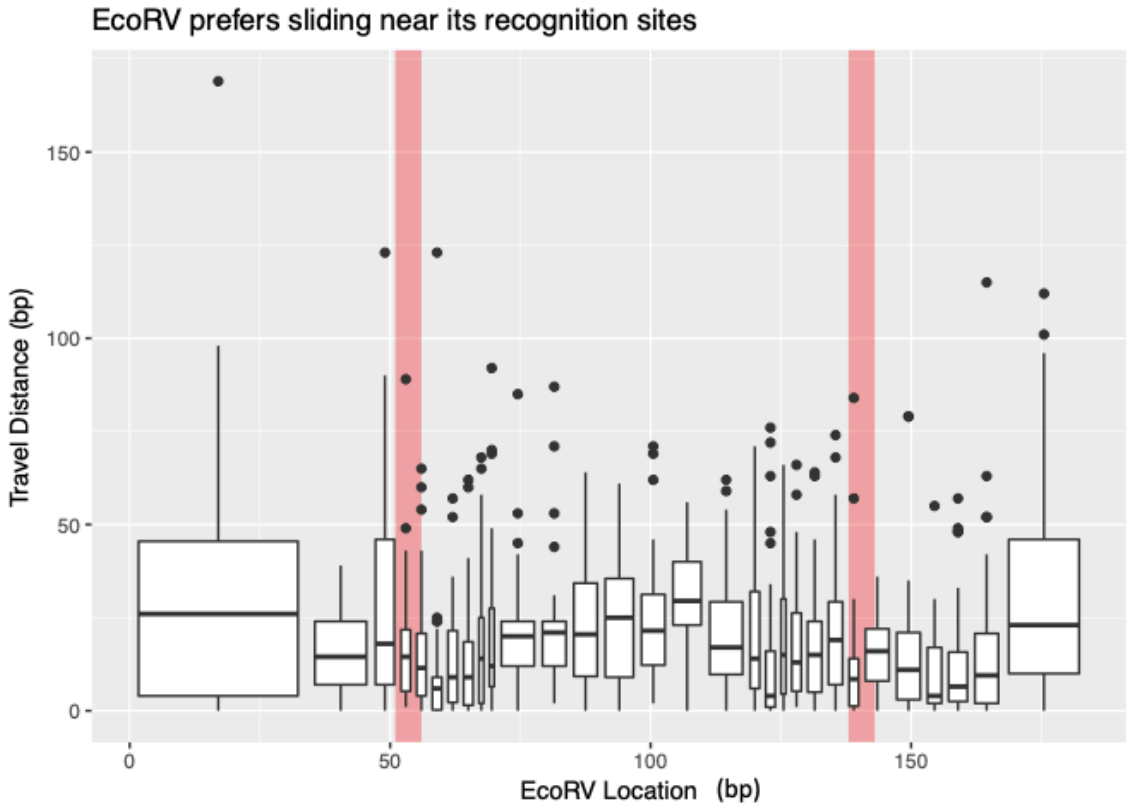


Figure 2.3.4iii **The figure shows a boxplot demonstrating the EcoRV travel distance per second with the different starting points.** This is to see whether being at certain location on the DNA would affect the EcoRV's next immediate translocation distance and translocation mechanisms (sliding or hopping). It can be seen from the boxplot that EcoRV tends to travel short distance using sliding when at or next to the recognition sites. It also seems that farther the EcoRV is from the recognition sites, longer the distance it travels per second in general. Note that long hops do happen at or near the recognition sites, which is shown by the outliers in the figure. The recognition sites are marked by pink bars in the figure. Each box contains the same number of measurements. The R script for this figure can be found in the Appendix.

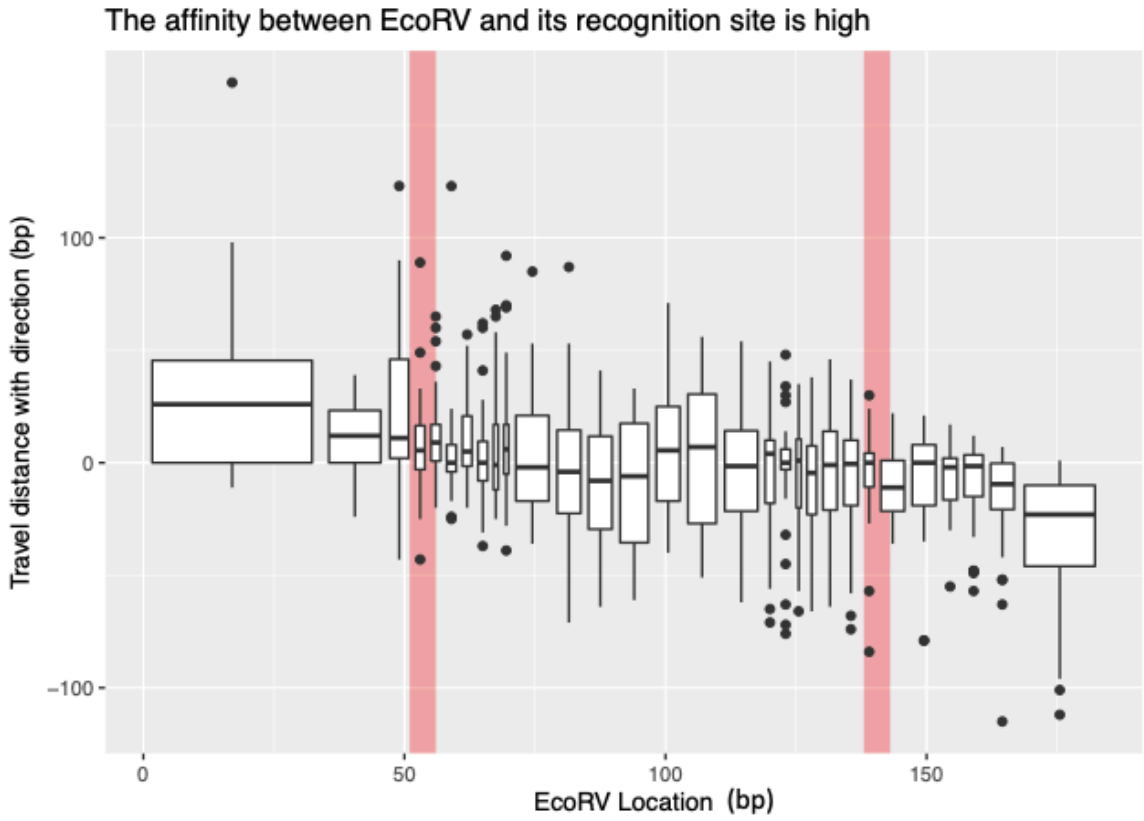


Figure 2.3.4iv **This boxplot shows the directed EcoRV travel distance per second with the different starting points of each travel.** This is to see whether the EcoRV is travelling toward its recognition sites. It can be seen from the longer whiskers of the boxplots that EcoRV tends to travel toward the recognition sites: EcoRV on the left of a recognition site tends to travel from left to right and EcoRV on the right of a recognition site tends to travel in the opposite way. The travel directions of the outliers suggest the same. The recognition sites are marked by pink bars in the figure. Each box contains the same number of measurements. R script for this figure can be found in the Appendix.

2.4 Conclusions

Through the distribution analysis and distance analysis of a large data set of 987 data points from 54 experiment groups of EcoRV's location at each time point during real-time continuous AFM fluid imaging, we have found that:

- EcoRV has high affinity for its recognition sites.
- Hops are clearly employed by EcoRV for translocating on DNA.
- Slides and short hops under 20 bp and slightly above 20 bp are more common than long hops or jumps, and travel distance larger than 40 bp are much rarer.
- EcoRV tends to travel short distance using sliding when at or next to the recognition sites.
- The general trend is that the farther the EcoRV is from the recognition sites, the faster it travels.
- The direction of travelling of an EcoRV at any time point is usually the direction of the nearest recognition site.

This study not only provides details regarding EcoRV's translocation, but introduces new quantitative methods for combining atomic force microscopy and DNA origami technique to study protein translocation on DNA.

For further studies it should be interesting to look for what force(s) hold the enzyme inside the frame and try to determine the diffusion rate of the enzyme using even faster scanning.

Chapter 3 A study of intersegmental transfer - protein complexes bringing together multiple DNA sites - with BcgI as the example

3.1 Introduction

3.1.1 The aim of the present work

The aim of the present work is to study the protein complexes that need two or more recognition sites on one DNA to conduct cleavage. The study intends to address why the proteins need multiple recognition sites, how the proteins form complexes, and how cleavage is conducted. This study used BcgI, a type IIB restriction enzyme, as a model protein. The study focuses on understanding the structure of the complex and how BcgI brings the two recognition sites together.

3.1.2 The synaptic complex in site recognition

Restriction enzymes play crucial roles in protecting bacteria from invading DNA, and precision on sequence recognition and cleavage is essential for the restriction enzymes to fulfil these critical functions. The strategies adopted by the restriction enzymes on sequence recognition and cleavage are honed by billions of years of evolution (Zaremba *et al.*, 2010).

The DNA site recognition and cleavage method used by the typical Type IIP restriction enzymes (e.g. EcoRV, EcoRI, and BamHI) are the best-known (See Section 2.1.3), as they are frequently used in experiments. Type IIP restriction enzymes typically require one recognition site on the DNA and conduct cleavage within the recognition site. The majority of protein-DNA interactions are, however, conducted by multimeric proteins interacting with multiple DNA sites which are distant from each other (Halford *et al.*, 2004). Examples can be found in a wide range of cellular processes: DNA replication, transcription, recombination, transposition and repair (Allen *et al*, 1997, Benkovic *et al*., 2001, Craig *et al.*, 2002, Lloyd *et al*., 2001, Ogata *et al*., 2003).

Studies have found that the majority of restriction enzymes need two or more recognition sites to function properly. Among restriction-modification system it is estimated, based on bacterial genome sequencing, that about 75% of all restriction enzymes need to interact with two DNA sites to conduct DNA cleavage (Kong *et al.*, 2000, Roberts *et al.*, 2003). All the endonucleases from the Type I and Type III restriction-modification systems interact with two sites in a DNA (Halford *et al.*, 2004). Both Type I and Type III restriction enzymes recognise asymmetric sites. While Type I enzymes recognise two sites in direct repeat, Type III enzymes recognise two sites in inverted orientation (Bourniquel and Bickle, 2002, Halford *et al.*, 2004, Peakman *et al*., 2003). Many Type II restriction enzymes also need two copies of their recognition site. Those include most Type IIB, Type IIE, Type IIF, and Type IIS restriction enzymes. Type IIB enzymes (e.g. BcgI) bind two recognition sites and cleave the two sites both upstream and downstream of them (Halford *et al.*, 2004). Type IIE enzymes form dimers and bring together two palindromic sites, each recognised by one of two clefts at the interface of the dimer with the DNA. Although only one of the clefts can cleave the DNA, both clefts must bind to the recognition sites for

the cleavage to happen (Halford *et al.*, 2004). Type IIF enzymes form tetramers, with each of the two dimers of the tetramer interacting with one recognition site; these enzymes cleave both sites concertedly (Halford *et al.*, 2004). Type IIS enzymes (e.g. FokI) often cleave at fixed positions downstream of two asymmetric recognition sites (Halford *et al.*, 2004).

Some proteins have been found to bring more than two DNA sites together. For example, this behaviour has been observed in the EcoRII-DNA triple synaptic complex (Reuter *et al.,* 1998). A three-site DNA-protein complex involving three MuA transposases has been found to act during Mu DNA transposition (Watson and Chaconas, 1996).

The synaptic complexes' ability to bring distant DNA sites together resembles that of other kinds of proteins such as transposases, recombinases and topoisomerases. Indeed, structural and sequential similarities have been discovered between them (Hickman *et al* 2000, Topal and Conrad,1993, Jo and Topal 1995, Vanamee *et al* 2005). There have been many efforts to modify the specificity of the restriction enzymes for more controlled gene editing (Alves *et al*, 1989, Heitman and Model, 1990, Osuna *et al*, 1991, Flores *et al*, 1995, Vanamee *et al*, 2005). Better understanding of how the restriction enzymes form the complex and how the complex bind multiple DNA sites will help understand how the high accuracy of the DNA cleavage is achieved by the complex.

DNA looping

One prominent feature during the process of the protein complexes bringing the multiple recognition sites together is DNA looping.

DNA looping happens quite often during protein-DNA interactions in many fundamental cellular processes (Saiz e Vilar, 2006). For example, in eukaryotic transcription, multiple proteins such as enhancers, silencers, and mediators which bind distal DNA sites are brought together to aid or interfere with the transcriptional complex (Saiz e Vilar, 2006). Through DNA looping, distal DNA regions affect each other or the proteins that bring them together (Saiz e Vilar, 2006). These effects include allowing the transfer of genetic information during recombination, regulating the length of telomeres at the end of chromosomes, and tightening the binding of regulatory proteins to their recognition sites (Saiz e Vilar, 2006).

Studies on thermodynamics have found that DNA loops can form extremely easily in the intracellular environment (Saiz e Vilar, 2006). The *in vivo* free energy required to bring together two sites which are more than one thousand bases apart is only marginally more than that required for two sites less than 100 bases apart (Saiz e Vilar, 2006). It might be easier than we expected for proteins to bind multiple sites on DNA thanks to the frequent occurrence of DNA supercoiling *in vivo*. The Sfil tetramer, for example, binds two DNA sites via two dimers angled at 60 degrees (or 120 degrees) from each other, an arrangement naturally preferred by supercoiling DNA (Emblenton *et al.*, 2004).

The stability of DNA binding is usually strongest for supercoiled DNA, weaker for *cis* DNA, where the two recognition sites are on the same DNA duplex, and still weaker for *trans* DNA, where the two recognition sites are on two DNA duplexes; this is based on the probability of juxtaposition of the two DNA sites in these scenarios (Halford *et al.*, 2004). The distance between the two recognition sites is also an important determining factor for juxtaposition. For example, for two sites *in cis* that are more than 500 base pairs apart, the

probability of juxtaposition is much lower than when they are closer (Halford *et al.*, 2004). The stability of binding two DNA sites is also affected by the helical structure of DNA because for both sites the requisite face of the DNA, which is usually the major groove, should be presented to the DNA-binding surfaces of the protein complex (Halford *et al.*, 2004). For example, the maximal stability of looping occurs at a staggered spacing of *n* plus or minus ½ turns (Watson *et al.*, 2000).

The average target search time drops as the DNA topology is changed from an open chain to a closed loop to a supercoil with increasing number of juxtaposition sites. In one approach, restriction enzymes were found more readily to locate to supercoiled plasmid than to the same plasmid at its relaxed state (Gowers and Halford, 2003). And according to a recent quantitative study, a DNA supercoil with the same number of recognition sites for a protein as a DNA open chain was found to interact with the protein approximately 23% faster than the DNA open chain (Mondal and Bhattacherjee, 2017). The variations in the number, position and relative distribution of the juxtaposition sites greatly impact the search dynamics (Mondal and Bhattacherjee, 2017). In general, the closed-loop and supercoiled DNA seem to promote protein-DNA interaction in two ways. First, they promote juxtaposition of distal DNA sites which facilitate protein complex which need multiple recognition sites for conducting cleavage. Second, they encourage three-dimensional translocation by increasing the chance of the protein re-associating with the DNA after dissociating it (Mondal and Bhattacherjee, 2017).

3.1.3 The complex forming mechanisms

There are two main pathways for proteins to bring two DNA sites together (Halford *et al.*, 2004). In one pathway, proteins first bind both sites of the DNA recognition sites and then collide to form a complex. Alternatively, proteins may bind one DNA recognition site and form a complex which holds the first recognition site as well as letting the adjacent DNA chain pass through (usually with energy from ATP hydrolysis). One feature to easily tell one pathway from the other is whether the loop length increases during the process. In the first pathway, when the proteins on the two DNA sites associate to form a complex, the DNA forms a loop of a fixed size. In the second pathway, the loop size increases as the DNA slides through the protein complex.

When the complex is a tetramer there are two possibilities: a dimer is formed at each of two DNA recognition sites and then collide resulting in a DNA loop of fixed size; or, a tetramer is formed at one recognition site which makes the DNA slide through, during which process the DNA loop increases in size, until the tetramer holds two recognition sites.

Restriction enzymes use both pathways. Previous studies have shown that the two Ecl18kI dimers bind to two target sites before associating into a transient tetramer and bringing the two sites together (Zaremba *et al*, 2010). SgrAI dimers also form transient tetramers for DNA cleavage (Zaremba *et al.,* 2010). On the other hand, some restriction enzymes form stable tetramers on one recognition site before searching for the other site. Restriction enzymes complexes that conduct loop capture rather than protein-protein association include SfiI, Cfr10I, Bse634I, NgoMIV, and Cfr42I (Zaremba *et al.*, 2010).

For the second pathway which involves DNA sliding through a protein complex that keeps holding one DNA recognition site, extra energy input (e.g. from ATP hydrolysis) is often required because the helical structure of DNA discourages such sliding (Halford *et al.*, 2004). If a subunit of the protein complex follows the helical path as it slides along the nonspecific DNA while the rest of the complex retains the hold of the initial specific site, the slide-through DNA will soon become severely twisted (Jeltsch *et al*., 1994). With the help of ATP hydrolysis, loop formation by one-dimensional translocation (i.e. sliding) is not only possible but common. Many proteins interacting with DNA have an ATPase function, including DNA recombinases, chromatin remodelling enzymes, and most of the Type I and Type III restriction enzymes (Halford *et al.*, 2004). For example, Type I restriction enzymes usually have ATPase and translocase as subunits (Halford *et al.*, 2004). Studies have shown that most Type I enzymes stay bound to their recognition sites while sliding adjacent nonspecific DNA past the stationary protein complex (Bourniquel and Bickle, 2002). The reason for the second pathway to be so commonly used by Type I enzymes might lie in their modular adaptability to new specificity. Type I R-M enzymes are naturally modulated to change specificities much more than other restriction enzymes. Using the second pathway, when the specificity of the enzyme complex is changed, the HsdR subunits that translocate and cleave the nonspecific DNA can remain unchanged (Halford *et al.*, 2004, Dybvig *et al.*, 1998).

The protein-protein association pathway is largely facilitated by Brownian movement of the DNA (Halford *et al.*, 2004). Brownian dynamic simulations on supercoiled DNA have shown that the juxtaposition of the two sites is likely to occur within a few milliseconds (Huang *et al.*, 2001, Jian *et al.*, 1998, Klenin and Langowski, 2001). Although the juxtaposition of two recognition sites by Brownian movement is reached much more

slowly in relaxed DNA, bringing two recognition sites together does not seem to be the rate-limiting step for restriction enzyme complexes (Halford *et al.*, 2004).

3.1.4 The need of two or more recognition sites

The complexes which conduct concerted cleavage on two or multiple recognition sites usually have the best efficiency when doing so. In comparison, their efficiency at a solitary recognition site is usually very low. For example, in a study where exogenic oligonucleotides were added to the reaction mixture of EcoRII and DNA oligonucleotides, while the one- and two-site cleavage were speeded up, the three-site cleavage, which had already been the fastest, was hardly affected (Tamulaitis *et al*, 2006). This suggested that the three-site cleavage was maximally efficient, and the efficiency of the one- and two-site cleavage had been suppressed (Tamulaitis *et al*, 2006). The difference on the cleavage efficiency has led to hypotheses on why the protein complexes are formed. One hypothesis, which has been supported by many studies, is that the synaptic complexes are formed to bring two or multiple recognition sites together before cleavage to increase the accuracy of the cleavage (Embleton *et al.*, 1999).

Some studies support this hypothesis. For example, Bse643I naturally forms tetramers, and the tetramers cut two recognitions sites with better efficiency than one recognition site. In the study, a site-directed mutation was introduced to convert the tetramer into dimers and the resulting dimers were found to cut a single recognition site as fast as a tetramer cut two sites (Zaremba *et al*, 2005). The authors suggest that down-regulation of cutting at a single-site for tetramers was due to crosstalk between the two dimers forming the tetramer (Zaremba *et al*, 2005). In other words, the tetramer construction prevents the component dimers from cutting at solitary recognition sites. A few other studies have also suggested that tetramers stabilize the dimers and increase the specificity of target site recognition (Zaremba *et al*, 2010).

The reasons of multiple site-binding are sometimes obvious, but not always. For example, it is clear why enzymes responsible for genetic information exchange between two DNAs need to cleave two DNA sites at once. If recombinases and transposases are active after a single binding, they might cause genetic information loss, with potentially lethal consequences. Other cases, such as for restriction enzymes, are less obvious. Some researchers have suggested that Type II restriction enzymes that need two DNA recognition sites to act effectively are related through evolution to the recombinases (Benkovic *et al.*, 2001, Taylor *et al*., 1990, Wentzell *et al.*, 1995). But so far insufficient evidence has been found to support this hypothesis (Halford *et al.*, 2004), and there are concrete differences between recombinases and restriction enzymes. When bound to a recombinase, two DNA strands tend to lie next to each other to be ready for the recombination, whereas when bound to a restriction enzyme complex, two DNA strands tend to be on the opposite side of the complex - perhaps to prevent strand transfer (Halford *et al.*, 2004). In addition, these two types of enzymes use distinct chemical mechanisms for DNA cleavage (Mizuuchi and Adzuma, 1991, Mizuuchi *et al.*, 1999). Although there is much speculation, we do not know why many restriction enzymes need to cut two or more recognition sites at once. It might be simply to improve the precision of cleavage.

For restriction enzymes, the fundamental roles are to destroy invading genomes and to protect the bacteria's own genome, so the bacteria must make sure the enzyme does not cleave any modified DNA sequence of its own (Halford *et al.*, 2004). In nature, DNA restriction is not normally followed by recombination, and so spurious cleavage frequently has lethal consequences. The precision of the 'double-check system' can distinguish small differences between recognition sites. *In vivo*, every copy of the recognition sequence in the host chromosome is protected by methyltransferase, making each host site slightly

different from the recognition site on the invading virus. Such ‘double-check system’ is useful in ensuring that sequences differing by only one or two base pairs from the recognition sequences are not cleaved by accident (Halford *et al.*, 2004, Bickle and Kruger, 1993). This ‘double-check system’ also lowers the risk of the restriction enzymes cleaving the host chromosome when the recognition sequences are incompletely methylated (Halford *et al.*, 2004).

Some restriction enzymes cut both recognition sites, some use one of the sites as an activator. For example, SfiI, a Type IIF enzyme, cleaves the two DNA sites at the same time whereas FokI, a Type IIS restriction enzyme, and most Type IIE enzymes cut only one DNA site, using the second site as an allosteric activator (Bilcock and Halford, 1999, Vanamee *et al.*, 2005, Vanamee *et al.,* 2007).

3.1.5 BcgI

BcgI is a non-standard Type II (Type IIB) restriction enzyme. Type IIB systems are particularly interesting because they have some features of both Type I and other Type II systems. Like Type I enzymes, Type IIB enzymes recognise bipartite sequences, but they cut at specified points near the sites, which makes them Type II enzymes (Marshall *et al*., 2007). BcgI is the archetype of Type IIB enzymes. The recognition sites of BcgI are 10/12 CGA N6 TGC 12/10, where N6 can be any six nucleotides between the fixed sequences CGA and TGC. BcgI cuts 10/12 nucleotides upstream from CGA and 12/10 nucleotides downstream from the TGC site. BcgI is known to cut two recognition sites *in cis* in one DNA molecule, or *in trans* in two DNA molecules. The enzyme cuts at eight phosphodiester bonds, releasing two 34bp fragments.

BcgI has a molecular weight of 182.4 kDa. It is made of two subunits: BcgIA (MW=71.6 kDa) and BcgIB (MW=39.2 kDa). The ratio of BcgIA to BcgIB is 2:1 (Kong *et al*., 1994). The A subunit has both DNA cleavage and DNA methylation activities, but only functions in the presence of the B subunit (Kong, 1998). The B subunit is thought to function in site recognition (Kong, 1998).

Previous studies have suggested that the two sites are cut by BcgI within the lifetime of a single DNA-enzyme complex. The two cleavages happen simultaneously, or near-simultaneously (Marshall *et al*., 2007). In simultaneous cleavage, eight phosphodiester bonds are cut at the same time, so the synaptic complex needs eight active sites. The formation of a synaptic complex that can potentially cut eight phosphodiester bonds at once is therefore of considerable interest.

Forming tetramers is one possible model. In 2013, Smith's group proposed a model of a BcgI tetramer (Smith *et al.,* 2013). Alternatively, it is possible that the complex is formed by BcgI dimers, and it uses its active sites more than once in the lifetime of the complex. In this case, the BcgI active sites might need to switch strands via a conformational change of the complex.

3.2 Materials and methods

3.2.1 The dry imaging experiment procedure

The DNA plasmid used for this experiment is circular pBR322 of 4361bp (0.5 µg/µl from Thermo Scientific). It has three BcgI recognition sites, at 708, 2063 and 3882 respectively (Figure 3.2.1). To make the plasmid linear, 1 µl of the plasmid was mixed with 1 µl of BamHI (20 units/µl, from New England Biolabs) and 18 µl NEBuffer3 (from New England Biolabs). The mixture was then placed in a thermal bath at 37^{o}C for 1 hour. The product which was the linear pBR322 was verified by electrophoresis in a 2% agarose gel at 90 V with circular pBR322 sample as the control. The verified linear plasmid was mixed with 1 µl of PstI restriction enzyme stock (20 units/µl, from New England Biolabs). This new mixture was maintained in the thermal bath at 37^{o}C for an hour followed by electrophoresis in a 2% agarose gel at 90 V. The plasmid fragment of 3232 bp with 2 BcgI recognition sites was extracted using a Qiagen gel extraction kit. This final product was stored at 4^{o}C. The gel purification results can be found in the Appendix.

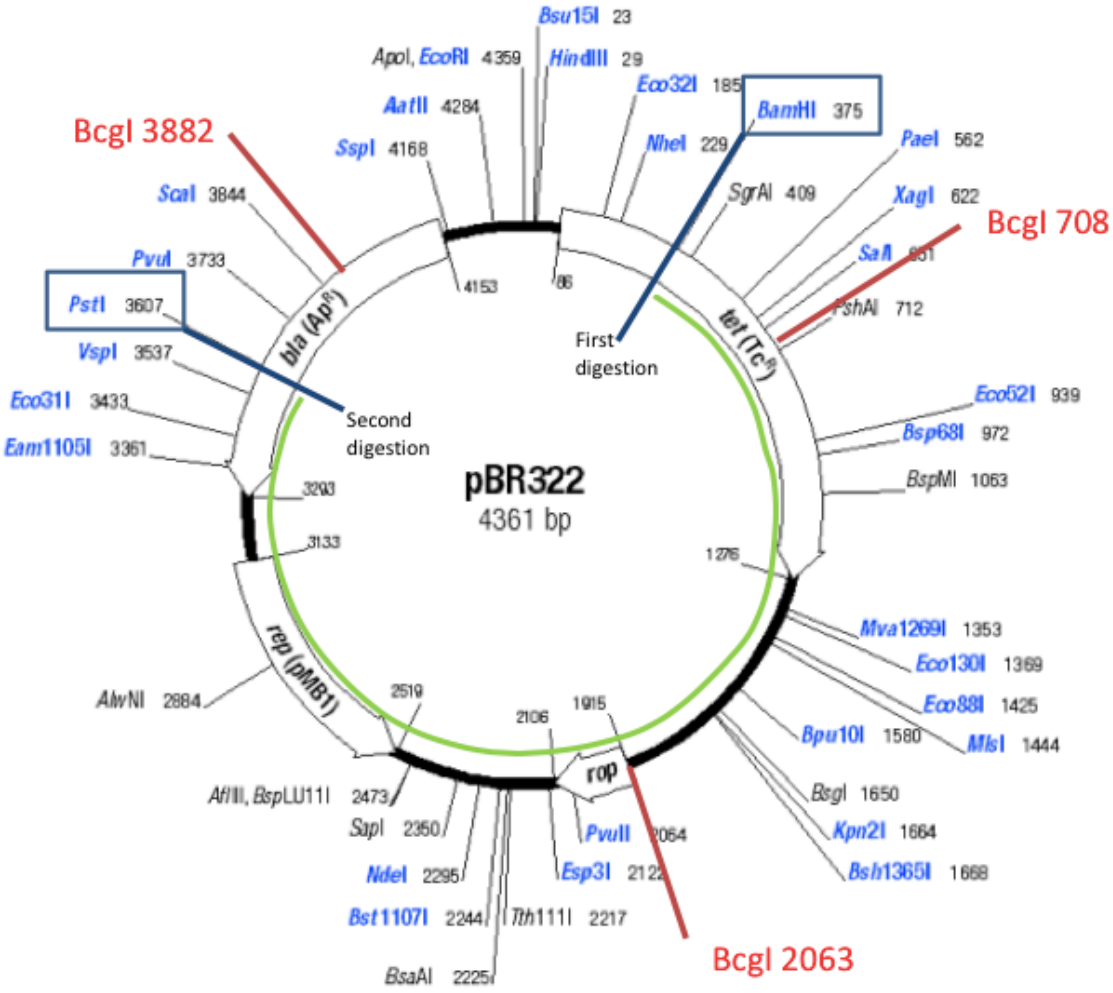


Figure 3.2.1 A figure showing plasmid pBR322, with marks at BcgI recognition sites and BamHI and PstI recognition sites. The green line indicates the part of the plasmid used for AFM dry imaging with BcgI after BamHI and PstI digesttion and gel purificaion. The seleted part has two BcgI recognnition sites – one is 333 bp from the neareat end, and the other is 1688 bp from the same end. The figure is adapted from ThermoFisher website.

The samples for the dry imaging were made as follows: 100 µl of the plasmid solution of 3232 bp with 2 BcgI recognition sites, 20 µl S-adenosylmethionine (1 mM) and 1 µl BcgI (from New England Biolabs) were mixed with BcgI reaction buffer (1x NEBuffer 3.1 (100 mM NaCl, 50 mM Tris-HCl, 10 mM $MgCl_2$ and 100 µg/ml BSA) plus 20 µM S-adenosylmethionine from New England Biolabs) to form a 1 ml mixture. 40 µl of this mixture was incubated at room temperature for 1 minute before being pipetted onto a mica disc. The loaded mica was left at room temperature for 2 minutes before the surface was washed by BPC water. This mica was then dried with nitrogen gas and then stored in a silica gel container for further drying.

Samples were taken out of the silica gel container to be observed under AFM in dry imaging mode. 1 μm^2 images were captured for data analyses. The data was processed using Scanning Probe Image Processor (SPIP), a software developed by Image Metrology for processing and analysing microscopic images, Microsoft Excel and R, which is a programming language used widely for statistical computing and graphics and originally developed by Ross Ihaka and Robert Gentleman among others.

3.2.2 The fluid imaging experiment procedure

For the fluid imaging with two BcgI recognition sites *in trans*, a DNA frame (See Section 3.2.3) was designed used CadNano to have two DNA duplexes (See Section 3.2.4) attach to it. To form the DNA origami frame, 2 µl of the M13mp 18ss DNA (10 nM), 2 µl of 10x origami folding buffer (20 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM EDTA, pH7.6, 1x), 5 µl of the DNA oligomers (200 nM) and 11 µl of BPC grade water (from Sigma-Aldrich) were mixed and placed in a thermal cycler from 80 °C to 25°C, descending 1°C every minute. The formed frames were stored at 4°C.

To form each of the two duplexes, 2 µl of the forward strand (100 µM), 2 µl of the reverse strand (100 µM) and 16 µl of 1x origami folding buffer were mixed and placed in a thermal bath starting at 90°C for 5 minutes and then descending 1°C every minute to 25°C. The product was then put to undergo electrophoresis in a 2% agarose gel and Tris-acetate-EDTA (TAE) buffer at 110V. The gel was then UV-dyed by gentle shaking, first for 30 mins in ethidium bromide solution of 100 ml deionised water and 6µl ethidium bromide, and then another 30 mins in 100ml deionised water. The gel was then illuminated under UV light and the target band of about 100 bp was cut off from the gel pad. The target DNA duplexes of about 100 bp were then extracted from the band using QIAquick gel extraction kit from Qiagen.

The obtained duplexes (about 20 µl of each of the two types) were then mixed with the formed frames (20 µl) and placed in a thermal cycler of 40°C to 15°C, descending 0.5°C every minute. The resulting products were put to undergo electrophoresis in a 0.5%

agarose gel and 1x TAE buffer with 10 mM $MgCl_2$. The dyeing and illumination steps were the same as described above. The band of the size of the target DNA products (duplexes attached to the frame) was then cut off from the gel pad and further purified using a Freeze and Squeeze purification kit from Bio-Rad with a spinning rate of 6,000 rpm. Sephacryl S-400 purification from GE Healthcare with elution limit of 271bp was also used to purify the mixture.

For AFM fluid imaging, each time, 40 µl of the complex solution was pipetted onto a mica disc and subsequently put under the AFM. Depending on the number of the frame and duplex complexes in one AFM image, the frame and duplex solution could be concentrated using Amicon Ultra centrifugal filters from Merck Millipore before being pipetted onto the mica disc.

During the imaging process, 1 ml of 5 nM BcgI solution (diluted using 1x BcgI reaction buffer) was introduced into the system through a 1 ml syringe attached to the scanner. The AFM images were taken at a rate of two seconds per image. The images were taken by Bruker Dimension FastScan AFM in the tapping in fluid mode. The model of all the probes used for this experiment was Fastscan-DX from Bruker. The AFM images were analysed using the Scanning Probe Image Processor (SPIP) software (Version 6.11) (Image Metrology 2014), and ImageJ.

3.2.3 The design of the DNA frame

The DNA origami frame is built using M13mp 18ss DNA (250µg/ml) from New England Biolabs and short DNA oligomers from Integrated DNA Technologies. The frame surrounds an empty space of 96 bp long and approximately 30 bp wide in the middle. The frame has two pairs of overhangs of 16 bp. Two overhangs of each of the pairs intrude into the empty space from opposite directions, resulting at a 64 bp distance between them. The two pairs are in parallel to each other.

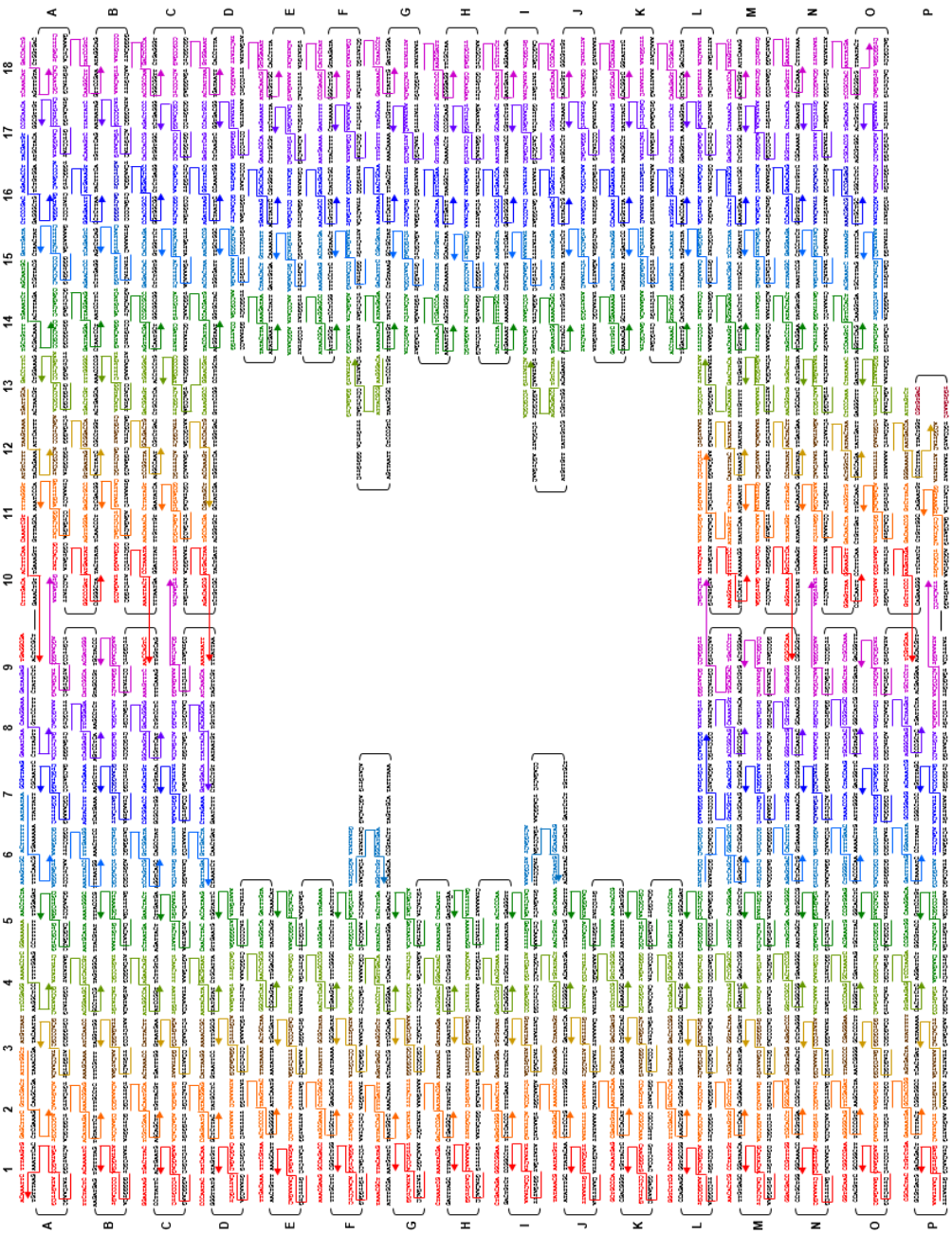


Figure 3.2.3 DNA origami frame for real-time fluid AFM imaging of BcgI forming complex. The DNA origami frame designed for this project is made by M13mp 18ss DNA (black) and short DNA oligomers (coloured). The frame surrounds an empty space of 96 bp long and approximately 30 bp wide in the middle. The frame has two pairs of overhangs of 16 bp. Two overhangs of each of the pairs intrude into the empty space from opposite directions, resulting at a 64 bp distance between them. The two pairs are parallel to each other. In the experiment, two DNA duplexes, each with one BcgI recognition site, are attached to the overhangs and are therefore also parallel to each other.

3.2.4 The design of the duplexes

Two DNA duplexes are designed. Both of them are of a length of 74 bp, which is slightly longer than the 64 bp distance between a pair of overhangs. Each of the duplexes has at both ends a single strand of 16 bp that are complementary to a pair of overhangs. Each of the two duplexes has one BcgI recognition site at about the same place of the duplex. The length of the duplexes gives them enough flexibility for the BcgI to bring together its two recognition sites.

The DNA duplex that is designed to attach to the DNA origami frame

Forward strand: 16 base sticky end +74 base sequence

Reverse strand: 74 base complementary sequence + 16 base sticky end

The underlined sequences are the BcgI recognition site CGA N6 TGC.

The upper duplex:

Forward strand 5'-3':

CTGTAGCT CAACATGT CGTTACCA GCCGATGT TAGCGATC TTACGCCG

ATCCAGGT GCGAGAGA AACGGATC CGTGCCCA TTTCAGTT AC (90)

Reverse strand 5'-3':

GACGGGAG AATTAACT GTAACTGA AATGGGCA CGGATCCG TTTCTCTC

GCACCTGG ATCGGCGT AAGATCGC TAACATCG GCTGGTAA CG (90)

The lower duplex:

Forward strand 5’-3’:

TTGCCTGA GAGTCTGG CGTTACCA GCCGATGT TAGCGATC TTACGCCG

ATCCAGGT GCGAGAGA AACGGATC CGTGCCCA TTTCAGTT AC (90)

Reverse strand 5’-3’:

CGACAATA AACAACAT GTAACTGA AATGGGCA CGGATCCG TTTCTCTC

GCACCTGG ATCGGCGT AAGATCGC TAACATCG GCTGGTAA CG (90)

3.3. Results and discussions

3.3.1 The dry images

The pBR322 plasmids in the suspension were fully digested by BamHI and PstI. The products were 1129 bp plasmid fragments with a single BcgI recognition site and 3232 bp plasmid fragments with two BcgI recognition sites (Figure 3.3.1i). After gel extraction, the 3232 bp fragments were isolated and used for interacting with BcgI for the dry images.

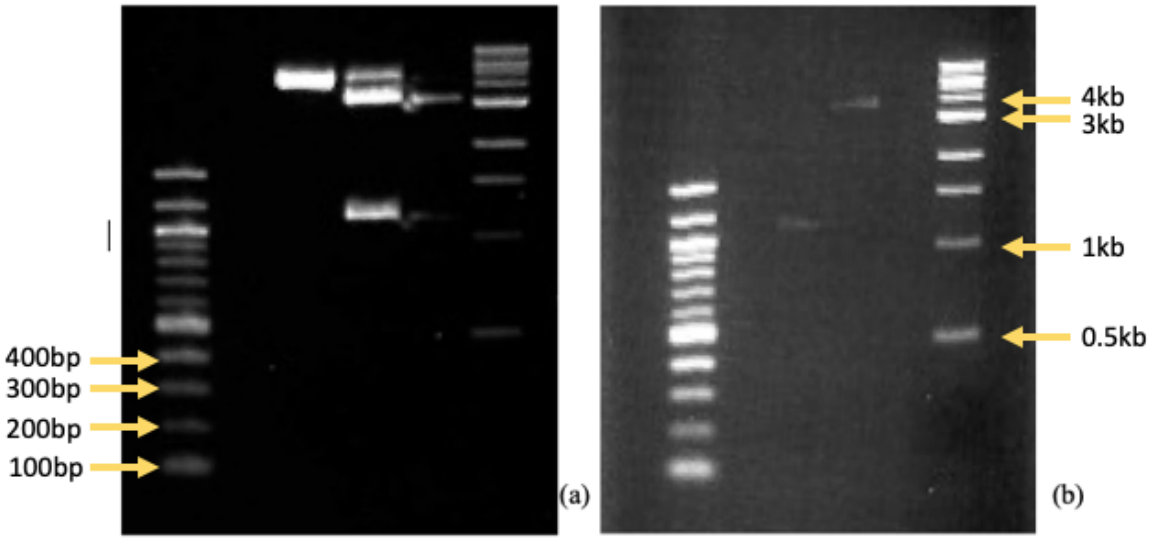


Figure 3.3.1i The purification results of the 3232bp fragment from the pBR322 plasmid. From (a), pBR322 plasmid (shown on the left lane), which is 4361bp long, is completely digested into a 3232bp fragment and a 1129 bp fragment (shown on the right lane) by BamHI and PstI. (b) shows that the 3232bp and 1129 bp have each been purified from the mixture. The 3232bp with two BcgI recognition sites are used in the following experiments to interact with BcgI for AFM dry images.

Hundreds of dry images of 15 sets of different 3232 bp plasmid fragments and BcgI's interactions were collected, showing different stages of the interaction. A panel of the images are shown in Figure 3.3.1ii (See the appendix for more images like this).

From looking at the images, we realised that the BcgI at different stages of interaction with the DNA were of different sizes – a sign of complex formation. We then conducted volume analysis to reveal more information about the components of BcgI complexes.

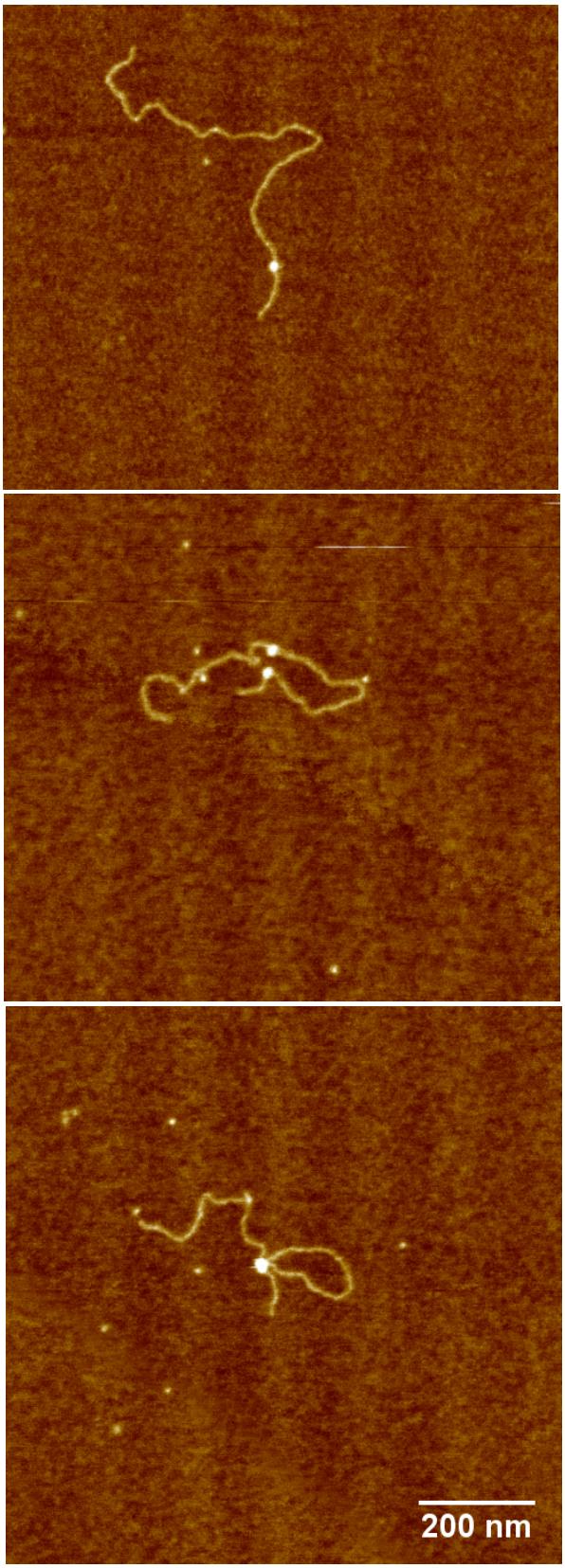


Figure 3.3.1ii **A panel of dry images of the interaction between the 3232 bp plasmid fragment and BcgI at different stages.** The scale bar applies to all three images. The scan rate is 18.6 Hz. The resolution is 512 x 512 pixels. The amplitude setpoint is 250 mV. The drive amplitude is 20 mV.

3.3.2 Volume analysis

The volume measurements of each BcgI (by itself or in complexes) on the DNA strand in the dry images was calculated as a segment of a sphere using the following equation:

$$Vm = (\frac{h}{6})(3r^2 + h^2)$$

Where r is the basal radius and h is the height.

We obtained 302 volume measurements, which followed the tri-modal distribution shown in Figure 3.3.2. The mean values of the three groups are 320 nm^3, 680 nm^3 and 1200 nm^3. They were compared with the volume of BcgI by itself, and in 2- and 4-molecule complexes calculated based on BcgI's molecular weight using the following equation:

$$Vc = \left(\frac{Mo}{No}\right)(V1 + dV2)$$

Where M_0 is the molecular weight, N_0 is Avogadro's number. V1 and V2 are the partial specific volume of particle (0.74 cm^3/g) and water (1 cm^3/g) respectively. Finally, d is the level of duration (normally using 0.4 g water/g protein) (Schneider et al., 1998).

The calculated values are 348 nm^3, 696 nm^3, and 1392 nm^3, respectively, in reasonably close accordance with the observed values (Figure 3.3.2).

Thus, analysis of the dry images suggests that BcgI forms 2-molecule and 4-molecule complexes to bind two recognition sites.

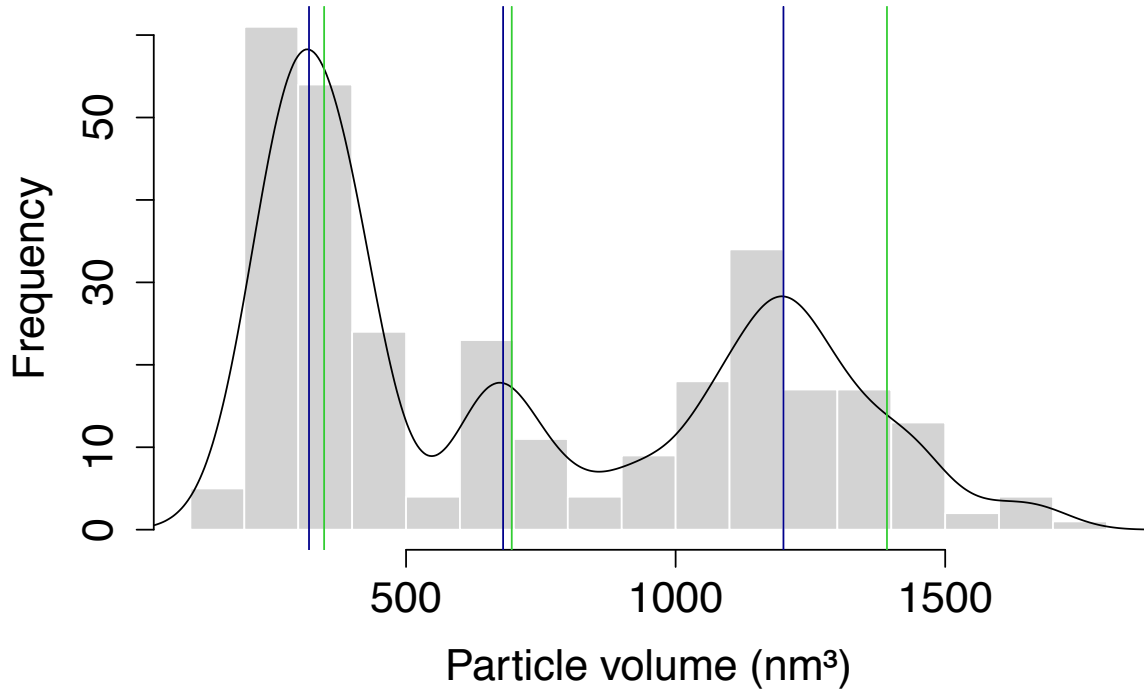
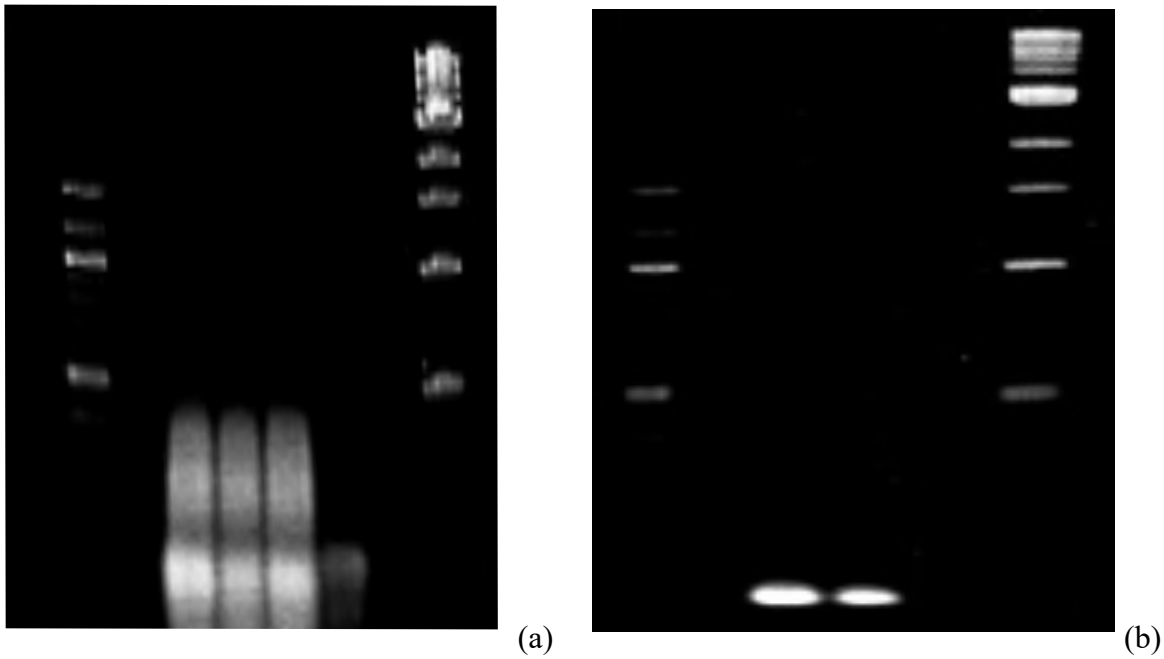


Figure 3.3.2 The histogram of the volume values of BcgI (complexes) in the AFM dry images shows that the values fall into three groups. The mean values of the three groups, which are indicated by dark blue lines in the figure are 320 nm^3, 680 nm^3 and 1200 nm^3. They are close to the expected volume values of one BcgI (348 nm^3), two-BcgI complexes (696 nm^3) and four-BcgI complexes (1392 nm^3) respectively. The expected volume values are calculated based on BcgI's molecular weight and are indicated by green lines in the figure.

3.3.3 The fluid images

A panel of fluid images (Figure 3.3.3ii, see the appendix for more images like this) shows the process of BcgI complex forming in real time. The two DNA duplexes, each carrying a BcgI recognition site (Figure 3.3.3i), were initially unoccupied. Then both of the DNA duplexes were occupied and eventually brought together (The scanning rate of every 2s used in the experiment was not fast enough to catch the process of the collision). It has been suggested previously that BcgI could bring two recognition sites together, whether the two sites are *in cis* or *in trans*. This is supported by our results from the dry images and fluid images combined.



(a) (b)

Figure 3.3.3i Purification results of the DNA strings with one BcgI recognition site. From (a), a gel segment containing the DNA products of approximately the size of the designated duplex (about 90bp) was extracted, redissolved and run on another agarose gel (b) where clear and narrow DNA bands of approximately the size of the duplex was observed. The duplex was then attached to the DNA frame.

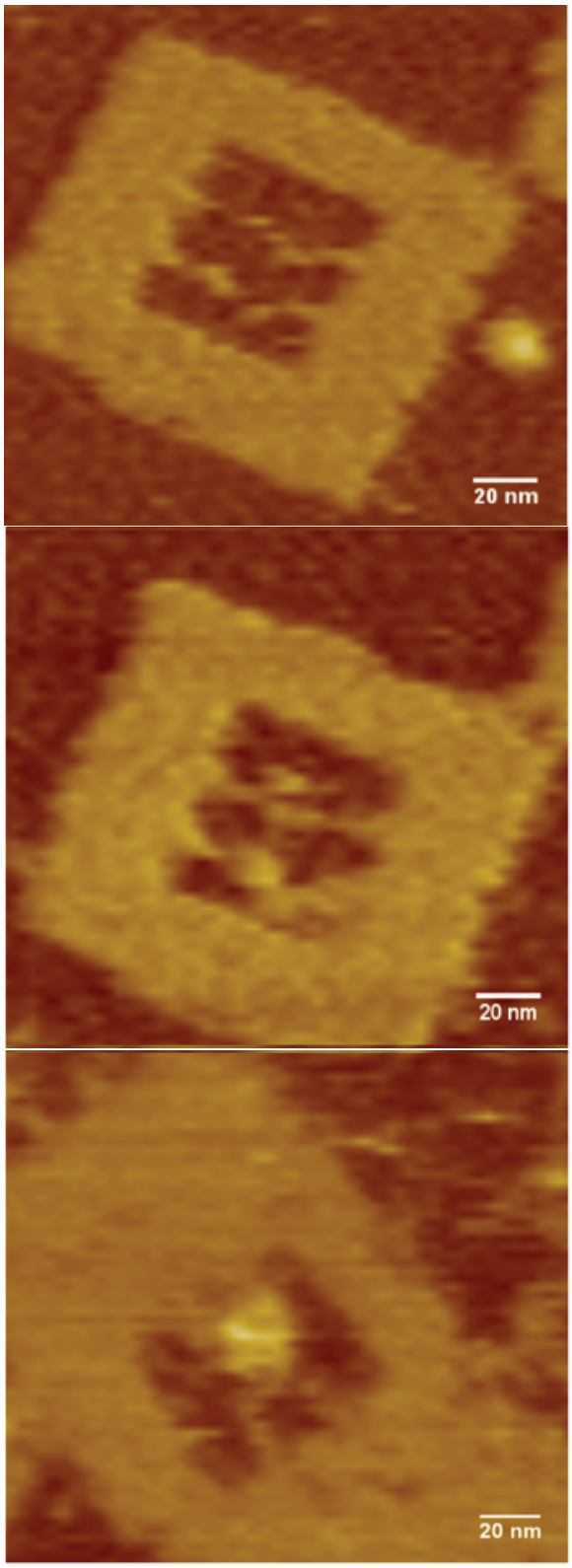


Figure 3.3.3ii **A panel of fluid images of the BcgI molecules interaction with two recognition sites *in trans*.** The AFM images were taken in fluid. The scan rate is 31.4 Hz. The resolution is 512 x 512 pixels. The amplitude setpoint is 210 mV. The drive amplitude is 250 mV.

3.3.4 The four-molecule complex model

We know from Section 3.1.5 that the BcgI complex cuts two DNA recognition sites within the lifetime of the protein-DNA complex, breaking eight phosphodiester bonds via one upstream double-strand break and one downstream double-strand break at each of the two sites. Based on the volume analysis results in Section 3.3.2, it is highly likely the cleavage of the two sites is conducted by a four-molecule complex. To cut all eight phosphodiester bonds within the lifetime of the protein-DNA complex, the four-molecule complex needs to have eight active sites to use for each phosphodiester bond.

Here in Figure 3.3.4 are shown the BcgI four-molecule complexes in both a dry AFM image and in a fluid AFM image and a structure model we propose for the four-molecule complexes. In the model, each molecule interacts at two points with an adjacent molecule, forming 8 active sites at those interacting points. Therefore, the formed four-molecule complexes can cut all eight phosphodiester bonds at once. The need for two BcgIs to form the one active site of the complex means that the cleavage would rarely happen without the completion of the four-molecule complex. Smith's group predicted a similar BcgI four-molecule complex structure (Smith *et al.*, 2013).

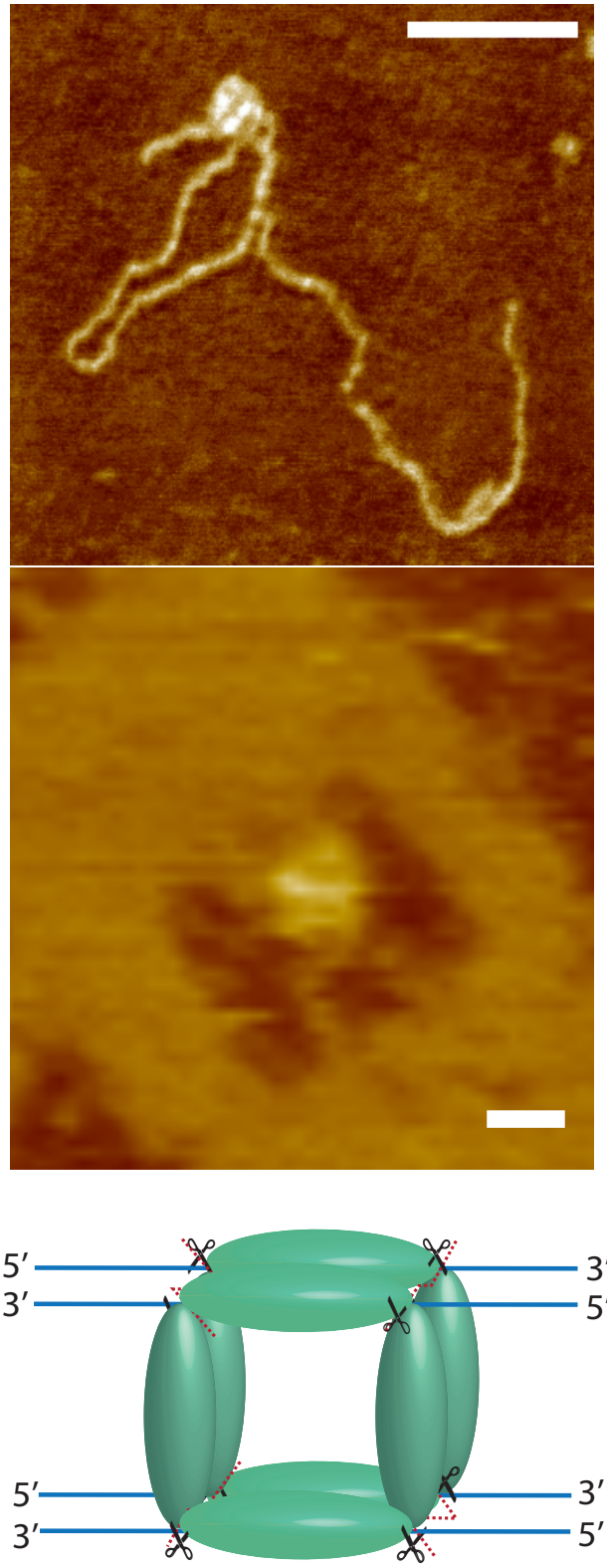


Figure 3.3.4 From the top to the bottom are a dry image and a fluid image of a BcgI complex with two recognition sites brought together and our proposed model of four-BcgI complex. The two recognition sites in the dry image are *in cis*, and those in the fluid image are *in trans*. The scale bar in the dry image represents 100 nm. The scale bar in the fluid image represents 20 nm. In the proposed model, each molecule (The the double-oval-shape represents two BcgIA subunits of a BcgI molecule. The BcgIB subunit is not shown here because it is believed to function to recognise the DNA sites rather than interact with other BcgI molecules (See Section 3.1.5).The oval shape is oversimplified and does not suggest the BcgIA is oval shaped) is interacting at two points with an adjacent molecule, forming 8 active sites at those interacting points. Therefore, the formed four-molecule complexes can cut all eight phosphodiester bonds at once.

3.3.5 BcgI complex-forming mechanism

Of all the DNA loops captured in the images, 94% have starting and ending points that coincide with the two BcgI recognition sites (Figure 3.3.5i). For a four-molecule complex that has the capacity to concertedly cleave two recognition sites, there are two possible mechanisms for the complex formation. The first (illustrated in Figure 3.3.5ii) is by first forming two-molecule BcgI complex at each recognition site, and then joining the two-molecule complexes into a single four-molecule complex that brings the two recognition sites together. The second (illustrated in Figure 3.3.5iii) is by forming the entire four-molecule complex at one recognition site, and then sliding the adjacent DNA through the complex until it catches on the second recognition site. In the latter case, we would expect to observe smaller loops before the sliding process is complete.

Based on the observation in this experiment that almost all the loops were of the maximum size and almost no smaller progressive loops were found, we suspect that the four-molecule BcgI complexes are formed by collision of two two-BcgI complexes, each having bound to a recognition site.

The cutting requires the BcgI four-molecule complex, whose active sites are formed at the meeting points of each two adjacent BcgI molecules, and the fact that the four-molecule complex rarely exists on a single recognition site ensure that cutting at any single recognition site is extremely rare and thus the accuracy of the site recognition is high.

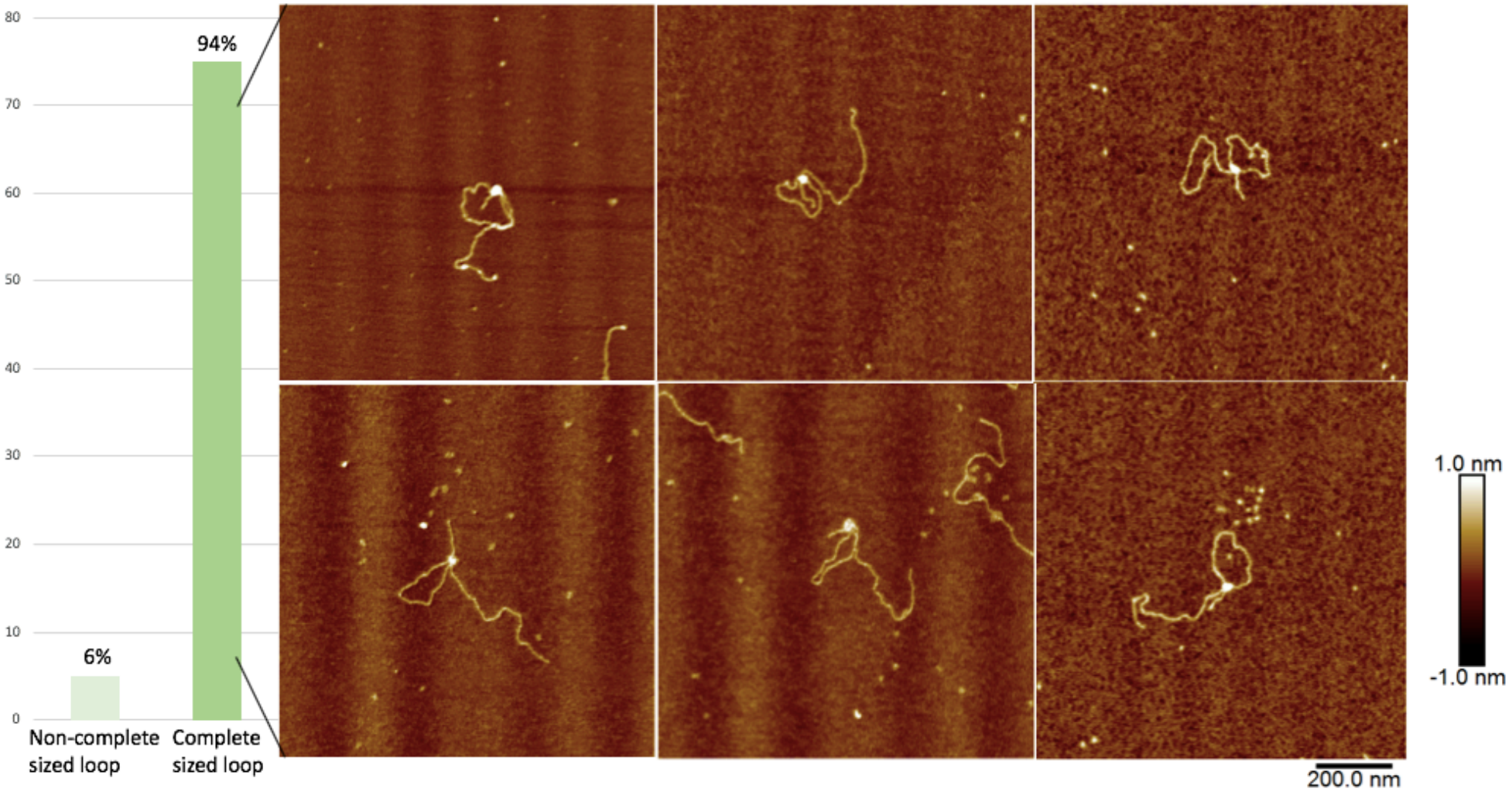


Figure 3.3.5i **A analysis of the loop size.** A study of 80 dry AFM images where loops are easily discerned shows that 94% of the loops have starting and ending points that coincide with the two BcgI recognition sites. The rest 6% has smaller loops.

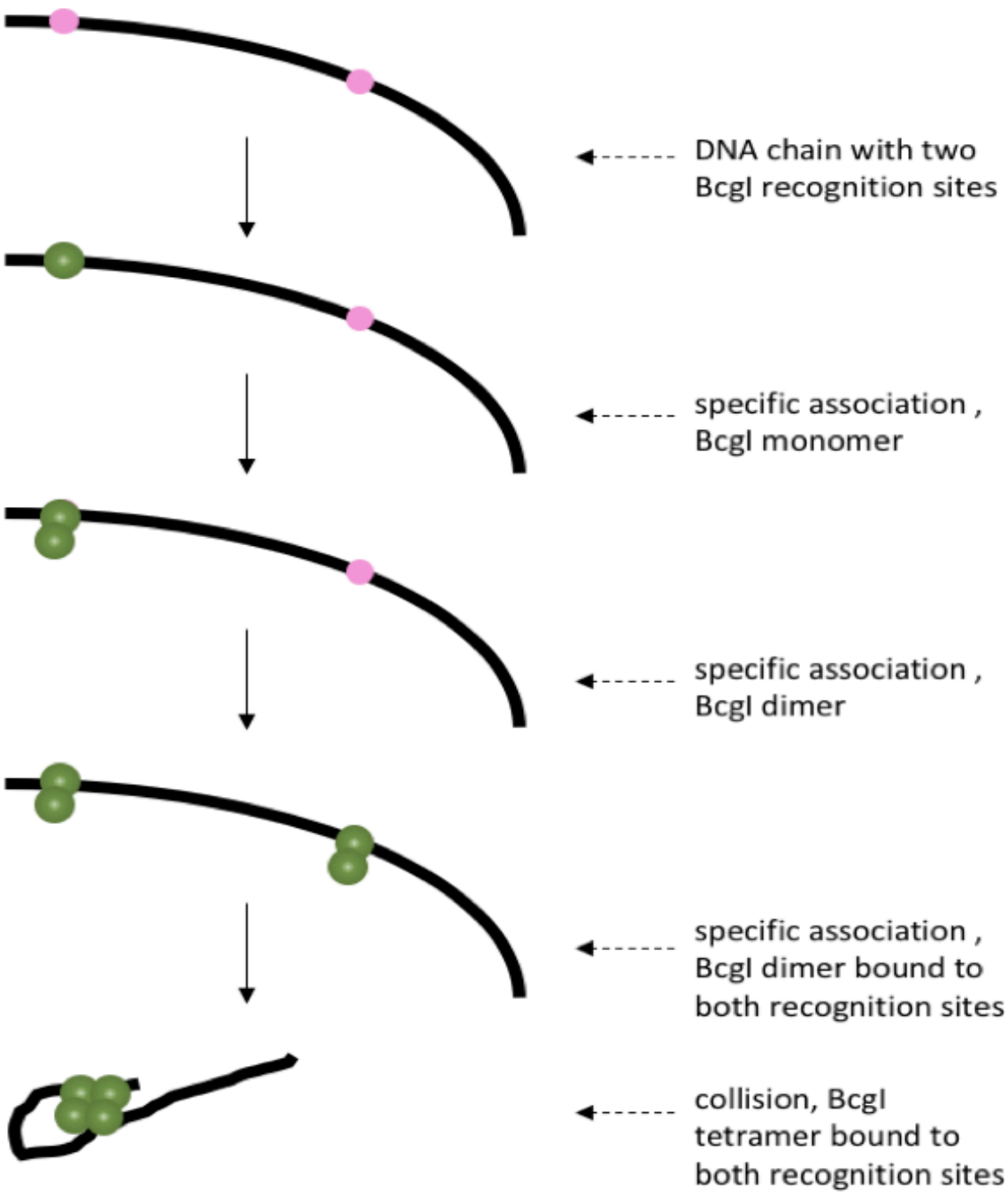


Figure 3.3.5ii One possible way of BcgI complex formation. In this way, first a two-molecule BcgI complex at each of the two recognition sites is formed. And then the two two-molecule complexes come together, forming a single four-molecule complex, during which process the two recognition sites are brought together. If BcgI uses this way, we should expect to observe a high proportion of complete-sized loops with the starting and ending points being the two BcgI recognition sites

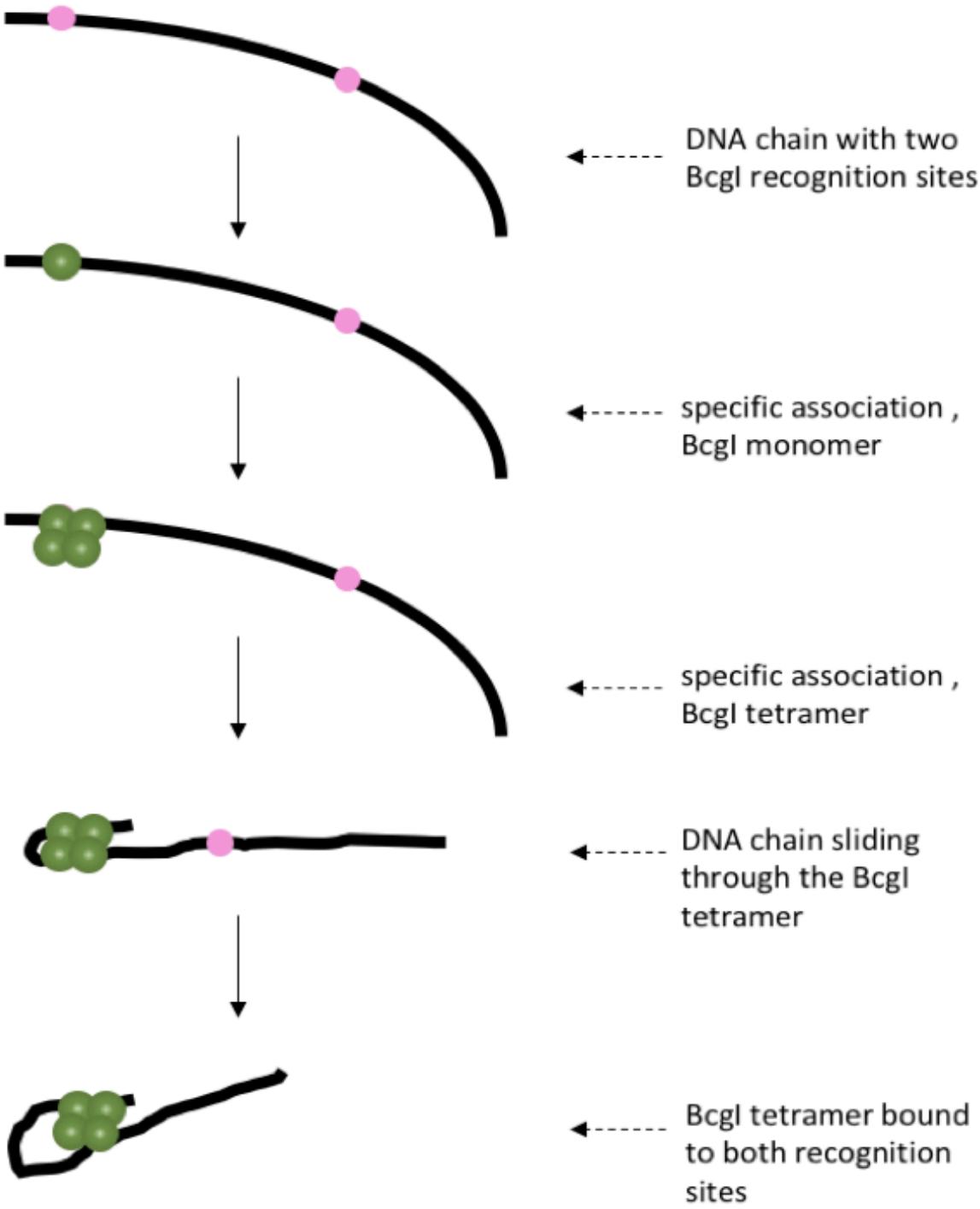


Figure 3.3.5iii The other possible way of BcgI complex formation. In this way the entire four-molecule complex is formed first at one recognition site. The complex then slides the adjacent DNA through it until it catches the second recognition site. If BcgI uses this way, we should expect to observe a high proportion of smaller loops among all the loops captured in the AFM images, because smaller loops are formed before the sliding process is complete.

3.4 Conclusions

Through analysing the dry AFM images of BcgI interacting with a DNA strand which has two BcgI recognition sites and the real-time fluid AFM images of BcgI interacting with two DNA strands each of which has a BcgI recognition site, we have found that BcgI forms four-molecule complex to bring two recognition sites together and cut at eight places concertedly. Through further analysis of the DNA loops observed in the dry AFM images, we have found that the BcgI forms the four-molecule complex by first forming a two-molecule BcgI complex at each recognition site, and then colliding the two-molecule complexes into a single four-molecule complex that brings the two recognition sites together.

As most restriction enzymes need multiple recognition sites on the DNA to conduct cleavage, understanding how BcgI forms the complex and binds two recognition sites should help understand how the high accuracy of the DNA cleavage is achieved by many restriction enzymes.

Chapter 4 Conclusions

For many years, because the best-studied Type IIP restriction enzymes recognize short and solitary base pairs, the applications of restriction enzymes in recombinant DNA technology has been limited to generating sticky DNA ends. Only recently have restriction enzymes been used in more complex gene editing tools. For example, artificial restriction enzymes (AREs) have been generated as a combination of one DNA recognition domain and a nuclease domain (Enghiad and Zhao, 2017). Broadly speaking, zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and clustered regularly interspaced short palindromic repeat associated enzyme 9 (CRISPR-Cas9) all fall into the artificial restriction enzyme category (Enghiad and Zhao, 2017). Compared with naturally occurring restriction enzymes, AREs tend to have higher and more designable specificity. However, their low activity and availability largely limit their applications (Enghiad and Zhao, 2017).

Recent work has shown that by using prokaryotic Argonaute of the archaeon Pyrococcus furiosus (PfAgo) as the platform, new AREs can be made with any sequence specificity and defined sticky ends of varying length (Enghiad and Zhao, 2017). Such PfAgo-based restriction enzymes use a single PfAgo protein plus short DNA guides for targeting any one or more sites (Enghiad and Zhao, 2017). Because high-level purification of the PfAgo protein is easily achievable and synthetic short DNA are inexpensive, the PfAgo-based AREs are likely to be a powerful tool in recombinant DNA technology (Enghiad and Zhao, 2017). Such PfAgo-based AREs have promising applications in DNA fingerprinting and DNA cloning. (Enghiad and Zhao, 2017). This is a big step towards making AREs have more desirable qualities of naturally occurring restriction enzymes.

Genome editing tools are increasingly applied in the medical field. ZFNs, TALENs and CRISPR-Cas9 have been used to modify T cells (Carroll, 2017, Tebas *et al.,* 2014, Menger *et al.,* 2016, Cyranoski, 2016, Kaiser, 2016). These attempts used a method called *ex vivo* treatment, in which the editing is conducted in the laboratory on cells derived from the patient and then the edited cells are transferred back to the patient (Carroll, 2017). However, *in vivo* treatment, which requires the genome editing tool and the DNA sequence donor to be delivered directly to the target site in the body, has yet to be achieved.

Safety concerns limit such *in vivo* applications. A major hurdle is off-target cleavage. ZFNs and TALENs tend to conduct the cleavage near the recognition sites rather than at the recognition sites, and CRISPR-Cas9 also has imperfect recognition. These off-target cleavage and mutagenesis are largely intolerable in many medical applications (Carroll, 2017).

Another major obstacle is the poor integration of the desired DNA sequence to the cleaved DNA strand. The cellular DNA repair machinery goes one of the two ways after the double-strand break. In homology-dependent repair (HDR), a donor sequence is copied and inserted to the break (Carroll, 2017). In non-homologous end joining (NHEJ), random base pairs may be inserted to the break with the sole purpose of bringing the ends back together (Carroll, 2017). The cellular DNA repair machinery employs NHEJ much more frequently than HDR (Carroll, 2017). The inclination towards NHEJ should probably be attributed to its speed. But this quick solution largely limits the more accurate HDR, and thus reduces the success rate of genome integration (Carroll, 2017). Despite many trials, only marginal progress has been made to increase the ratio of HDR to NHEJ (Beumer *et al.,* 2008, Bozas *et al.*, 2009, Chu *et al.,* 2015, Maruyama *et al.*, 2015).

A better understanding of the genome editing tools will be important not only for their medical application, but also for agriculture and other world health issues (For a good summary of these issues, see Kohn *et al.,* 2016). Genome editing on crops and food animals has been conducted with much less caution than in medical practice. As a result, we already have wheat that is disease resistant, potatoes that don't sweeten on storage, and food animals that carry more muscle mass (Carroll, 2017). Bolder and more recent approaches aim at sterilising female mosquitoes or inactivating the mosquito genes required for parasite growth (Hammond *et al.*, 2016, Gantz *et al.*, 2015). These approaches can largely reduce disease transmission in areas where disease treatment is challenging, but they might at the same time eliminate a whole species and damage the ecosystem. Much caution is required.

The study of restriction enzymes is important for broader and safer use in genome editing, and for understand more complex genome editing tools to facilitate their application. For example, the concerted cleavage of the DNA demonstrated by restriction enzymes such as BcgI normally allows better precision in site recognition and may provide a template for solving the off-target problems of more complex tools.

We have introduced very powerful methods for understanding the mechanics of protein-DNA interactions. The combined HS-AFM and DNA origami technique are powerful for quantitative studies as they provide high spatial and temporal resolution (e.g. my experiments are the first quantification of EcoRV translocation distance at the sub 40bp length-scale). Furthermore, the analytic pipelines developed using ImageJ and R provide a template for further quantitative studies of protein-DNA interactions using AFM.

For future studies of protein-DNA interactions, studies about CRISPR-Cas9 should be of immediate importance. CRISPR-Cas9 is a bacterial adaptive immune system that utilises the uptake of foreign DNA from invasive genetic elements such as viruses and plasmids (For recent reviews see Zhang *et al*., 2014, Doudna and Charpentie, 2014). There are some important questions to be answered. First, how CRISPR-Cas9 distinguishes self from non-self remains unsolved. This causes danger in more complicated applications (Deveau *et al*., 2008). Second, although it has been known for decades that dsDNA breaks in mammalian cells trigger homologous recombination, the design of a template to direct the recombination still requires further development (Cong *et al*., 2013, Smith *et al*., 1995). Moreover, off-target effects are a significant obstacle, as Cas9 cleavage has been found to target sequences partially homologous to the guiding RNAs (Fu *et al*., 2013). Solving these problems is the priority of many seeking to use CRISPR-Cas9 for gene therapy. Some methods have already been proposed which require the modification of the enzyme Cas9 (Mali *et al*., 2013, Ran *et al*., 2013).

Some studies of CRISPR-Cas9 using HS-AFM have already extended our understanding. For example, Shibata's group shed light on many aspects of the complex tool, including the structural stability of Cas9-RNA, Cas9 domain conformational change upon DNA binding, the dissociation of Cas9-RNA from the DNA after cleavage, and the DNA target site search by Cas9-RNA (Shibata *et al.*, 2017).

As the role of HS-AFM becomes more prominent in studying protein-DNA interactions, I dare to suggest that there will be a growing number of studies on CRISPR using HS-AFM. I hope that this work will encourage and inform these studies.

Overall, our studies have increased our knowledge on restriction enzymes, especially their mechanisms used for translocation on the DNA. We have learnt that:

- EcoRV has high affinity to its recognition sites.
- Hops are clearly employed by EcoRV for translocating on DNA.
- Slides and short hops under 20 bp and slightly above 20 bp are more common than long hops or jumps, and travel distance larger than 40 bp is much rarer.
- EcoRV tends to travel short distance using sliding when at or next to the recognition sites.
- The general trend is the farther the EcoRV is from the recognition sites, the longer the distance it travels per second.
- The direction of travelling of an EcoRV at any time point is usually the direction of the nearest recognition site.
- BcgI forms four-molecule complex.
- BcgI forms the four-molecule complex by first forming two-molecule BcgI complex at each recognition site, and then colliding the two-molecule complexes into a single four-molecule complex that brings the two recognition sites together.

In addition, our quantitative study of Protein translocation on DNA, which use atomic force microscopy and DNA origami and data analysis tools including image processing tools and programming language, has proved effective. I have diligently recorded all steps I have taken in the Material and Methods sections and all the codes in the Appendix. I hope they will be of help to those who would also like to make AFM image analysis more real-time and quantitative.

Epilogue

It was only in the 19th century that medicine developed as a science, when the microscope revealed that cells are fundamental building blocks of the body and diseases are transmitted by recognizable entities of pathology. Despite this shift in attention, the microscope itself hardly became the centre of attention. The seemingly lackluster field of microscopy, however, has attracted some of the most curious and creative minds in history. They have devoted their passion and energy into improving the technique, and to diversifying and refining its use.

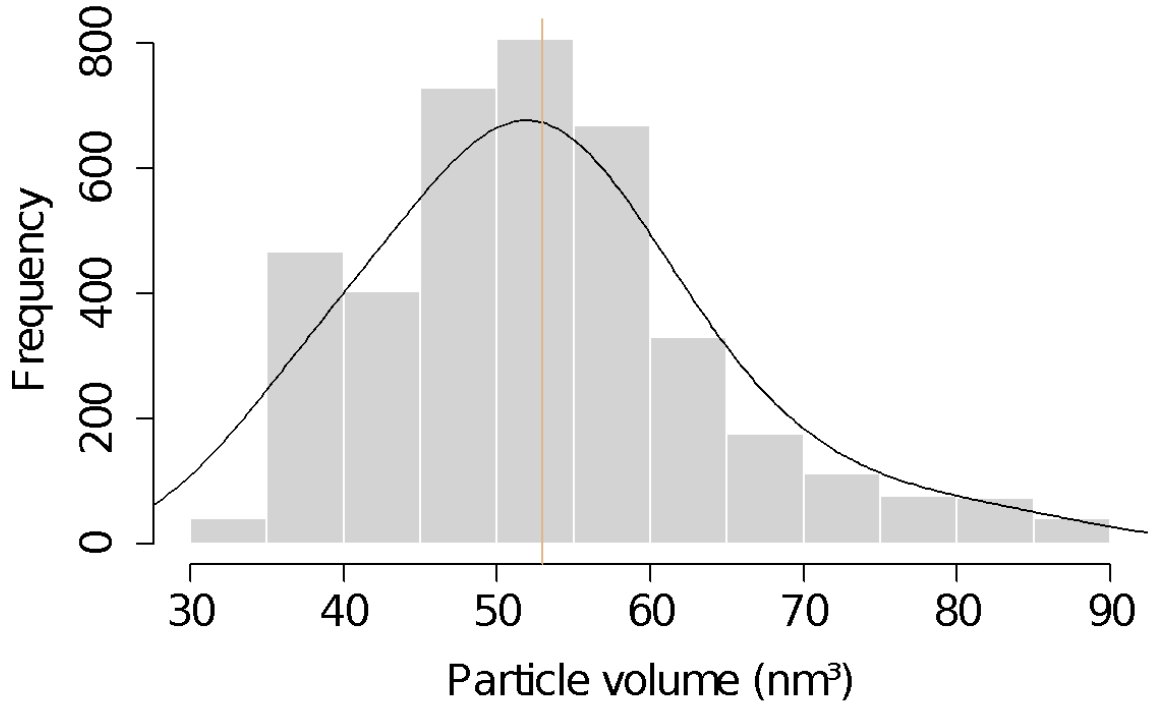
Thanks to these pioneers, microscopy has grown to encompass many capabilities. It seems that microscopy will have a prominent position in biological development for a long time to come. As a PhD student, I am thrilled and humbled to be able to contribute to this field by conducting quantitative research on protein-DNA interaction using fast-scan atomic force microcopy, DNA origami, and analysis platforms.

Appendix

1. Histogram of EcoRV volume measurements from the AFM dry images
2. Independent Samples T-Test Assuming Different Variances for comparing the mean of the particle's volume in the fluid images with the mean of EcoRV in dry images
3. R scripts for data-analysis figures in 2.3.1
4. R scripts for data-analysis figures in 2.3.3
5. R scripts for data-analysis figures in 3.3.2
6. More images like Figure 2.3.1iii
7. More images like Figure 2.3.1iv and 2.3.2
8. More images like Figure 3.3.1ii
9. More images like Figure 3.3.3ii

Appendix 1

Histogram of EcoRV volume measurements from the AFM dry images



A brown line marks the mean of 53nm^3

Appendix 2

Independent Samples T-Test Assuming Different Variances for comparing the mean of the particle's volume in the fluid images with the mean of EcoRV in dry images

t-Test: Two-Sample Assuming Unequal Variances

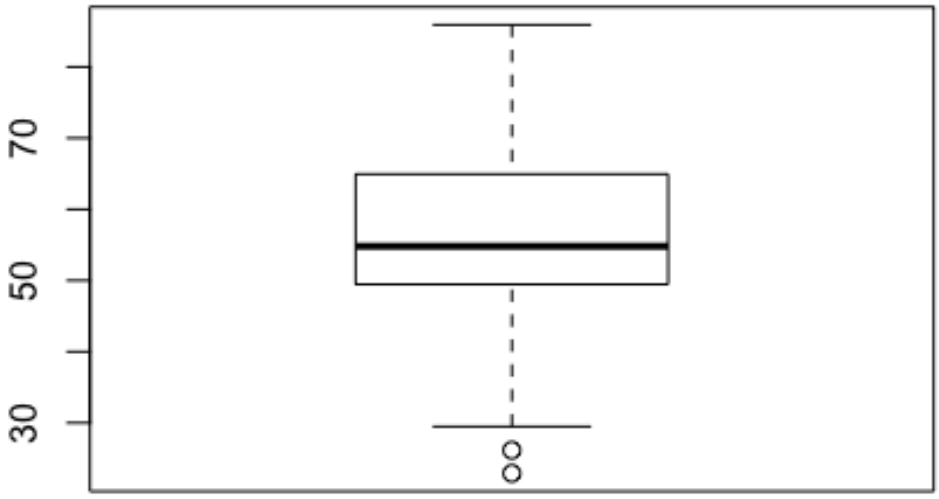
	Volume dry	*Volume fluid*
Mean	56.4178421	56.0758116
Variance	346.412342	121.425176
Observations	2897	106
Hypothesized Mean Difference	0	
df	128	
t Stat	0.30409063	
P(T<=t) one-tail	0.38077632	
t Critical one-tail	1.65684523	
P(T<=t) two-tail	0.76155264	
t Critical two-tail	1.97867085	

Null hypothesis: There is no significant difference between the two means
P=0.76>0.05
Null hypothesis therefore cannot be rejected

Appendix 3

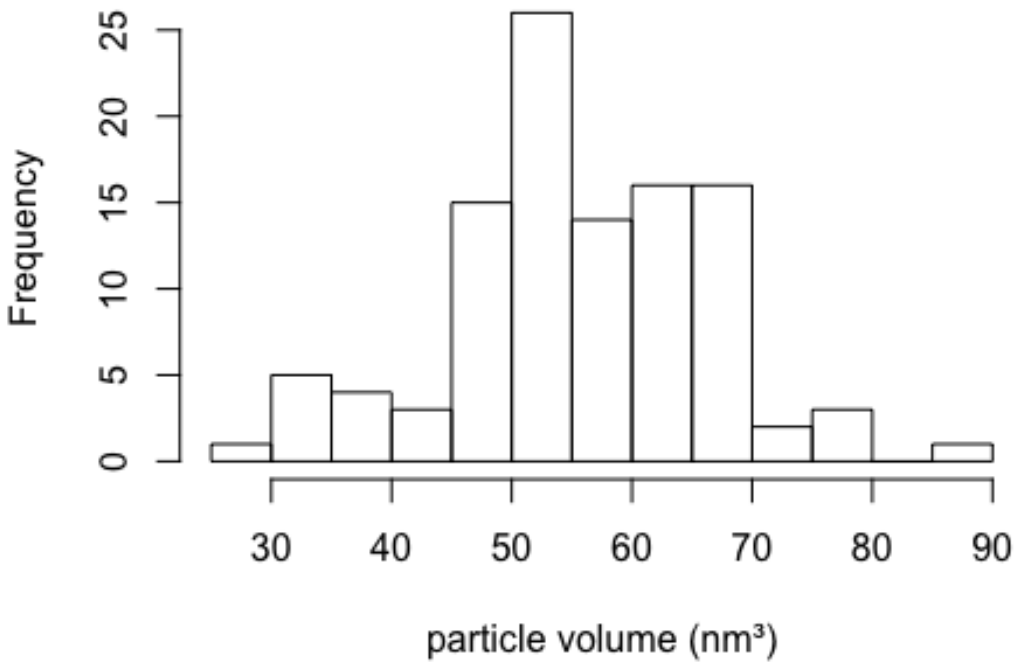
R scripts for data-analysis figures in 2.3.1

```
##EcoRV volume analysis - EcoRV on the DNA string
setwd("/Users/dandanliu/Desktop/Rplots")
data1<-read.csv('volume analysis EcoRVs on the DNA strings.csv')
colnames(data1)<-c('volume','particle')
summary(data1$volume)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   22.91   49.45   54.81   55.49   64.94   85.92       1

boxplot(data1$volume)
```



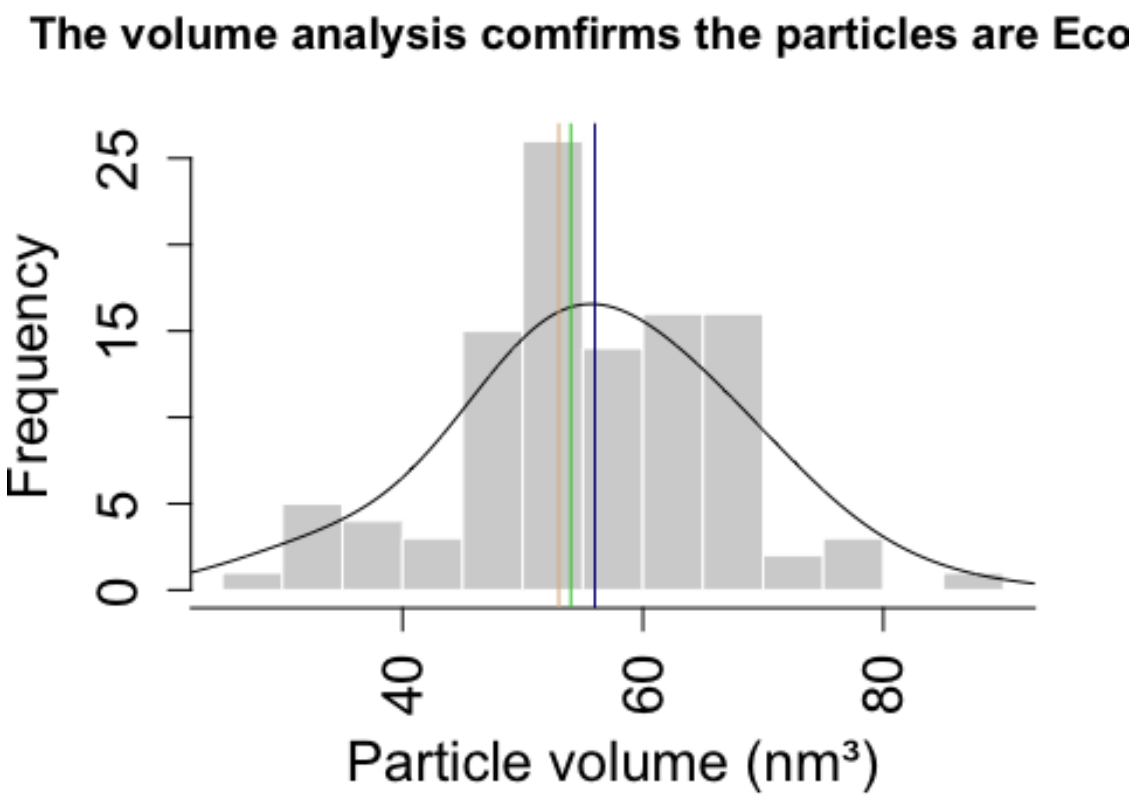

```
boxplot.stats(data1$volume)

## $stats
## [1] 29.48257 49.45500 54.80850 64.94357 85.91661
##
## $n
## [1] 108
##
## $conf
## [1] 52.45369 57.16332
##
```

```
## $out
## [1] 22.91330 26.15201

## 22.91330 and 26.15201 are outliers


volume1<-data1$volume[data1$volume>=29.48257]
data2<- data1[data1$volume>=29.48257,]
summary(data2$volume)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   29.48   49.45   54.81   56.08   64.94   85.92       1

myhist<-hist1<-hist(volume1, xlab = 'particle volume (nm³)',
                    main=NA,breaks=20)
```




```
multiplier<-myhist$counts/myhist$density
mydensity<-density(na.omit(volume1), adjust=2)
mydensity$y<-mydensity$y*multiplier[1]
par(xpd=FALSE)
plot(myhist,col="light gray", border = "white", xlab = 'Particle volume
(nm³)', main="The volume analysis comfirms the particles are EcoRV",
xaxt="n", cex.axis=1.5, cex.lab=1.5)
axis(1,at=seq(0,100,by=20),las=2, cex.axis=1.5)
lines(mydensity, col="black")
##mean
abline(v=56, col="dark blue")
##reference volume
```

```
abline(v=54, col="limegreen")
##mean from dry image volume analysis
abline(v=53, col="burlywood")
```



Raw data used for the previous R script

Particle volume (nm^3)

22.9132996	53.0115733	64.9435733
26.1520133	53.0115733	64.9435733
29.4825721	53.0115733	64.9435733
30.615	53.0115733	64.9435733
31.7592029	53.0115733	67.1133788
31.7592029	53.0115733	67.1133788
32.9155733	53.0115733	67.1133788
32.9155733	53.0115733	67.1133788
35.2663867	53.0115733	67.1133788
35.2663867	53.0115733	67.1133788
36.4616146	54.8085038	67.7716667
38.8936754	54.8085038	67.9255267
44.2074713	54.8085038	69.2961367
44.2074713	54.8085038	69.2961367
44.2074713	54.8085038	69.2961367
45.9455267	56.3860267	69.2961367
45.9455267	56.3860267	69.2961367
49.455	56.6183867	69.2961367
49.455	56.6183867	69.2961367
49.455	56.6183867	69.2961367
49.455	56.6183867	71.4922396
49.455	56.6183867	73.70208
49.455	56.6183867	78.1315733
49.455	56.6183867	78.1315733
49.455	56.6183867	78.1645433
49.455	58.4416146	85.9166146
49.455	58.4416146	
49.455	58.4416146	
49.455	58.4416146	
49.455	60.27858	
51.2272029	60.27858	
51.2272029	60.64125	
51.2272029	60.64125	
51.2272029	60.64125	
51.2272029	60.64125	
53.0115733	62.7863279	
53.0115733	62.7863279	
53.0115733	62.7863279	
53.0115733	63.9952933	
53.0115733	64.9435733	
53.0115733	64.9435733	

Appendix 4

R scripts for data-analysis figures in 2.3.3

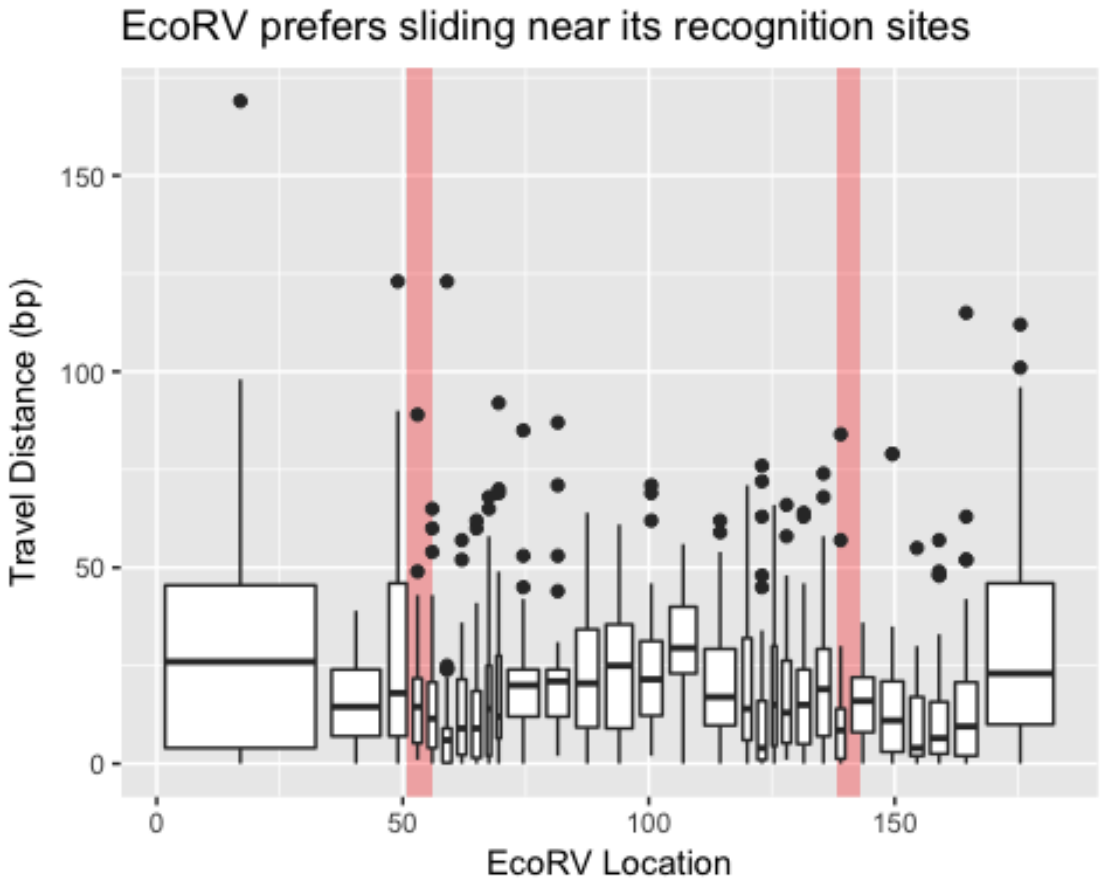
```
library(tidyverse)

## ── Attaching packages ─────────────────────────────────── tidyverse
1.2.1 ──

## ✓ ggplot2 2.2.1     ✓ purrr   0.2.5
## ✓ tibble  1.4.2     ✓ dplyr   0.7.5
## ✓ tidyr   0.8.1     ✓ stringr 1.3.1
## ✓ readr   1.1.1     ✓ forcats 0.3.0

## ── Conflicts ────────────────────────────────────
tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()

setwd("/Users/dandanliu/Desktop/Rplots")
traj=read_csv('trajectory analysis.csv')

## Parsed with column specification:
## cols(
##   sampling.unit = col_integer(),
##   ecorv.location = col_integer()
## )

#make groups
na.indices=which(is.na(traj$sampling.unit))

traj<-mutate(traj,
             obs=seq_along(sampling.unit) %>%
               map_dbl(~sum(na.indices<.))) %>%
  mutate(obs=obs+1)%>%
  filter(!is.na(sampling.unit))


traj<-traj %>%
  group_by(obs)%>%
  mutate(jump=lead(ecorv.location)-ecorv.location,
         dist=abs(jump))

traj<-traj %>%
  group_by(obs)%>%
  mutate(time=1:n(),scaled_time=time/max(time))

ggplot()+
  geom_rect(aes(xmin=c(51,138),xmax=c(56,143),ymin=-
Inf,ymax=Inf),fill='red',alpha=0.3)+
  geom_boxplot(traj,mapping=aes(x=ecorv.location,
                                y=dist,
```

```
                                group=cut_number(ecorv.location,30)))+
  labs(x='EcoRV Location',y='Travel Distance (bp)',
       title='EcoRV prefers sliding near its recognition sites')+
  ggsave('dist_boxplot.pdf')
```

```
## Saving 5 x 4 in image
```

```
## Warning: Removed 54 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 54 rows containing non-finite values (stat_boxplot).
```

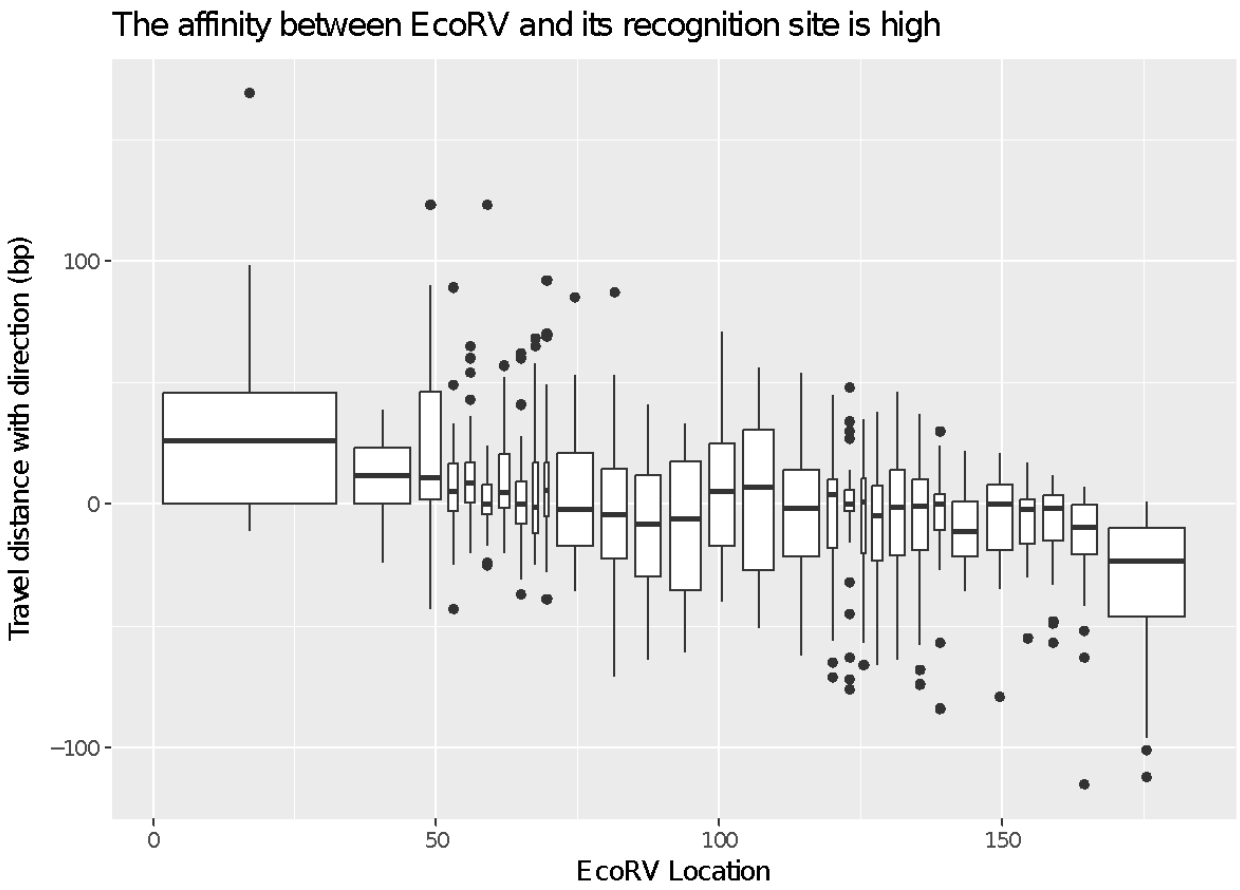



```
ggplot()+
  geom_rect(aes(xmin=c(51,138),xmax=c(56,143),ymin=-
Inf,ymax=Inf),fill='red',alpha=0.3)+
  geom_boxplot(traj,mapping=aes(x=ecorv.location,
                                y=jump,
                                group=cut_number(ecorv.location,30)))+
  labs(x='EcoRV Location',y='Travel distance with direction (bp)',
       title='The affinity between EcoRV and its recognition site is
high')+
  ggsave('jump_boxplot.pdf')
```

```
## Saving 5 x 4 in image
```

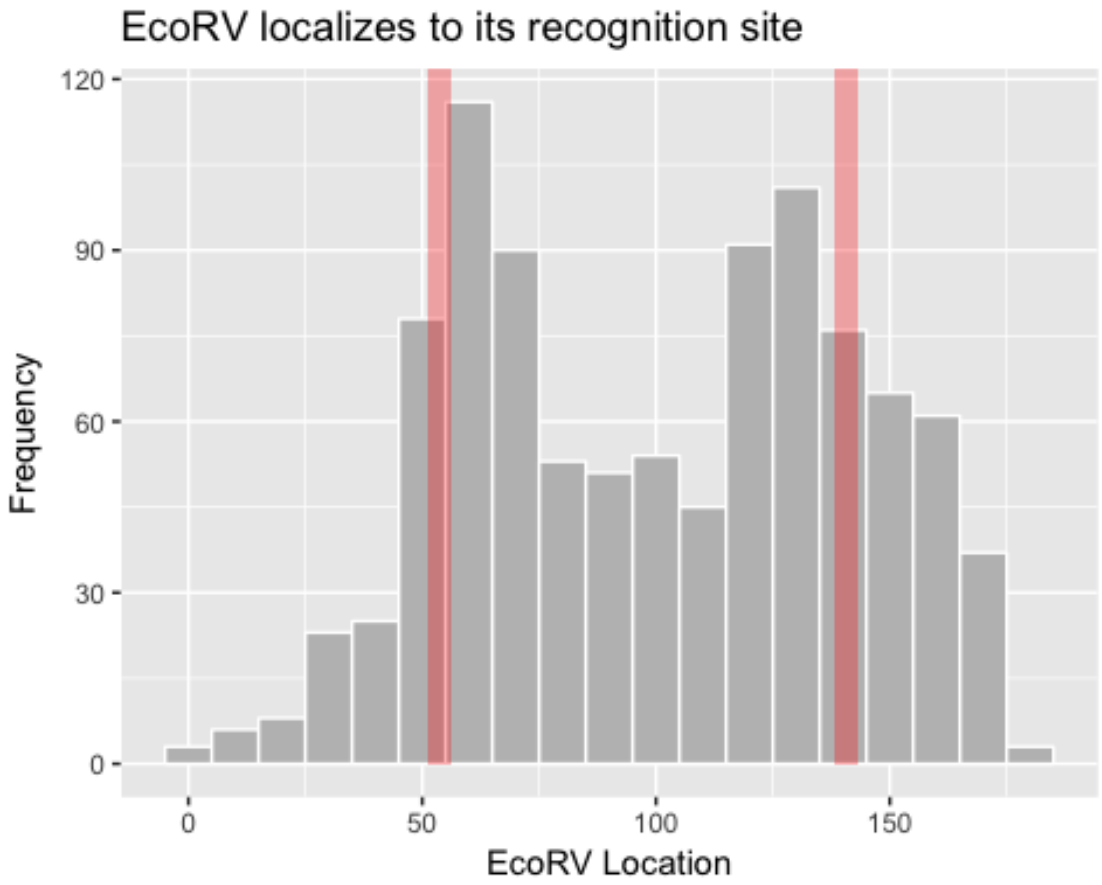
```
## Warning: Removed 54 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 54 rows containing non-finite values (stat_boxplot).
```

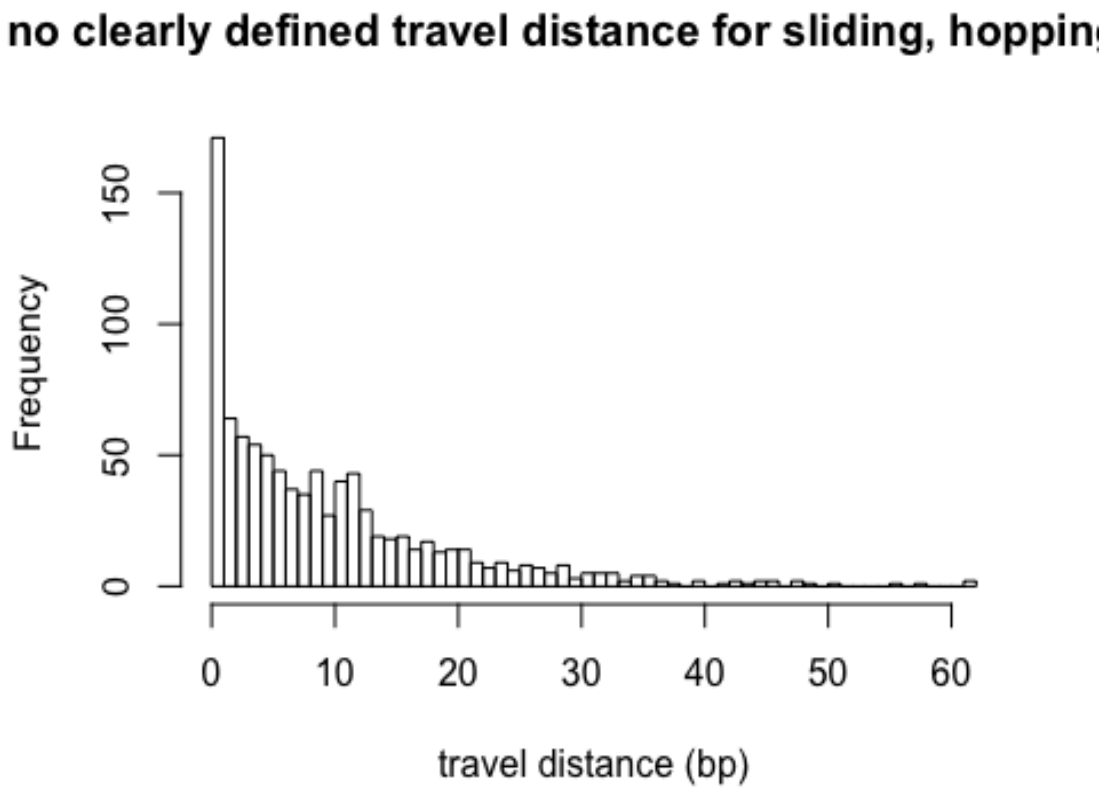



```
ggplot()+
  geom_histogram(data=traj,aes(x=ecorv.location),binwidth=10,
                 color='white',fill='gray')+
  geom_rect(aes(xmin=c(51,138),xmax=c(56,143),ymin=0,ymax=Inf),
            fill='red',
            alpha=0.3)+

  labs(title='EcoRV localizes to its recognition site',
       x='EcoRV Location',y='Frequency')+
  ggsave('EcoRV_occupancy.pdf')
```

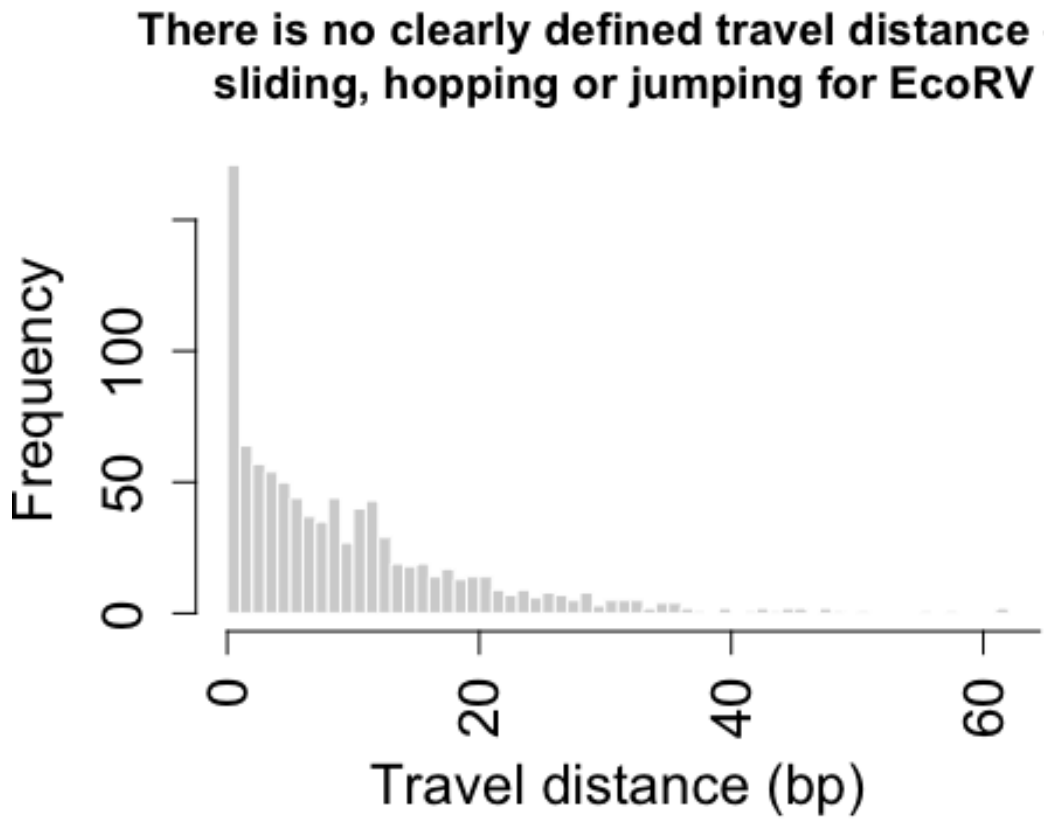
```
## Saving 5 x 4 in image
```




```
##EcoRV travel distance histogram
setwd("/Users/dandanliu/Desktop/Rplots")
data1<-read.csv('Travel distance.csv')

summary(data1)

##      X84.5
##  Min.   : 0.000
##  1st Qu.: 2.000
##  Median : 7.000
##  Mean   : 9.712
##  3rd Qu.:13.500
##  Max.   :61.500

myhist<-hist1<-hist(data1[,1], xlab = 'travel distance (bp)',
                    main="There is no clearly defined travel distance
for sliding, hopping or jumping",breaks=70)
```




```
plot(myhist,col="light gray", border = "white", xlab = 'Travel distance
(bp)', main="There is no clearly defined travel distance of \n sliding,
hopping or jumping for EcoRV", xaxt="n", cex.axis=1.5, cex.lab=1.5)
axis(1,at=seq(0,180,by=20),las=2, cex.axis=1.5)
```


There is no clearly defined travel distance of sliding, hopping or jumping for EcoRV
Frequency
0
50
100
0
20
40
60
Travel distance (bp)

Raw data used for the previous R script

Base pair point (bp)

Set 1	126	94	52	62	47	69	62	79	69	66	29
	29	121	50	7	7	18	7	5	62		
Set 2	34	67	42	50	140	154	156	153			
Set 3	0	169	57	100	82	99	168	133			
Set 4	132	128	123	133	148	69	30				
Set 5	120	106	70	68	50	173	122	152	131	128	107
	163	147	144	158	162	120	118				
Set 6	144	121	126	134	104	132	99	59	58	51	99
	131	177	166	103	174	152	166	51	100	106	129
Set 7	124	121	127	104	89	105	99	69	43	56	57
	56	70	76	52	62						
Set 8	34	87	80	57	93	52	74	127	69	78	
Set 9	153	136	68	136	147	149	167	151	116	115	113
	111	84	67	55							
Set 10	140	145	166	173	147	68	68	63	63	72	52
	53	57	37	54	35	47	99	104	145		
Set 11	128	120	147	139	131	110	150	150	161	160	143
	164	145	146	122	122	125	131	139	154		
Set 12	164	151	172	140	164	144	159	150	160	160	151
	132	88	64	54	11	6	31	70	88		
Set 13	88	88	66	57	70	80	65	77	98	160	111
	103	139	147	137	170	156	133				
Set 14	140	140	113	113	92	101	88	103	90	26	69
	76	40	58	67	92	52	67				
Set 15	133	175	164	163	145	123	91	85	53	58	34
	34	34	60	43	50	91	54	57	57		

Set 16	77	64	59	64	79	91	73	50	52	34	36
	61	66	59	63	57	59	54	60	64		
Set 17	136	78	52	85	57	64	69	139	138	127	131
	105	136									
Set 18	65	65	65	65	65	65	65	65	65	38	19
	115	99	121	166	156	101	98	67	65	72	56
	60										
Set 19	122	122	122	122	122	122	122	122	122	46	22
	22	31	32	60	68	69	69	81	93		
Set 20	2	53	45	68	69	118	68	122	106	66	63
	61	70	89	81	60	183	183	173	167		
Set 21	154	171	110	82	51	61	61	70	114	120	129
	137	157	169	146	133	129					
Set 22	82	58	52	57	67	65	93	123	130	124	172
	172	126	95	112	50	54	56				
Set 23	62	91	55	44	82	105	96	87	51	52	37
	76	93	111	95	70	50	72	72	60		
Set 24	61	113	167	158	167	158	163	166	165	165	161
	161	167	167	150	128	124					
Set 25	130	135	139	127	121	127	104	128	134	161	165
	167	167	167	115	134	155	130	130	138		
Set 26	126	144	145	109	86	61	59	34	82	89	59
	73	96	61	50	50	69	57	69	52		
Set 27	37	39	78	74	52	73	51	58	54	68	126
	126	103	106	140	154	145	128	113	98		
Set 28	120	90	86	127	123	127	100	102	123	150	126
	135	127	103	86	99	125	101	111	124	130	129
	140										
Set 29	79	65	52	69	74	95	54	34	120	156	154
	165	144	166	114	143	159	157	109	161		
Set 30	154	143	133	147	141	125	137	138	128	105	105
	88	69	71	113	131	132	140	138	133		
Set 31	154	153	129	126	127	155	158	157	153	154	158
	160	155	153	155	151	155					

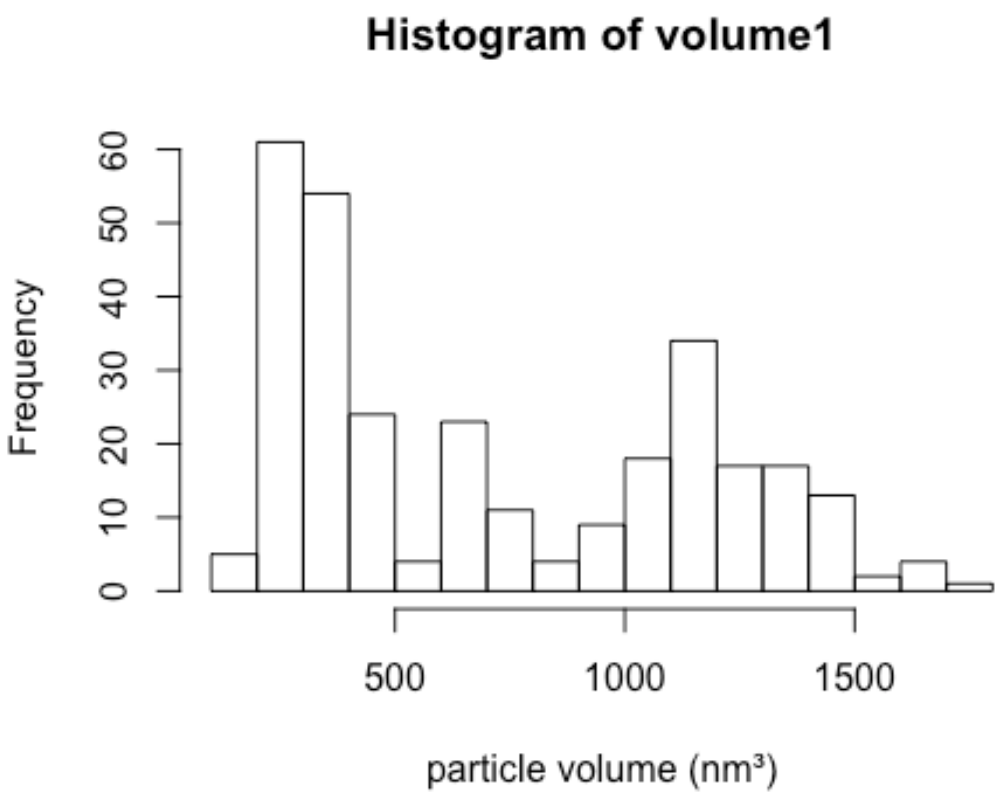
Set 32	68	85	48	28	17	60	60	60	60	60	60
	60	84	13	111	81	168	169	68			
Set 33	135	114	141	121	137	140	126	129	167	125	120
	135	121	97	130	129	147	126	69	138		
Set 34	126	158	158	125	121	56	121	131	68	54	63
	120	128	123	51	84	110	145	140			
Set 35	105	54	81	134	115	67	132	153	157	140	131
	67	122	127	61	61	41	38	50	97	48	55
	63										
Set 36	56	116	153	137	136	173	77	122	120	64	124
	112	121	66	128	161	153	127	125	154	124	123
	137	130									
Set 37	51	125	140	56	71	156	134	135			
Set 38	50	96	77	70	81	83	104	66	51	120	162
	160	159	152	165	126	152	164	162	122		
Set 39	143	143	143	137	144	145	126	134	134	112	102
	99	82	59	71	107	124	128	129	151		
Set 40	140	140	145	123	120	119	151	151	151	151	151
	167	163	154	152	155	153	158				
Set 41	150	154	156	156	156	157	162	163	165	165	151
	139	128	80								
Set 42	56	110	69	161	104	129	95	92	121	125	160
	169										
Set 43	54	143	132	152	131	112	127	138	81	101	75
	95	88	106	82	59	50	53	50	21		
Set 44	48	51	48	55	56	52	101	108	78	69	100
	124	158	135	125	126	124	122	120	121	128	120
	133	110	152								
Set 45	89	51	39	37	50	37	37	37	38	55	61
	83	127	126	109	138	139	140	140	141		
Set 46	146	120	126	125	93	125	59	66	56	86	99
	124	61	61	58	67	58	58	52	60	60	60
Set 47	140	140	170	170	135	167	146	135	61	97	36
	60	69	85	117	131	107	121	114	120		

Set 48	133	149	149	157	109	65	58	58	59	50	61
	94	76	100	114	55	68	51				
Set 49	121	94	122	77	94	89	81	59	55	78	90
	98	60	72	55	40	33	30	24	29		
Set 50	30	60	68	66	61	63	68	69	69	41	25
	65	80	91	98							
Set 51	140	142	141	120	120	102	85	83	89	54	29
	69	66	35	68	45	82	79	63	89		
Set 52	77	79	53	50	98	123	120	124	134	115	92
	116	142	117	128	139	114	75	73	50	68	99
	145	120	66	107							
Set 53	65	87	90	128	89	51	62	68	50	51	57
	57	67	68	54	87	68	67	72	70	81	57
Set 54	69	42	50	67	59	81	77	96	96	92	123
	130	144									

Appendix 5

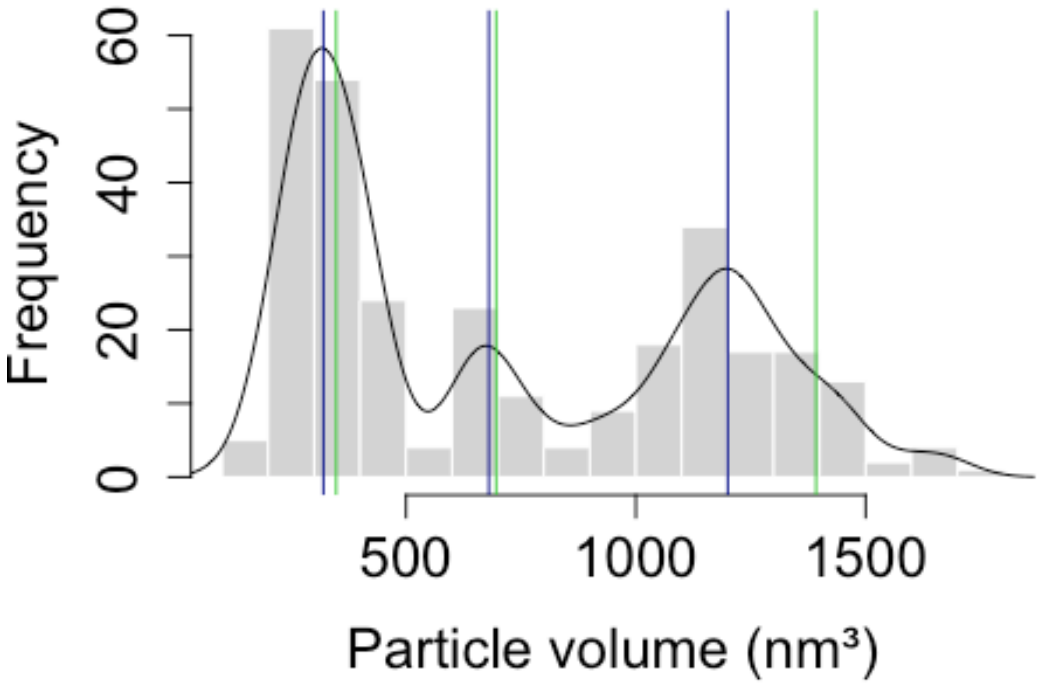
R scripts for data-analysis figures in 3.3.2

```
##BcgI volume analysis
setwd("/Users/dandanliu/Desktop/Rplots")
data1<-read.csv("BcgI complexes .csv")
colnames(data1)<-c('particle','volume')
summary(data1)

##     particle       volume
##  Min.   :  2   Min.   : 122.7
##  1st Qu.: 77   1st Qu.: 315.4
##  Median :152   Median : 642.9
##  Mean   :152   Mean   : 719.0
##  3rd Qu.:227   3rd Qu.:1163.3
##  Max.   :302   Max.   :1705.9

volume1<-data1$volume

myhist<-hist1<-hist(volume1, xlab = 'particle volume (nm³)',
                    breaks=20)
```




```
multiplier<-myhist$counts/myhist$density
mydensity<-density(volume1, adjust=0.5)
mydensity$y<-mydensity$y*multiplier[1]
```

```
par(xpd=FALSE)
plot(myhist,col="light gray", border = "white", xlab = 'Particle volume
(nm³)', main=NA, cex.lab=1.5, cex.axis=1.5)
lines(mydensity, col="black")
abline(v=c(320,680,1200), col="dark blue")
abline(v=c(348,696,1392), col="limegreen")
```



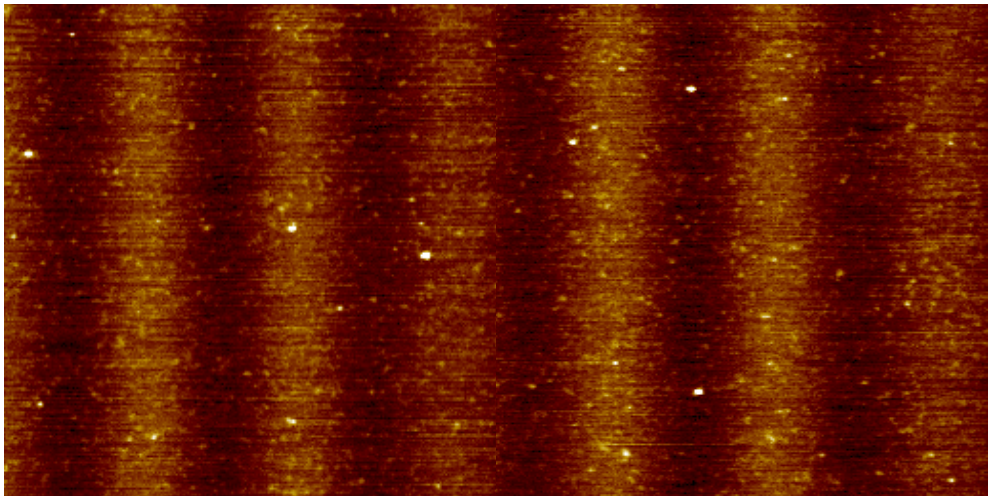
Raw data used for the previous R script

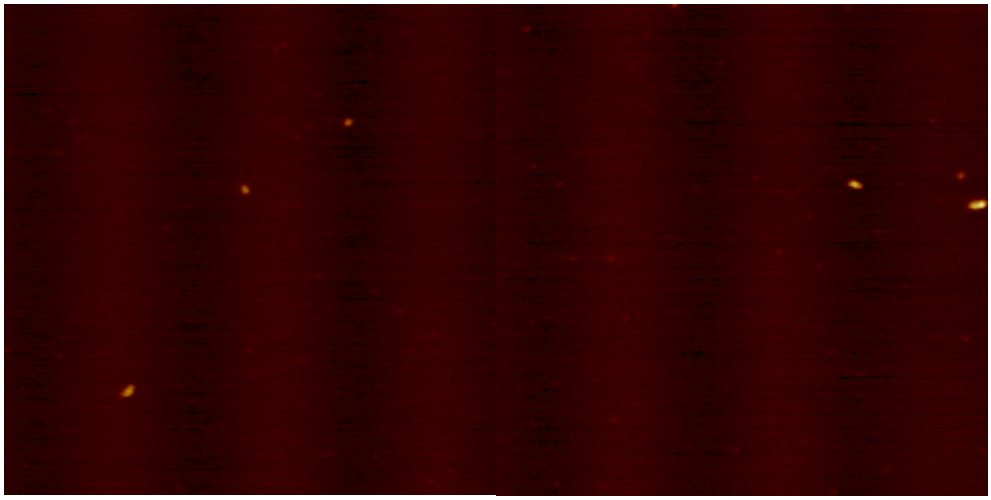
Volume (nm^3)

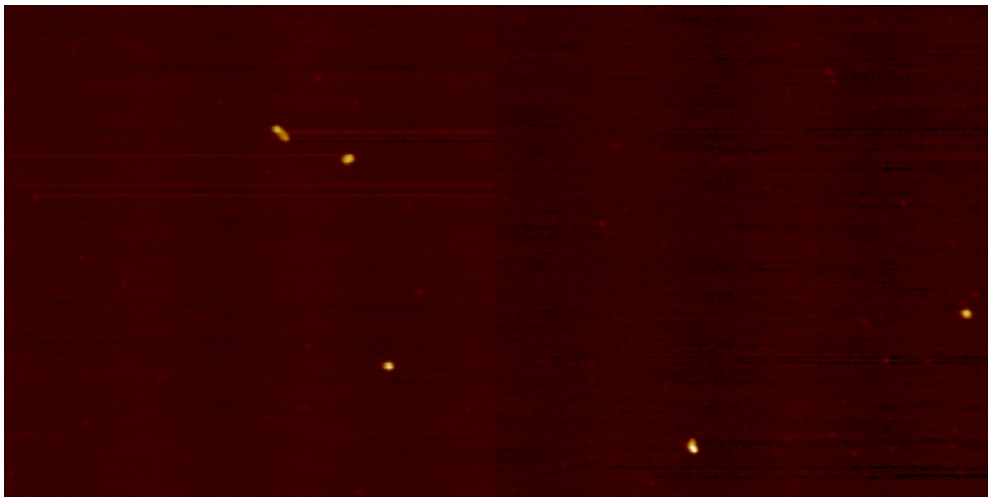
642.915	342.213947	243.765527	722.985	292.121527	229.092307	1440.01656	295.27932
402.021527	925.145527	402.021527	729.206387	344.26332	286.655833	1124.21839	213.781667
859.051667	1457.11595	643.798387	320.056013	320.056013	208.155833	1124.21839	336.748253
1309.37058	722.985	771.443573	419.602387	292.121527	453.53532	1236.31639	231.244253
315.442307	642.915	917.133293	1257.19529	313.215	717.29532	299.382253	200.472253
1705.90077	1260.18667	916.036387	344.873527	373.122013	699.682013	1198.17167	385.07076
1323.64659	642.915	419.602387	720.255293	1650.95129	642.915	1078.85167	415.608307
1455.65167	384.917947	1190.67858	1137.87529	1137.87529	295.27932	344.873527	315.442307
384.917947	250.06332	295.27932	902.45484	772.43058	807.765	363.798307	556.548253
385.07076	250.06332	1077.63858	394.643573	1309.37058	957.331573	402.021527	342.213947
377.715833	295.27932	394.643573	1351.50833	758.571667	1190.67858	295.27932	1273.84357
1458.82359	1242.23215	1024.52129	664.371667	1318.95595	188.661667	1257.19529	1190.67858
674.573527	664.371667	1190.67858	344.873527	1258.53084	271.072013	1163.31557	229.173947
1415.355	320.056013	1187.70395	1198.17167	257.90076	208.155833	270.540307	1236.31639
362.10951	1124.21839	369.735	1458.82359	208.155833	1077.63858	1236.31639	384.917947
344.26332	427.955833	1077.63858	430.965	245.835833	957.331573	1323.77167	1656.39658
430.133947	339.85476	720.255293	402.021527	1190.67858	1353.75239	344.26332	295.27932
645.70332	1190.67858	1318.95595	513.82332	1257.19529	286.655833	295.27932	427.955833
385.07076	250.06332	394.643573	1406.26261	630.765293	270.540307	1190.67858	599.841527
430.270013	1309.37058	315.442307	369.735	1198.17167	256.57254	295.27932	535.31276
1318.95595	373.122013	286.655833	292.121527	1078.85167	257.90076	229.092307	377.715833
315.442307	402.021527	1137.87529	397.01532	271.072013	1382.48129	1077.63858	686.035573
605.72484	1124.21839	270.540307	313.215	221.16276	330.615833	1190.67858	
1482.80011	1236.31639	1017.45839	1014.55284	229.092307	957.331573	122.721667	
642.915	1190.67858	643.798387	1078.85167	415.608307	286.655833	1353.75239	
1388.92667	397.01532	292.121527	1078.85167	1198.17167	196.517947	1190.67858	
315.442307	1561.37861	643.798387	1406.26261	1014.55284	336.748253	257.90076	
1370.83608	248.110763	402.021527	1563.71058	394.643573	369.735	297.46476	
427.955833	1187.70395	686.035573	1257.19529	270.540307	674.573527	302.021947	
1370.83608	777.240013	664.371667	1198.17167	645.70332	1077.63858	171.898253	
344.26332	315.442307	430.965	1198.17167	415.608307	1257.19529	173.615833	
1242.23215	1673.95284	373.122013	1187.70395	286.655833	427.955833	1433.71458	
481.335833	1017.46153	686.035573	837.225527	295.27932	1163.31557	315.442307	
1163.31557	286.655833	674.573527	319.30032	1137.87529	271.072013	339.85476	
1458.82359	1057.81157	1433.71458	1057.81157	221.16276	208.155833	1455.65167	
470.872307	292.121527	320.056013	369.735	264.214253	1236.31639	245.835833	
231.244253	674.573527	402.021527	1133.24484	250.06332	917.133293	858.880013	
1137.87529	1302.315	1296.83256	1137.87529	772.43058	1257.19529	295.27932	
1302.315	957.331573	295.27932	1198.17167	415.608307	270.540307	342.213947	
373.122013	642.915	295.27932	1602.18395	1078.85167	1077.63858	257.90076	

Appendix 6

More images like Figure 2.3.1iii

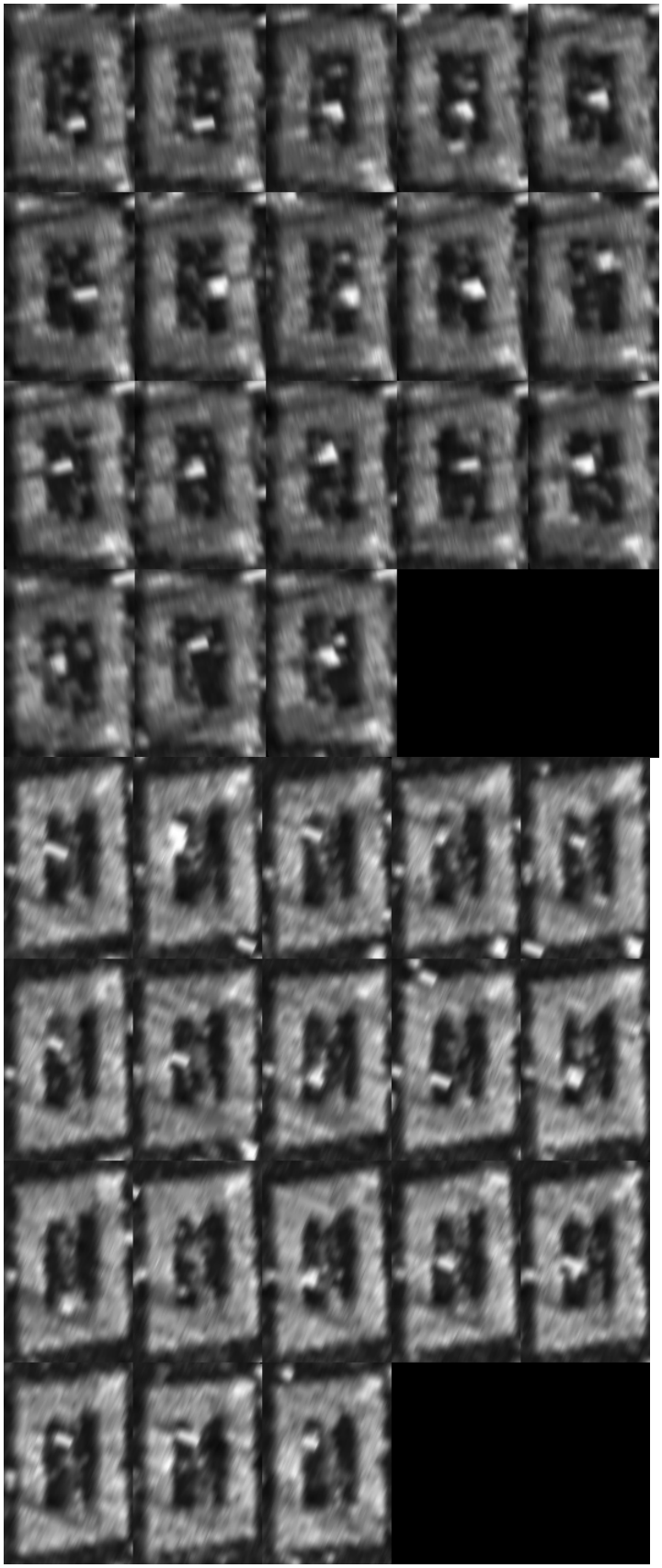


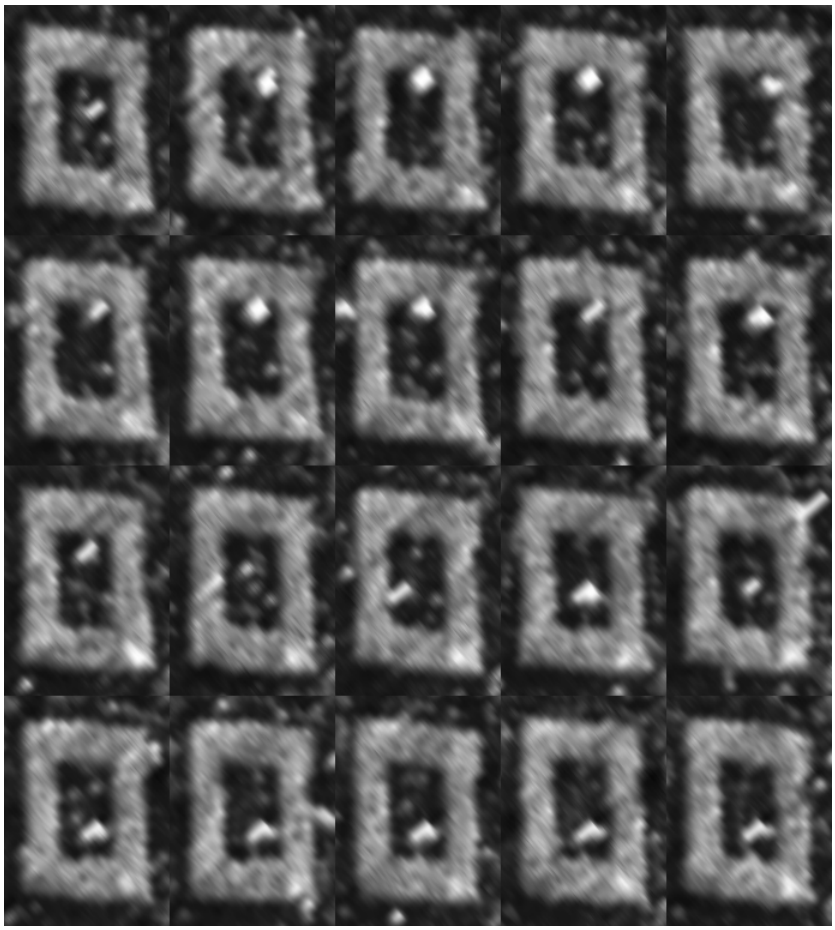


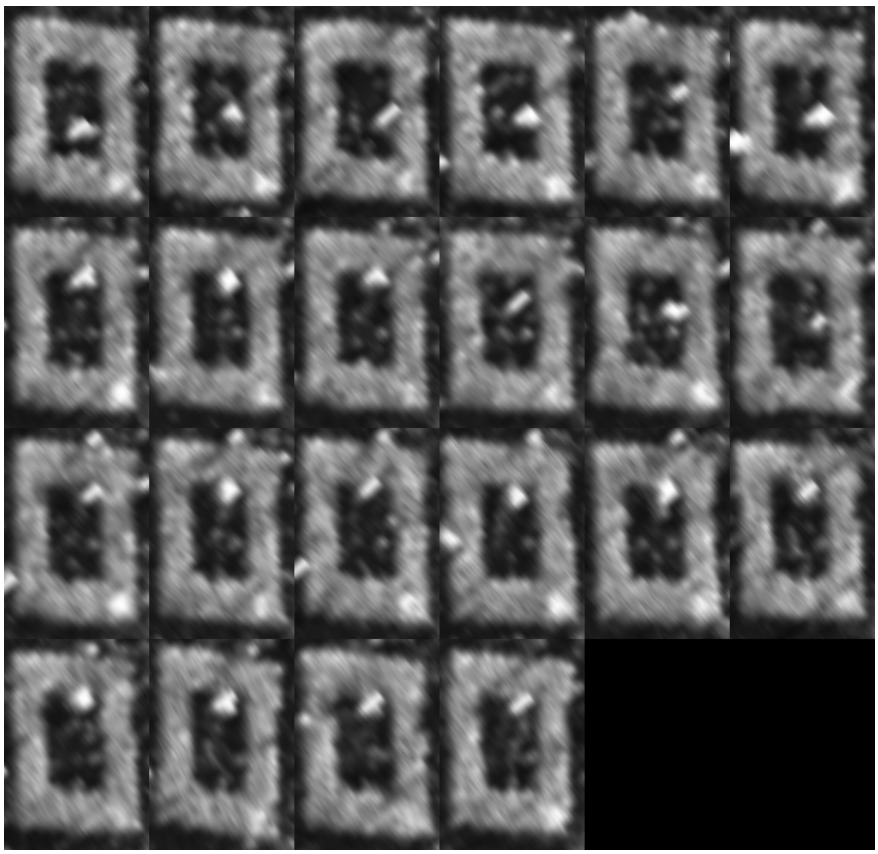


Appendix 7

More images like Figure 2.3.1iv and Figure 2.3.2

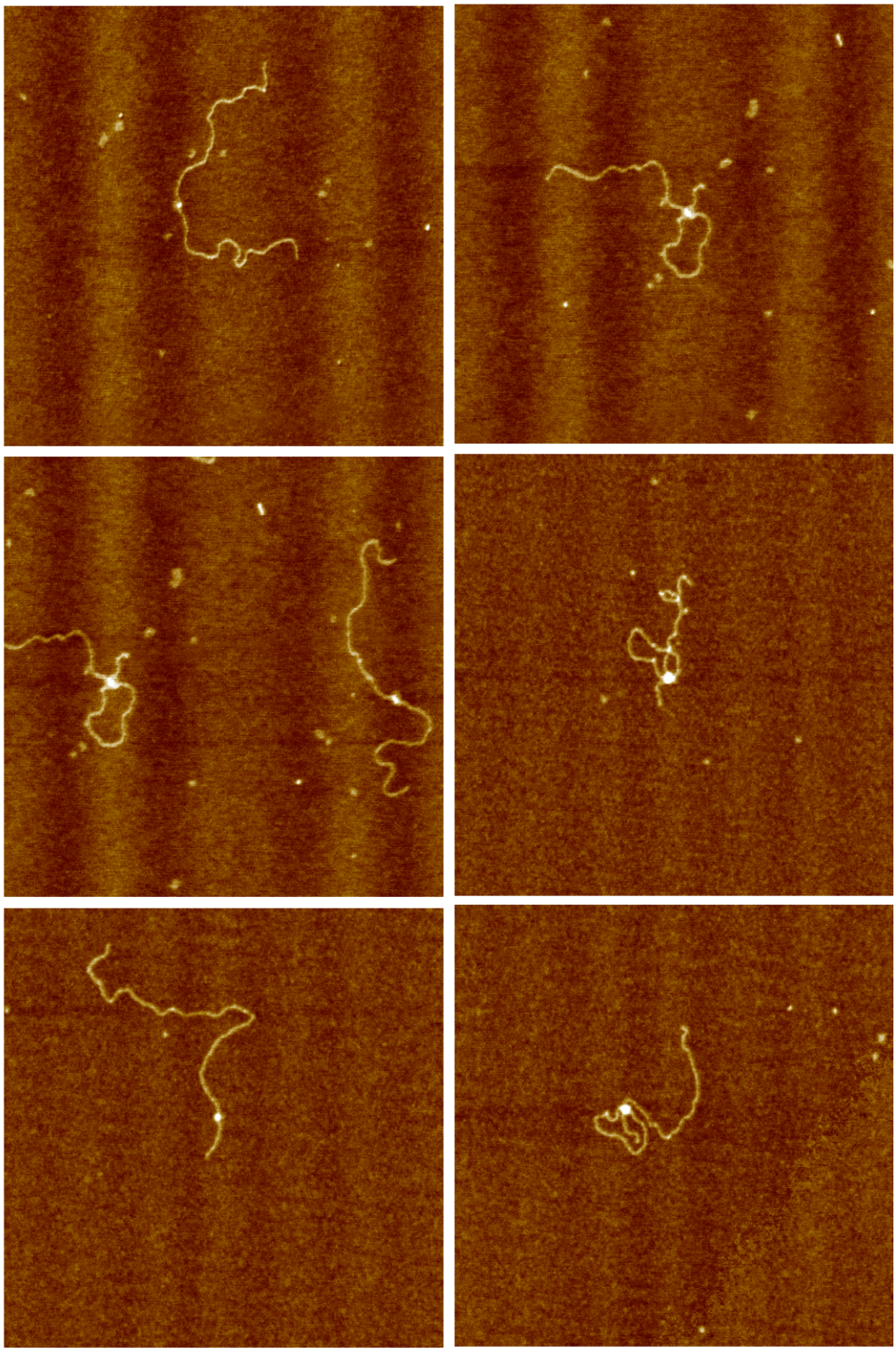


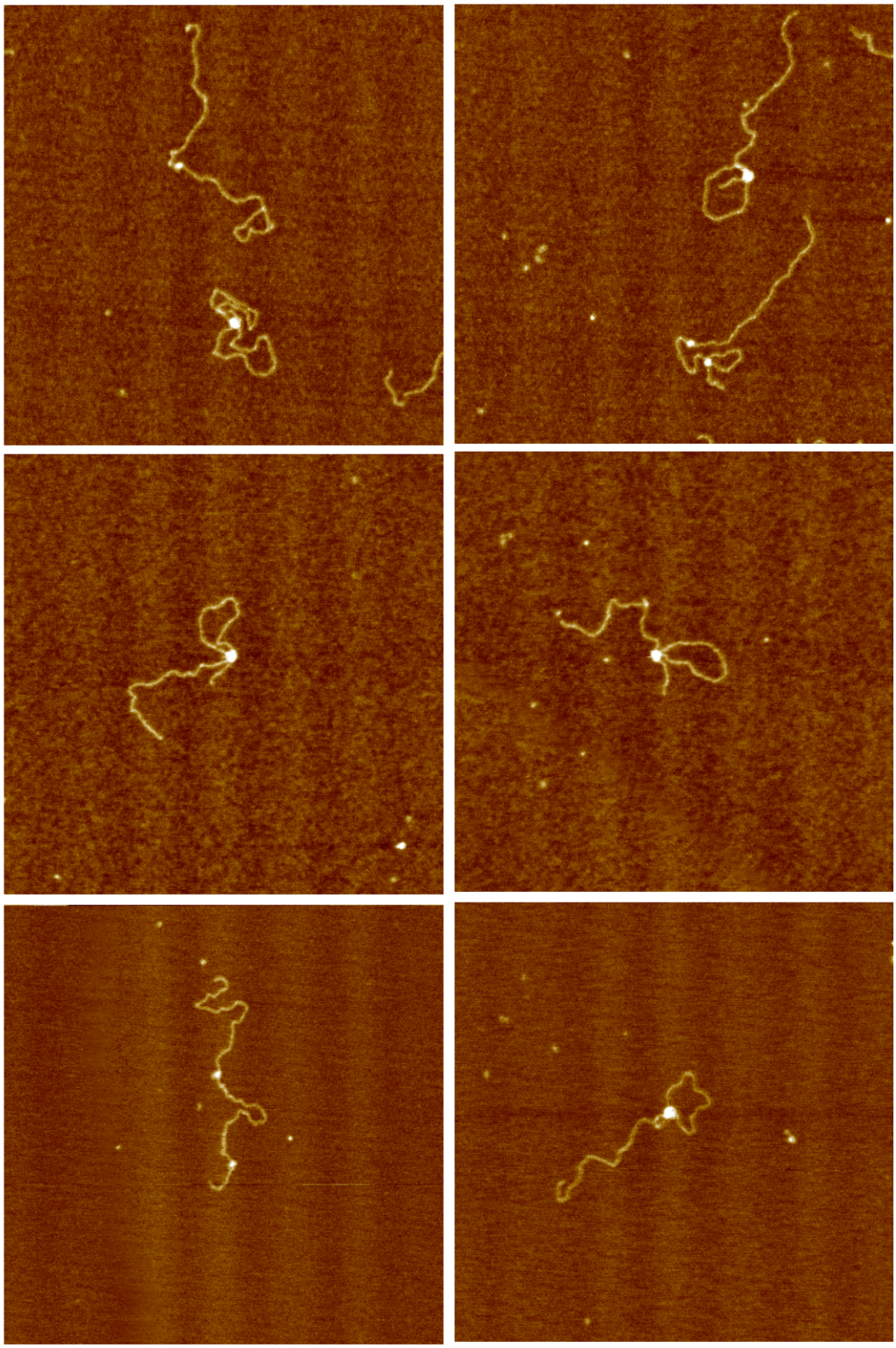




Appendix 8

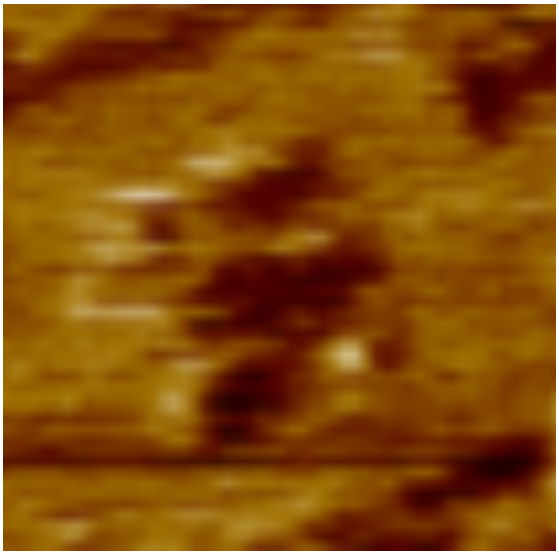
More images like Figure 3.3.1ii

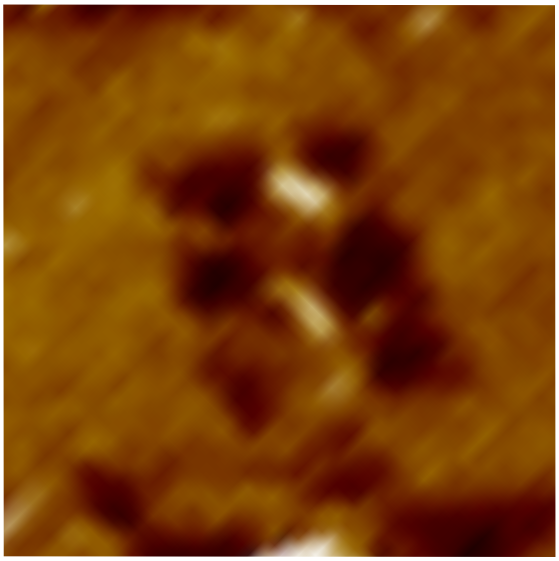


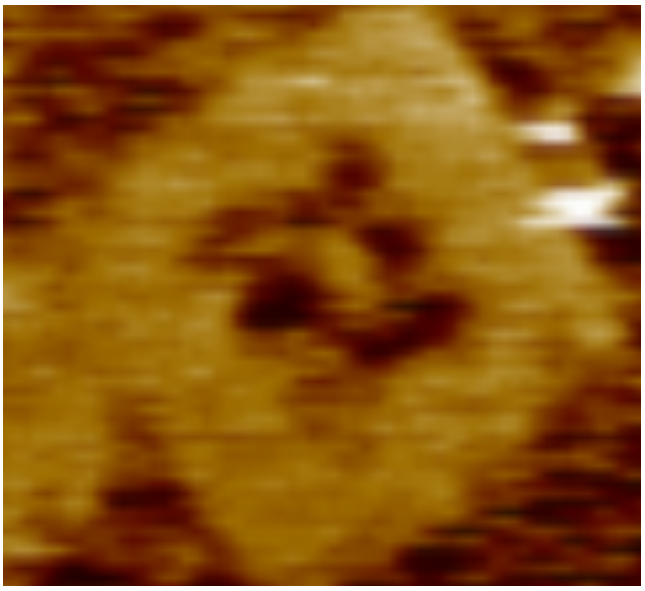


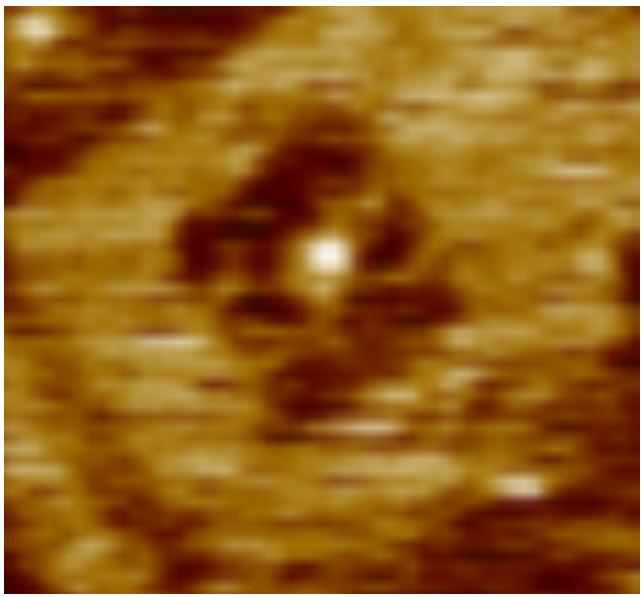
Appendix 9

More images like Figure 3.3.3ii









Bibliography

Abe, M., Yoshizawa, M., Sugita, N., Tanaka, A., Chiba, S., Yambe, T. and Nitta, S. (2007) 'Estimation of effects of visually-induced motion sickness using independent component analysis', *Conf Proc IEEE Eng Med Biol Soc*, 2007, pp. 5501-4.

Aggarwal, A. K. and Wah, D. A. (1998) 'Novel site-specific DNA endonucleases', *Curr Opin Struct Biol*, 8(1), pp. 19-25.

Allen, D. J., Makhov, A., Grilley, M., Taylor, J. and Thresher, R. (1997) 'MutS mediates heteroduplex loop formation by a translocation mechanism', *EMBO J*, 16, pp. 4467–76.

Alves, J., Rüter, T., Geiger, R., Fliess, A., Maass, G. and Pingoud, A. (1989) 'Changing the hydrogen-bonding potential in the DNA binding site of EcoRI by site-directed mutagenesis drastically reduces the enzymatic activity, not, however, the preference of this restriction endonuclease for cleavage within the site-GAATTC-', *Biochemistry*, 28(6), pp. 2678-84.

Andersen, E. S., Dong, M., Nielsen, M. M., Jahn, K., Lind-Thomsen, A., Mamdouh, W., Gothelf, K. V., Besenbacher, F. and Kjems, J. (2008) 'DNA origami design of dolphin-shaped structures with flexible tails', *ACS Nano*, 2(6), pp. 1213-8.

Andersen, E. S., Dong, M., Nielsen, M. M., Jahn, K., Subramani, R., Mamdouh, W., Golas, M. M., Sander, B., Stark, H., Oliveira, C. L., Pedersen, J. S., Birkedal, V., Besenbacher, F., Gothelf, K. V. and Kjems, J. (2009) 'Self-assembly of a nanoscale DNA box with a controllable lid', *Nature*, 459(7243), pp. 73-6.

Anderson, E. S. and Felix, A. (1952) 'Variation in Vi-phage II of Salmonella typhi', *Nature,* 170(4325), pp. 492-4.

Anderson, J. E. (1993). Restriction endonucleases and modification methylases. *Current Opinion in Structural Biology*, 3(1), 24-30.

Ando, T. (2012) ‘High-speed atomic force microscopy coming of age', *Nanotechnology*, 23(6), pp. 062001.

Ando, T. (2017) 'Directly watching biomolecules in action by high-speed atomic force microscopy', *Biophys Rev*, 9(4), pp. 421-429.

Ando, T., Kodera, N., Takai, E., Maruyama, D., Saito, K. and Toda, A. (2001) 'A high-speed atomic force microscope for studying biological macromolecules', *Proc Natl Acad Sci U S A*, 98(22), pp. 12468-72.

Ando, T., Uchihashi, T., Kodera, N., Yamamoto, D., Miyagi, A., Taniguchi, M. and Yamashita, H. (2008) 'High-speed AFM and nano-visualization of biomolecular processes', *Pflugers Arch*, 456(1), pp. 211-25.

Arber, W. and Linn, S. (1969) 'DNA modification and restriction', *Annu Rev Biochem*, 38, pp. 467-500.

Ashby, P. D. and Lieber, C. M. (2004) 'Brownian force profile reconstruction of interfacial 1-nonanol solvent structure', *J Am Chem Soc*, 126(51), pp. 16973-80.

Auerbach, C., Robson, J. M. and Carr, J. G. (1947) 'The Chemical Production of Mutations', *Science,* 105(2723), pp. 243-7.

Avery, O. T., Macleod, C. M. and McCarty, M. (1944) ‘Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a deoxyribonucleic acid fraction isolated from pneumococcus Type III’, *J Exp Med*, 79(2), pp. 137-58.

Benkovic, S. J., Valentine, A. M., Salinas, F. (2001) ‘Replisome-mediated DNA replication. Annu’, *Rev. Biochem*, 70, pp. 181–208.

Bennett, S. E., Sanderson, R. J. and Mosbaugh, D. W. (1995) 'Processivity of Escherichia coli and rat liver mitochondrial uracil-DNA glycosylase is affected by NaCl concentration', *Biochemistry*, 34(18), pp. 6109-19.

Berg, J. M., Tymoczko, J. L., Gatto, Jr, G. J. and Stryer, L. (2015). Biochemistry. 8th Edition. New York: W. H. Freeman and Company.

Berg, O. G. and Blomberg, C. (1976) 'Association kinetics with coupled diffusional flows. Special application to the lac repressor--operator system', *Biophys Chem*, 4(4), pp. 367-81.

Bertani, G. and Weigle, J. J. (1953) 'Host controlled variation in bacterial viruses', *J Bacteriol*, 65(2), pp. 113-21.

Beumer, K. J., Trautman, J. K., Bozas, A., Liu, J. L., Rutter, J., Gall, J. G. and Carroll, D. (2008) 'Efficient gene targeting in Drosophila by direct embryo injection with zinc-finger nucleases', *Proc Natl Acad Sci U S A*, 105(50), pp. 19821-6.

Bickle, T. A. and Krüger, D. H. (1993) 'Biology of DNA restriction', *Microbiol Rev*, 57(2), pp. 434-50.

Biebricher, A., Wende, W., Escudé, C., Pingoud, A. and Desbiolles, P. (2009) 'Tracking of single quantum dot labeled EcoRV sliding along DNA manipulated by double optical tweezers', *Biophys J,* 96(8), pp. L50-2.

Bilcock, D. T. and Halford, S. E. (1999) 'DNA restriction dependent on two recognition sites: activities of the SfiI restriction-modification system in Escherichia coli', *Mol Microbiol*, 31(4), pp. 1243-54.

Binnig, G., Quate, C. F. and Gerber, C. (1986) 'Atomic force microscope', *Phys Rev Lett*, 56(9), pp. 930-933.

Blainey, P. C., van Oijen, A. M., Banerjee, A., Verdine, G. L. and Xie, X. S. (2006) 'A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA', *Proc Natl Acad Sci U S A,* 103(15), pp. 5752-7.

Bonnet, I., Biebricher, A., Porté, P. L., Loverdo, C., Bénichou, O., Voituriez, R., Escudé, C., Wende, W., Pingoud, A. and Desbiolles, P. (2008) 'Sliding and jumping of single EcoRV restriction enzymes on non-cognate DNA', *Nucleic Acids Res*, 36(12), pp. 4118-27.

Bourniquel, A. A. and Bickle, T. A. (2002) 'Complex restriction enzymes: NTP-driven molecular motors', *Biochimie*, 84, pp. 1047–59.

Bozas, A., Beumer, K. J., Trautman, J. K. and Carroll, D. (2009) 'Genetic analysis of zinc-finger nuclease-induced gene targeting in Drosophila', *Genetics*, 182(3), pp. 641-51.

Bukhari, A. I. and Froshauer, S. (1978) 'Insertion of a transposon for chloramphenicol resistance into bacteriophage Mu', *Gene,* 3(4), pp. 303-14.

Bustamante, C., Smith, S. B., Liphardt, J. and Smith, D. (2000) 'Single-molecule studies of DNA mechanics', *Curr Opin Struct Biol,* 10(3), pp. 279-85.

Carroll, D. (2017) 'Genome Editing: Past, Present, and Future', *Yale J Biol Med*, 90(4), pp. 653-659.

Chu, V. T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K. and Kühn, R. (2015) 'Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells', *Nat Biotechnol*, 33(5), pp. 543-8.

Cleveland, J. P., Schäffer, T. E. and Hansma, P. K. (1995) 'Probing oscillatory hydration potentials using thermal-mechanical noise in an atomic-force microscope', *Phys Rev B Condens Matter*, 52(12), pp. R8692-R8695.

Cohen, S. N., Chang, A. C., Boyer, H. W. and Helling, R. B. (1973) 'Construction of biologically functional bacterial plasmids in vitro', *Proc Natl Acad Sci U S A,* 70(11), pp. 3240-4.

Cong, Y., Zarlenga, D. S., Richt, J. A., Wang, X., Wang, Y., Suo, S., Wang, J., Ren, Y. and Ren, X. (2013) 'Evolution and homologous recombination of the hemagglutinin-esterase gene sequences from porcine torovirus', *Virus Genes*, 47(1), pp. 66-74.

Craig NL, Craigie R, Gellert M, Lambowitz AM, eds. 2002. Mobile DNA II. Washington, DC: ASM. 1204 pp.

Crampton, N., Roes, S., Dryden, D. T., Rao, D. N., Edwardson, J. M. and Henderson, R. M. (2007) 'DNA looping and translocation provide an optimal cleavage mechanism for the type III restriction enzymes', *EMBO J,* 26(16), pp. 3815-25.

Crick, F. (1970) 'Central dogma of molecular biology', *Nature,* 227(5258), pp. 561-3.

Coppey, M., Bénichou, O., Voituriez, R. and Moreau, M. (2004) 'Kinetics of target site localization of a protein on DNA: a stochastic approach', *Biophys J*, 87(3), pp. 1640-9.

Cyranoski, D. (2016) 'Chinese scientists to pioneer first human CRISPR trial', *Nature*, 535(7613), pp. 476-7.

Czajkowsky, D. M. and Shao, Z. (2003) 'Inhibition of protein adsorption to muscovite mica by monovalent cations', *J Microsc*, 211(Pt 1), pp. 1-7.

Dahm, R. (2005) 'Friedrich Miescher and the discovery of DNA', *Dev Biol*, 278(2), pp. 274-88.

Darwin, C. (1859). On the Origin of Species. London: John Murray.

Deveau, H., Barrangou, R., Garneau, J. E., Labonté, J., Fremaux, C., Boyaval, P., Romero, D. A., Horvath, P. and Moineau, S. (2008) 'Phage response to CRISPR-encoded resistance in Streptococcus thermophilus', *J Bacteriol*, 190(4), pp. 1390-400.

Dietz, H., Douglas, S. M. and Shih, W. M. (2009) 'Folding DNA into twisted and curved nanoscale shapes', *Science*, 325(5941), pp. 725-30.

Doudna, J. A. and Charpentier, E. (2014) 'Genome editing. The new frontier of genome engineering with CRISPR-Cas9', *Science*, 346(6213), pp. 1258096.

Douglas, S. M., Dietz, H., Liedl, T., Högberg, B., Graf, F. and Shih, W. M. (2009) 'Self-assembly of DNA into nanoscale three-dimensional shapes', *Nature*, 459(7245), pp. 414-8.

Dubochet, J. (2012) 'Cryo-EM--the first thirty years', *J Microsc*, 245(3), pp. 221-4.

Dybvig, K., Sitaraman, R. and French, C. T. (1998) 'A family of phase-variable restriction enzymes with differing specificities generated by high-frequency gene rearrangements', *Proc Natl Acad Sci U S A*, 95(23), pp. 13923-8.

Eaton, P and West, P. (2010). Atomic Force Microscopy. Oxford: Oxford University Press

Eghiaian, F., Rico, F., Colom, A., Casuso, I. and Scheuring, S. (2014) 'High-speed atomic force microscopy: imaging and force spectroscopy', *FEBS Lett*, 588(19), pp. 3631-8.

El Kirat, K., Burton, I., Dupres, V. and Dufrene, Y. F. (2005) 'Sample preparation procedures for biological atomic force microscopy', *J Microsc*, 218(Pt 3), pp. 199-207.

Elf, J., Li, G. W. and Xie, X. S. (2007) 'Probing transcription factor dynamics at the single-molecule level in a living cell', *Science*, 316(5828), pp. 1191-4.

Embleton, M. L., Vologodskii, A. V. and Halford, S. E. (2004) ‘Dynamics of DNA loop capture by the SfiI restriction endonuclease on supercoiled and relaxed DNA’, *J Mol Biol*, 339, pp. 53–66.

Embleton, M. L., Williams, S. A., Watson, M. A. and Halford, S. E. (1999) 'Specificity from the synapsis of DNA elements by the Sfi I endonuclease', *J Mol Biol*, 289(4), pp. 785-97.

Endo, M., Katsuda, Y., Hidaka, K. and Sugiyama, H. (2010a) 'A versatile DNA nanochip for direct analysis of DNA base-excision repair', *Angew Chem Int Ed Engl,* 49(49), pp. 9412-6.

Endo, M., Katsuda, Y., Hidaka, K. and Sugiyama, H. (2010b) 'Regulation of DNA methylation using different tensions of double strands constructed in a defined DNA nanostructure'*, J Am Chem Soc,* 132(5), pp. 1592-7.

Endo, M. and Sugiyama, H. (2014) 'Single-molecule imaging of dynamic motions of biomolecules in DNA origami nanostructures using high-speed atomic force microscopy', *Acc Chem Res,* 47(6), pp. 1645-53.

Endo, M., Tatsumi, K., Terushima, K., Katsuda, Y., Hidaka, K., Harada, Y. and Sugiyama, H. (2012) 'Direct visualization of the movement of a single T7 RNA polymerase and transcription on a DNA nanostructure', *Angew Chem Int Ed Engl*, 51(35), pp. 8778-82.

Enghiad, B. and Zhao, H. (2017) 'Programmable DNA-Guided Artificial Restriction Enzymes', *ACS Synth Biol*, 6(5), pp. 752-757.

Fan, L., Potter, D. and Sulchek, T. (2012) 'Constant tip-surface distance with atomic force microscopy via quality factor feedback', *Rev Sci Instrum*, 83(2), pp. 023706.

Fett, T and Thun, G. (1998) ‘Determination of room-temperature tensile creep of PZT’, *Journal of Materials Science Letters,* 17, pp. 1929-1931

Flores, H., Osuna, J., Heitman, J. and Soberón, X. (1995) 'Saturation mutagenesis of His114 of EcoRI reveals relaxed-specificity mutants', *Gene*, 157(1-2), pp. 295-301.

Foster, T. J. (1977) 'Insertion of the tetracycline resistance translocation unit Tn10 in the lac operon of Escherichia coli K12', *Mol Gen Genet*, 154(3), pp. 305-9.

Frank, J. (2017) 'Time-resolved cryo-electron microscopy: Recent progress', *J Struct Biol*, 200(3), pp. 303-306.

Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., Fleischmann, R. D., Bult, C. J., Kerlavage, A. R., Sutton, G., Kelley, J. M., Fritchman, R. D., Weidman, J. F., Small, K. V., Sandusky, M., Fuhrmann, J., Nguyen, D., Utterback, T. R., Saudek, D. M., Phillips, C. A., Merrick, J. M., Tomb, J. F., Dougherty, B. A., Bott, K. F., Hu, P. C., Lucier, T. S., Peterson, S. N., Smith, H. O., Hutchison, C. A. and Venter, J. C. (1995) 'The minimal gene complement of Mycoplasma genitalium', *Science*, 270(5235), pp. 397-403.

Friedman, A. M., Fischmann, T. O. and Steitz, T. A. (1995) 'Crystal structure of lac repressor core tetramer and its implications for DNA looping', *Science*, 268(5218), pp. 1721-7.

Fu, Y., Foden, J. A., Khayter, C., Maeder, M. L., Reyon, D., Joung, J. K. and Sander, J. D. (2013) 'High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells', *Nat Biotechnol*, 31(9), pp. 822-6.

Fung, Y.C. (1993). Biomechanics: Mechanical Properties of Living Tissues. New York: Springer.

Gaj, T., Gersbach, C. A. and Barbas, C. F. (2013) 'ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering', *Trends Biotechnol*, 31(7), pp. 397-405.

Gantz, V. M., Jasinskiene, N., Tatarenkova, O., Fazekas, A., Macias, V. M., Bier, E. and James, A. A. (2015) 'Highly efficient Cas9-mediated gene drive for population

modification of the malaria vector mosquito Anopheles stephensi', *Proc Natl Acad Sci U S A*, 112(49), pp. E6736-43.

Gemmen, G. J., Millin, R. and Smith, D. E. (2006) 'DNA looping by two-site restriction endonucleases: heterogeneous probability distributions for loop size and unbinding force', *Nucleic Acids Res,* 34(10), pp. 2864-77.

Gill, P., Jeffreys, A. J. and Werrett, D. J. (1985) 'Forensic application of DNA 'fingerprints'', *Nature*, 318(6046), pp. 577-9.

Gowers, D. M. and Halford, S. E. (2003) 'Protein motion from non-specific to specific DNA by three-dimensional routes aided by supercoiling', *EMBO J*, 22(6), pp. 1410-8.

Gowers, D. M., Wilson, G. G. and Halford, S. E. (2005) 'Measurement of the contributions of 1D and 3D pathways to the translocation of a protein along DNA', *Proc Natl Acad Sci U S A*, 102(44), pp. 15883-8.

Guillaume-Gentil, O., Potthoff, E., Ossola, D., Franz, C. M., Zambelli, T. and Vorholt, J. A. (2014) 'Force-controlled manipulation of single cells: from AFM to FluidFM', *Trends Biotechnol*, 32(7), pp. 381-8.

Gusella, J. F., Wexler, N. S., Conneally, P. M., Naylor, S. L., Anderson, M. A., Tanzi, R. E., Watkins, P. C., Ottina, K., Wallace, M. R. and Sakaguchi, A. Y. (1983) 'A polymorphic DNA marker genetically linked to Huntington's disease', *Nature*, 306(5940), pp. 234-8.

Halford, S. E. (2001) 'Hopping, jumping and looping by restriction enzymes', *Biochem Soc Trans*, 29(4), pp. 363-74.

Halford, S. E. (2009) 'An end to 40 years of mistakes in DNA-protein association kinetics?', *Biochem Soc Trans*, 37(Pt 2), pp. 343-8.

Halford, S. E., Gowers, D. M. and Sessions, R. B. (2000) 'Two are better than one', *Nat Struct Biol*, 7(9), pp. 705-7.

Halford, S. E. and Marko, J. F. (2004) 'How do site-specific DNA-binding proteins find their targets?', *Nucleic Acids Res*, 32(10), pp. 3040-52.

Halford, S. E. and Szczelkun, M. D. (2002) 'How to get from A to B: strategies for analysing protein motion on DNA', *Eur Biophys J*, 31(4), pp. 257-67.

Halford, S. E., Welsh, A. J. and Szczelkun, M. D. (2004) 'Enzyme-mediated DNA looping', *Annu Rev Biophys Biomol Struct*, 33, pp. 1-24.

Hammar, P., Leroy, P., Mahmutovic, A., Marklund, E. G., Berg, O. G. and Elf, J. (2012) 'The lac repressor displays facilitated diffusion in living cells', *Science*, 336(6088), pp. 1595-8.

Hammond, A., Galizi, R., Kyrou, K., Simoni, A., Siniscalchi, C., Katsanos, D., Gribble, M., Baker, D., Marois, E., Russell, S., Burt, A., Windbichler, N., Crisanti, A. and Nolan, T. (2016) 'A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae', *Nat Biotechnol*, 34(1), pp. 78-83.

Han, D., Pal, S., Nangreave, J., Deng, Z., Liu, Y. and Yan, H. (2011) 'DNA origami with complex curvatures in three-dimensional space', *Science*, 332(6027), pp. 342-6.

Hansma, H. G., Browne, K. A., Bezanilla, M. and Bruice, T. C. (1994) 'Bending and straightening of DNA induced by the same ligand: characterization with the atomic force microscope', *Biochemistry*, 33(28), pp. 8436-41.

Hansma, H. G. and Hoh, J. H. (1994) 'Biomolecular imaging with the atomic force microscope', *Annu Rev Biophys Biomol Struct*, 23, pp. 115-39.

Hansma, H. G. and Laney, D. E. (1996) 'DNA binding to mica correlates with cationic radius: assay by atomic force microscopy', *Biophys J,* 70(4), pp. 1933-9.

Hards, A., Zhou, C., Seitz, M., Bräuchle, C. and Zumbusch, A. (2005) 'Simultaneous AFM manipulation and fluorescence imaging of single DNA strands', *Chemphyschem*, 6(3), pp. 534-40.

Heitman, J. and Model, P. (1990) 'Mutants of the EcoRI endonuclease with promiscuous substrate specificity implicate residues involved in substrate recognition', *EMBO J*, 9(10), pp. 3369-78.

Henderson, R., Baldwin, J. M., Ceska, T. A., Zemlin, F., Beckmann, E. and Downing, K. H. (1990) 'Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy', *J Mol Biol*, 213(4), pp. 899-929.

Hershfield, V., Boyer, H. W., Yanofsky, C., Lovett, M. A. and Helinski, D. R. (1974) 'Plasmid ColEl as a molecular vehicle for cloning and amplification of DNA', *Proc Natl Acad Sci U S A*, 71(9), pp. 3455-9.

Hickman, A. B., Li, Y., Mathew, S. V., May, E. W., Craig, N. L. and Dyda, F. (2000) 'Unexpected structural diversity in DNA recombination: the restriction endonuclease connection', *Mol Cell*, 5(6), pp. 1025-34.

Horton, N. C. and Perona, J. J. (1998) 'Recognition of flanking DNA sequences by EcoRV endonuclease involves alternative patterns of water-mediated contacts', J Biol Chem, 273(34), pp. 21721-9.

Housman, D. and Gusella, J. F. (1983) 'Application of recombinant DNA techniques to neurogenetic disorders', *Res Publ Assoc Res Nerv Ment Dis*, 60, pp. 167-72.

Houtsmuller, A. B., Rademakers, S., Nigg, A. L., Hoogstraten, D., Hoeijmakers, J. H. and Vermeulen, W. (1999) 'Action of DNA repair endonuclease ERCC1/XPF in living cells', *Science*, 284(5416), pp. 958-61.

Hu, T., Grosberg, A. Y. and Shklovskii, B. I. (2006) 'How proteins search for their specific sites on DNA: the role of DNA conformation', *Biophys J,* 90(8), pp. 2731-44.

Huang, J., Schlick, T. and Vologodskii, A. (2001) 'Dynamics of site juxtaposition in supercoiled DNA', *Proc Natl Acad Sci U S A*, 98(3), pp. 968-73.

Ikai, A., Afrin, R., Saito, M. and Watanabe-Nakayama, T. (2018) 'Atomic force microscope as a nano- and micrometer scale biological manipulator: A short review', *Semin Cell Dev Biol,* 73, pp. 132-144.

Jackson, D. A., Symons, R. H. and Berg, P. (1972) 'Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of Escherichia coli', *Proc Natl Acad Sci U S A,* 69(10), pp. 2904-9.

Jeltsch, A., Alves, J., Wolfes, H., Maass, G. and Pingoud, A. (1994) 'Pausing of the restriction endonuclease EcoRI during linear diffusion on DNA', *Biochemistry*, 33(34), pp. 10215-9.

Jeltsch, A. and Pingoud, A. M. (2001) 'Methods for determining activity and specificity of DNA binding and DNA cleavage by class II restriction endonucleases', *Methods Mol Biol*, 160, pp. 287-308.

Jian, H., Schlick, T. and Vologodskii, A. (1998) 'Internal motion of supercoiled DNA: brownian dynamics simulations of site juxtaposition', *J Mol Biol*, 284(2), pp. 287-96.

Jiang, Q., Song, C., Nangreave, J., Liu, X., Lin, L., Qiu, D., Wang, Z. G., Zou, G., Liang, X., Yan, H. and Ding, B. (2012) 'DNA origami as a carrier for circumvention of drug resistance', *J Am Chem Soc, 134(32),* pp. 13396-403.

Jo, K. and Topal, M. D. (1995) 'DNA topoisomerase and recombinase activities in Nae I restriction endonuclease', *Science*, 267(5205), pp. 1817-20.

Kaiser, Y., Lepzien, R., Kullberg, S., Eklund, A., Smed-Sörensen, A. and Grunewald, J. (2016) 'Expanded lung T-bet+RORγT+ CD4+ T-cells in sarcoidosis patients with a favourable disease phenotype', *Eur Respir J*, 48(2), pp. 484-94.

Ke, Y., Douglas, S. M., Liu, M., Sharma, J., Cheng, A., Leung, A., Liu, Y., Shih, W. M. and Yan, H. (2009a) 'Multilayer DNA origami packed on a square lattice', *J Am Chem Soc*, 131(43), pp. 15903-8.

Ke, Y., Sharma, J., Liu, M., Jahn, K., Liu, Y. and Yan, H. (2009b) 'Scaffolded DNA origami of a DNA tetrahedron molecular container', *Nano Lett*, 9(6), pp. 2445-7.

Kelly, T. J. and Smith, H. O. (1970) 'A restriction enzyme from Hemophilus influenzae. II', *J Mol Biol,* 51(2), pp. 393-409.

King, G. M., Carter, A. R., Churnside, A. B., Eberle, L. S. and Perkins, T. T. (2009) 'Ultrastable atomic force microscopy: atomic-scale stability and registration in ambient conditions', *Nano Lett,* 9(4), pp. 1451-6.

Klenin, K. V. and Langowski, J. (2001) 'Intrachain reactions of supercoiled DNA simulated by Brownian dynamics', *Biophys J*, 81(4), pp. 1924-9.

Klenin, K. V., Merlitz, H., Langowski, J. and Wu, C. X. (2006) 'Facilitated diffusion of DNA-binding proteins', *Phys Rev Lett,* 96(1), pp. 018104.

Kodera, N., Yamamoto, D., Ishikawa, R. and Ando, T. (2010) 'Video imaging of walking myosin V by high-speed atomic force microscopy', *Nature*, 468(7320), pp. 72-6.

Kohn, D. B., Porteus, M. H. and Scharenberg, A. M. (2016) 'Ethical and regulatory aspects of genome editing', *Blood*, 127(21), pp. 2553-60.

Kolata, G. (1983) 'Huntington's disease gene located', *Science*, 222(4626), pp. 913-5.

Kong, H. (1998) 'Analyzing the functional organization of a novel restriction modification system, the BcgI system', *J Mol Biol*, 279(4), pp. 823-32.

Kong, H., Lin, LF., Porter, N., Stickel, S. and Byrd, D. (2000) ‘Functional analysis of putative restriction-modification system genes in the Helicobacter pylori J99 genome’, *Nucleic Acids Res*, 28, pp. 3216–23.

Kong, H., Roemer, S. E., Waite-Rees, P. A., Benner, J. S., Wilson, G. G. and Nwankwo, D. O. (1994) 'Characterization of BcgI, a new kind of restriction-modification system', *J Biol Chem*, 269(1), pp. 683-90.

Korayem, M. H., Badkoobeh Hezaveh, H. and Taheri, M. (2014) 'Dynamic modeling and simulation of rough cylindrical micro/nanoparticle manipulation with atomic force microscopy', *Microsc Microanal*, 20(6), pp. 1692-707.

Kuzuya, A., Kimura, M., Numajiri, K., Koshi, N., Ohnishi, T., Okada, F. and Komiyama, M. (2009) 'Precisely programmed and robust 2D streptavidin nanoarrays by using periodical nanometre-scale wells embedded in DNA origami assembly', *Chembiochem,* 10(11), pp. 1811-5.

Kuzuya, A. and Komiyama, M. (2009) 'Design and construction of a box-shaped 3D-DNA origami', *Chem Commun (Camb),* (28), pp. 4182-4.

Kuzyk, A., Laitinen, K. T. and Törmä, P. (2009) 'DNA origami as a nanoscale template for protein assembly', *Nanotechnology,* 20(23), pp. 235305.

Last, J. A., Russell, P., Nealey, P. F. and Murphy, C. J. (2010) 'The applications of atomic force microscopy to vision science', *Invest Ophthalmol Vis Sci*, 51(12), pp. 6083-94.

Lever, M. A., Th'ng, J. P., Sun, X. and Hendzel, M. J. (2000) 'Rapid exchange of histone H1.1 on chromatin in living human cells', *Nature*, 408(6814), pp. 873-6.

Lin, C., Liu, Y., Rinker, S. and Yan, H. (2006) 'DNA tile based self-assembly: building complex nanoarchitectures', *Chemphyschem,* 7(8), pp. 1641-7.

Lloyd, G., Landini, P. and Busby, S. (2001) ‘Activation and repression of transcription initiation in bacteria’ *Essays Biochem*, 37, pp. 17–31.

Loenen, W. A., Dryden, D. T., Raleigh, E. A. and Wilson, G. G. (2014a) 'Type I restriction enzymes and their relatives', *Nucleic Acids Res*, 42(1), pp. 20-44.

Loenen, W. A., Dryden, D. T., Raleigh, E. A., Wilson, G. G. and Murray, N. E. (2014b) 'Highlights of the DNA cutters: a short history of the restriction enzymes', Nucleic Acids Res, 42(1), pp. 3-19.

Lomholt, M. A., Ambjörnsson, T. and Metzler, R. (2005) 'Optimal target search on a fast-folding polymer chain with volume exchange', *Phys Rev Lett*, 95(26), pp. 260603.

Long, F., Wang, C., Lü, M., Zhang, F., Sun, J. and Hu, J. (2011) 'Optimizing single DNA molecules manipulation by AFM', J Microsc, 243(2), pp. 118-23.

Lundin, S., Jemt, A., Terje-Hegge, F., Foam, N., Pettersson, E., Käller, M., Wirta, V., Lexow, P. and Lundeberg, J. (2015) 'Endonuclease specificity and sequence dependence of type IIS restriction enzymes', *PLoS One,* 10(1), pp. e0117059.

Luria, S. E. (1953) 'Host-induced modifications of viruses', *Cold Spring Harb Symp Quant Biol,* 18, pp. 237-44.

Luria, S. E. and Human, M. L. (1952) 'A nonhereditary, host-induced variation of bacterial viruses', *J Bacteriol*, 64(4), pp. 557-69.

Luscombe, N. M., Laskowski, R. A. and Thornton, J. M. (2001) 'Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level', *Nucleic Acids Res*, 29(13), pp. 2860-74.

Machnicka, M. A., Kaminska, K. H., Dunin-Horkawicz, S. and Bujnicki, J. M. (2015) 'Phylogenomics and sequence-structure-function relationships in the GmrSD family of Type IV restriction enzymes', *BMC Bioinformatics*, 16, pp. 336.

Mahmood, I. A. and Moheimani, S. O. (2009) 'Making a commercial atomic force microscope more accurate and faster using positive position feedback control', *Rev Sci Instrum*, 80(6), pp. 063705.

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E. and Church, G. M. (2013) 'RNA-guided human genome engineering via Cas9', *Science*, 339(6121), pp. 823-6.

Marshall, J. J., Gowers, D. M. and Halford, S. E. (2007) 'Restriction endonucleases that bridge and excise two recognition sites from DNA', *J Mol Biol,* 367(2), pp. 419-31.

Marshall, J. J. and Halford, S. E. (2010) 'The type IIB restriction endonucleases', *Biochem Soc Trans*, 38(2), pp. 410-6.

Martin, A. M., Sam, M. D., Reich, N. O. and Perona, J. J. (1999) 'Structural and energetic origins of indirect readout in site-specific DNA cleavage by a restriction endonuclease', *Nat Struct Biol*, 6(3), pp. 269-77.

Martin, Y., Williams, C. C. and Wickramasinghe, H. K. (1987) 'Atomic force microscope - force mapping and profilling on a sub 100 - Å scale'. *Journal of Applied Physics*, 61(4723).

Maruyama, T., Dougan, S. K., Truttmann, M. C., Bilate, A. M., Ingram, J. R. and Ploegh, H. L. (2015) 'Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining', *Nat Biotechnol*, 33(5), pp. 538-42.

Matthews, B. W., Ohlendorf, D. H., Anderson, W. F. and Takeda, Y. (1982) 'Structure of the DNA-binding region of lac repressor inferred from its homology with cro repressor', *Proc Natl Acad Sci U S A,* 79(5), pp. 1428-32.

Maune, H. T., Han, S. P., Barish, R. D., Bockrath, M., Goddard, W. A., Rothemund, P. W. and Winfree, E. (2010) 'Self-assembly of carbon nanotubes into two-dimensional geometries using DNA origami templates', *Nat Nanotechnol,* 5(1), pp. 61-6.

Mendel, Gregor Johann. (1866) ‘Versuche über Pflanzen-Hybriden [Experiments Concerning Plant Hybrids]’, *In Verhandlungen des naturforschenden Vereines in Brünn [Proceedings of the Natural History Society of Brünn],* IV, pp. 3–47.

Menger, L., Sledzinska, A., Bergerhoff, K., Vargas, F. A., Smith, J., Poirot, L., Pule, M., Hererro, J., Peggs, K. S. and Quezada, S. A. (2016) 'TALEN-Mediated Inactivation of PD-1 in Tumor-Reactive Lymphocytes Promotes Intratumoral T-cell Persistence and Rejection of Established Tumors', *Cancer Res*, 76(8), pp. 2087-93.

Menozzi, C., Carlo Gazzadi, G., Alessandrini, A. and Facci, P. (2005) 'Focused ion beam-nanomachined probes for improved electric force microscopy', *Ultramicroscopy*, 104(3-4), pp. 220-5.

Misteli, T., Gunjan, A., Hock, R., Bustin, M. and Brown, D. T. (2000) 'Dynamic binding of histone H1 to chromatin in living cells', *Nature*, 408(6814), pp. 877-81.

Mizuuchi, K. and Adzuma, K. (1991) 'Inversion of the phosphate chirality at the target site of Mu DNA strand transfer: evidence for a one-step transesterification mechanism', *Cell*, 66(1), pp. 129-40.

Mizuuchi, K., Nobbs, T. J., Halford, S. E., Adzuma, K. and Qin, J. (1999) 'A new method for determining the stereochemistry of DNA cleavage reactions: application to the SfiI and HpaII restriction endonucleases and to the MuA transposase', *Biochemistry*, 38(14), pp. 4640-8.

Mokaberi, B. and Requicha, A. A. G. (2006) ‘Drift compensation for automatic nanomanipulation with scanning probe microscopes’, *IEEE Transactions on Automation Science and Engineering*, 3(3), pp. 199-207.

Mondal, A. and Bhattacherjee, A. (2017) 'Understanding the Role of DNA Topology in Target Search Dynamics of Proteins', *J Phys Chem B*, 121(40), pp. 9372-9381.

Morris, V. J., Kirby, A. R. and Gunning, A. P. (1999) ‘Using Atomic Force Microscopy to Probe Food Biopolymer Functionality’, *Scanning Vol.* 21, pp. 287–292.

Muller, D. J. (2008) 'AFM: a nanotool in membrane biology', *Biochemistry*, 47(31), pp. 7986-98.

Müller, D. J., Amrein, M. and Engel, A. (1997) 'Adsorption of biological molecules to a solid support for scanning probe microscopy', *J Struct Biol*, 119(2), pp. 172-88.

Muller, H. J. (1927) 'ARTIFICIAL TRANSMUTATION OF THE GENE', *Science*, 66(1699), pp. 84-7.

Nakayama, Y. (2002) 'Scanning probe microscopy installed with nanotube probes and nanotube tweezers', *Ultramicroscopy*, 91(1-4), pp. 49-56.

Nangreave, J., Han, D., Liu, Y. and Yan, H. (2010) 'DNA origami: a history and current perspective', Curr Opin Chem Biol, 14(5), pp. 608-15.

Neaves, K. J., Cooper, L. P., White, J. H., Carnally, S. M., Dryden, D. T., Edwardson, J. M. and Henderson, R. M. (2009) 'Atomic force microscopy of the EcoKI Type I DNA restriction enzyme bound to DNA shows enzyme dimerization and DNA looping', Nucleic Acids Res, 37(6), pp. 2053-63.

Plamont, M. A., Billon-Denis, E., Maurin, S., Gauron, C., Pimenta, F. M., Specht, C. G., Shi, J., Quérard, J., Pan, B., Rossignol, J., Moncoq, K., Morellet, N., Volovitch, M., Lescop, E., Chen, Y., Triller, A., Vriz, S., Le Saux, T., Jullien, L. and Gautier, A. (2016) 'Small fluorescence-activating and absorption-shifting tag for tunable protein imaging in vivo', Proc Natl Acad Sci U S A, 113(3), pp. 497-502.

Neaves, K. J., Cooper, L. P., White, J. H., Carnally, S. M., Dryden, D. T., Edwardson, J. M. and Henderson, R. M. (2009) 'Atomic force microscopy of the EcoKI Type I DNA

Nelluri, S. and Srilakshmi, E. (2015) 'Sample preparation methods for Scanning Probe Microsocpy', *South Asian Journal of Engineeting and Technology*, 1(1), pp. 35-40.

Nguyen, J. and Harbison,A. (2017) 'Scanning Electron Microscopy Sample Preparation and Imaging', *Molecular Profiling*, pp. 71-84.

Nievergelt, A. P., Erickson, B. W., Hosseini, N., Adams, J. D. and Fantner, G. E. (2015) 'Studying biological membranes with extended range high-speed atomic force microscopy', *Sci Rep*, 5, pp. 11987.

Obataya, I., Nakamura, C., Han, S., Nakamura, N. and Miyake, J. (2005a) 'Mechanical sensing of the penetration of various nanoneedles into a living cell using atomic force microscopy', *Biosens Bioelectron*, 20(8), pp. 1652-5.

Obataya, I., Nakamura, C., Han, S., Nakamura, N. and Miyake, J. (2005b) 'Nanoscale operation of a living cell using an atomic force microscope with a nanoneedle', *Nano Lett*, 5(1), pp. 27-30.

Ogata, K., Sato, K. and Tahirov, T. (2003) 'Eukaryotic transcriptional regulatory complexes: cooperativity from near and afar', *Curr. Opin. Struct. Biol*,13, pp. 40–48.

Osuna, J., Flores, H. and Soberón, X. (1991) 'Combinatorial mutagenesis of three major groove-contacting residues of EcoRI: single and double amino acid replacements retaining methyltransferase-sensitive activities', *Gene*, 106(1), pp. 7-12.

Pastré, D., Piétrement, O., Fusil, S., Landousy, F., Jeusset, J., David, M. O., Hamon, L., Le Cam, E. and Zozime, A. (2003) 'Adsorption of DNA to mica mediated by divalent counterions: a theoretical and experimental study', *Biophys J*, 85(4), pp. 2507-18.

Peakman, L. J., Antognozzi, M., Bickle, T. A., Janscak, P. and Szczelkun, M. D. (2003) 'S-Adenosyl methionine prevents promiscuous DNA cleavage by the EcoP1I type III restriction enzyme', *J. Mol. Biol*, 333, pp. 321–35.

Perrault, S. D. and Shih, W. M. (2014) 'Virus-inspired membrane encapsulation of DNA nanostructures to achieve in vivo stability', *ACS Nano*, 8(5), pp. 5132-40.

Pingoud, A. and Jeltsch, A. (2001) 'Structure and function of type II restriction endonucleases', *Nucleic Acids Res*, 29(18), pp. 3705-27.

Piontek, M. C. and Roos, W. H. (2018) 'Atomic Force Microscopy: An Introduction', *Methods Mol Biol,* 1665, pp. 243-258.

Radmacher, M. (2007) 'Studying the mechanics of cellular processes by atomic force microscopy', *Methods Cell Biol*, 83, pp. 347-72.

Raghavendra, N. K., Bheemanaik, S. and Rao, D. N. (2012) 'Mechanistic insights into type III restriction enzymes', *Front Biosci (Landmark Ed),* 17, pp. 1094-107.

Raghavendra, N. K. and Rao, D. N. (2004) 'Unidirectional translocation from recognition site and a necessary interaction with DNA end for cleavage by Type III restriction enzyme', *Nucleic Acids Res*, 32(19), pp. 5703-11.

Rajendran, A., Endo, M., Hidaka, K. and Sugiyama, H. (2013a) 'Direct and real-time observation of rotary movement of a DNA nanomechanical device', *J Am Chem Soc*, 135(3), pp. 1117-23.

Rajendran, A., Endo, M., Hidaka, K., Tran, P. L., Mergny, J. L., Gorelick, R. J. and Sugiyama, H. (2013b) 'HIV-1 nucleocapsid proteins as molecular chaperones for tetramolecular antiparallel G-quadruplex formation', *J Am Chem Soc*, 135(49), pp. 18575-85.

Rajendran, A., Endo, M., Hidaka, K., Tran, P. L., Mergny, J. L. and Sugiyama, H. (2013c) 'Controlling the stoichiometry and strand polarity of a tetramolecular G-quadruplex structure by using a DNA origami frame', *Nucleic Acids Res*, 41(18), pp. 8738-47.

Ramanathan, S. P., van Aelst, K., Sears, A., Peakman, L. J., Diffin, F. M., Szczelkun, M. D. and Seidel, R. (2009) 'Type III restriction enzymes communicate in 1D without looping between their target sites', *Proc Natl Acad Sci U S A*, 106(6), pp. 1748-53.

Ran, F. A., Hsu, P. D., Lin, C. Y., Gootenberg, J. S., Konermann, S., Trevino, A. E., Scott, D. A., Inoue, A., Matoba, S., Zhang, Y. and Zhang, F. (2013) 'Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity', *Cell*, 154(6), pp. 1380-9.

Reuter, M., Kupper, D., Meisel, A., Schroeder, C. and Krüger, D. H. (1998) 'Cooperative binding properties of restriction endonuclease EcoRII with DNA recognition sites', *J Biol Chem*, 273(14), pp. 8294-300.

Riggs, A. D., Bourgeois, S. and Cohn, M. (1970) 'The lac repressor-operator interaction. 3. Kinetic studies', *J Mol Biol*, 53(3), pp. 401-17.

Rinker, S., Ke, Y., Liu, Y., Chhabra, R. and Yan, H. (2008) 'Self-assembled DNA nanostructures for distance-dependent multivalent ligand-protein binding', *Nat Nanotechnol*, 3(7), pp. 418-22.

Roberts, R. J., Belfort, M., Bestor, T., Bhagwat, AS. and Bickle, T. A. (2003) ‘A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes’, *Nucleic Acids Res*, 31, pp. 1805–12.

Roberts, R. J., Vincze, T., Posfai, J. and Macelis, D. (2003) 'REBASE: restriction enzymes and methyltransferases', *Nucleic Acids Res*, 31(1), pp. 418-20.

Robinson, R. S. (1996), ‘Interactive computer correction of piezoelectric creep in scanning tunneling microscopy images’, *J. Comput.-Assist. Microsc*, 2(1), pp. 53 – 58.

Rothemund, P. W. (2006) 'Folding DNA to create nanoscale shapes and patterns', *Nature,* 440(7082), pp. 297-302.

Rouzina, I. and Bloomfield, V. A. (1996) 'Influence of ligand spatial organization on competitive electrostatic binding to DNA', *Journal of Physical Chemistry*, 100, pp. 4305-4313.

Saccà, B. and Niemeyer, C. M. (2012) 'DNA origami: the art of folding DNA', *Angew Chem Int Ed Engl*, 51(1), pp. 58-66.

Saito, M., Watanabe-Nakayama, T., Machida, S., Osada, T., Afrin, R. and Ikai, A. (2015) 'Spectrin-ankyrin interaction mechanics: A key force balance factor in the red blood cell membrane skeleton', *Biophys Chem*, 200-201, pp. 1-8.

Saiz, L. and Vilar, J. M. (2006) 'DNA looping: the consequences and its control', *Curr Opin Struct Biol,* 16(3), pp. 344-50.

Sanger, F., Nicklen, S. and Coulson, A. R. (1977) 'DNA sequencing with chain-terminating inhibitors', *Proc Natl Acad Sci U S A*, 74(12), pp. 5463-7.

Sannohe, Y., Endo, M., Katsuda, Y., Hidaka, K. and Sugiyama, H. (2010) 'Visualization of dynamic conformational switching of the G-quadruplex in a DNA nanostructure', *J Am Chem Soc*, 132(46), pp. 16311-3.

Schaeffer, T. E., Viani, M., Walters D. A., Drake B., Runge, E.K., Cleveland J.P., Wendman, M. A., Hansma, P. K. (1997) 'Atomic force microscope for small cantilevers', *SPIE. Proceedings,* 3009, pp. 48-52.

Schneider, S. W., Lärmer, J., Henderson, R. M. and Oberleithner, H. (1998) 'Molecular weights of individual proteins correlate with molecular volumes measured by atomic force microscopy', *Pflugers Arch*, 435(3), pp. 362-7.

Schwarz, F. W., van Aelst, K., Tóth, J., Seidel, R. and Szczelkun, M. D. (2011) 'DNA cleavage site selection by Type III restriction enzymes provides evidence for head-on protein collisions following 1D bidirectional motion', *Nucleic Acids Res*, 39(18), pp. 8042-51.

Seeman, N. C. (1982) 'Nucleic acid junctions and lattices', *J Theor Biol*, 99(2), pp. 237-47.

Seidel, R., van Noort, J., van der Scheer, C., Bloom, J. G., Dekker, N. H., Dutta, C. F., Blundell, A., Robinson, T., Firman, K. and Dekker, C. (2004) 'Real-time observation of DNA translocation by the type I restriction modification enzyme EcoR124I', *Nat Struct Mol Biol*, 11(9), pp. 838-43.

Shibata, M., Nishimasu, H., Kodera, N., Hirano, S., Ando, T., Uchihashi, T. and Nureki, O. (2017) 'Real-space and real-time dynamics of CRISPR-Cas9 visualized by high-speed atomic force microscopy', *Nat Commun*, 8(1), pp. 1430.

Shibata, M., Uchihashi, T., Ando, T. and Yasuda, R. (2015) 'Long-tip high-speed atomic force microscopy for nanometre-scale imaging in live cells', *Sci Rep*, 5, pp. 8724.

Shih, W. M. and Lin, C. (2010) 'Knitting complex weaves with DNA origami', *Curr Opin Struct Biol,* 20(3), pp. 276-82.

Silberberg, Y. R., Pelling, A. E., Yakubov, G. E., Crum, W. R., Hawkes, D. J. and Horton, M. A. (2008) 'Mitochondrial displacements in response to nanomechanical forces', *J Mol Recognit*, 21(1), pp. 30-6.

Slutsky, M. and Mirny, L. A. (2004) 'Kinetics of protein-DNA interaction: facilitated target location in sequence-dependent potential', *Biophys J,* 87(6), pp. 4021-35.

Smith, G. R., Amundsen, S. K., Dabert, P. and Taylor, A. F. (1995) 'The initiation and control of homologous recombination in Escherichia coli', *Philos Trans R Soc Lond B Biol Sci*, 347(1319), pp. 13-20.

Smith, R. M., Marshall, J. J., Jacklin, A. J., Retter, S. E., Halford, S. E. and Sobott, F. (2013) 'Organization of the BcgI restriction-modification protein for the cleavage of eight phosphodiester bonds in DNA', *Nucleic Acids Res*, 41(1), pp. 391-404.

Stanford, N. P., Szczelkun, M. D., Marko, J. F. and Halford, S. E. (2000) 'One- and three-dimensional pathways for proteins to reach specific DNA sites', *EMBO J*, 19(23), pp. 6546-57.

Suzuki, Y., Endo, M., Katsuda, Y., Ou, K., Hidaka, K. and Sugiyama, H. (2014) 'DNA origami based visualization system for studying site-specific recombination events', *J Am Chem Soc*, 136(1), pp. 211-8.

Szybalski, W., Kim, S. C., Hasan, N. and Podhajska, A. J. (1991) 'Class-IIS restriction enzymes--a review', *Gene*, 100, pp. 13-26.

Tafvizi, A., Mirny, L. A. and van Oijen, A. M. (2011) 'Dancing on DNA: kinetic aspects of search processes on DNA', *Chemphyschem*, 12(8), pp. 1481-9.

Tamulaitis, G., Sasnauskas, G., Mucke, M. and Siksnys, V. (2006) 'Simultaneous binding of three recognition sites is necessary for a concerted plasmid DNA cleavage by EcoRII restriction endonuclease', *J Mol Biol*, 358(2), pp. 406-19.

Taylor, J. D., Goodall, A. J., Vermote, C. L. M. and Halford, S. E. (1990) 'Fidelity of DNA recognition by the EcoRV restriction- modification system in vivo', *Biochemistry*, 29, pp. 10727– 10733.

Tebas, P., Stein, D., Tang, W. W., Frank, I., Wang, S. Q., Lee, G., Spratt, S. K., Surosky, R. T., Giedlin, M. A., Nichol, G., Holmes, M. C., Gregory, P. D., Ando, D. G., Kalos, M., Collman, R. G., Binder-Scholl, G., Plesa, G., Hwang, W. T., Levine, B. L. and June, C. H. (2014) 'Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV', *N Engl J Med,* 370(10), pp. 901-10.

Terry, B. J., Jack, W. E. and Modrich, P. (1985) 'Facilitated diffusion during catalysis by EcoRI endonuclease. Nonspecific interactions in EcoRI catalysis', *J Biol Chem*, 260(24), pp. 13130-7.

Thuring, R. W., Sanders, J. P. and Borst, P. (1975) 'A freeze-squeeze method for recovering long DNA from agarose gels', *Anal Biochem*, 66(1), pp. 213-20.

Topal, M. D. and Conrad, M. (1993) 'Changing endonuclease EcoRII Tyr308 to Phe abolishes cleavage but not recognition: possible homology with the Int-family of recombinases', *Nucleic Acids Res*, 21(11), pp. 2599-603.

Tranvouez, E., Boer-Duchemin, E., Comtet, G. and Dujardin, G. (2007) 'Active drift compensation applied to nanorod manipulation with an atomic force microscope', *Rev Sci Instrum*, 78(11), pp. 115103.

Vanamee, E. S., Berriman, J. and Aggarwal, A. K. (2007) 'An EM view of the FokI synaptic complex by single particle analysis', *J Mol Biol,* 370(2), pp. 207-12.

Vanamee, E. S., Viadiu, H., Kucera, R., Dorner, L., Picone, S., Schildkraut, I. and Aggarwal, A. K. (2005) 'A view of consecutive binding events from structures of tetrameric endonuclease SfiI bound to DNA', *EMBO J*, 24(23), pp. 4198-208.

van den Broek, B., Noom, M. C. and Wuite, G. J. (2005) 'DNA-tension dependence of restriction enzyme activity reveals mechanochemical properties of the reaction pathway', *Nucleic Acids Res*, 33(8), pp. 2676-84.

Venclovas, C., Timinskas, A. and Siksnys, V. (1994) 'Five-stranded beta-sheet sandwiched with two alpha-helices: a structural link between restriction endonucleases EcoRI and EcoRV', *Proteins*, 20(3), pp. 279-82.

Vipond, I. B. and Halford, S. E. (1995) 'Specific DNA recognition by EcoRV restriction endonuclease induced by calcium ions', *Biochemistry*, 34(4), pp. 1113-9.

Voigt, N. V., Tørring, T., Rotaru, A., Jacobsen, M. F., Ravnsbaek, J. B., Subramani, R., Mamdouh, W., Kjems, J., Mokhir, A., Besenbacher, F. and Gothelf, K. V. (2010) 'Single-molecule chemical reactions on DNA origami', *Nat Nanotechnol*, 5(3), pp. 200-3.

von Hippel, P. H. and Berg, O. G. (1989) 'Facilitated target location in biological systems', *J Biol Chem*, 264(2), pp. 675-8.

Watanabe, H., Uchihashi, T., Kobashi, T., Shibata, M., Nishiyama, J., Yasuda, R. and Ando, T. (2013) 'Wide-area scanner for high-speed atomic force microscopy', *Rev Sci Instrum*, 84(5), pp. 053702.

Watanabe-Nakayama, T., Machida, S., Afrin, R. and Ikai, A. (2010) 'Microscoop for manipulation of micro-objects: use of fabricated cantilever with atomic force microscope', *Small*, 6(24), pp. 2853-7.

Watson, J. D. and Crick, F. H. (1953) 'The structure of DNA', *Cold Spring Harb Symp Quant Biol*, 18, pp. 123-31.

Watson, M. A. and Chaconas, G. (1996) 'Three-site synapsis during Mu DNA transposition: a critical intermediate preceding engagement of the active site', *Cell*, 85(3), pp. 435-45.

Watson, M. A., Gowers, D. M. and Halford, S. E. (2000) 'Alternative geometries of DNA looping: an analysis using the SfiI endonuclease', *J Mol Biol*, 298(3), pp. 461-75.

Wentzell, L. M., Nobbs, T. J. and Halford, S. E. (1995) 'The SfiI restriction endonuclease makes a four-strand DNA break at two copies of its recognition sequence', *J Mol Biol*, 248(3), pp. 581-95.

West, P. E. (2007) 'AFM image artifacts. Introduction to atomic force microscopy: theory, practice, application', *Santa Clara, CA: Pacific Nanotechnology*. pp. 110–125.

Willemsen, O. H., Snel, M. M., Cambi, A., Greve, J., De Grooth, B. G. and Figdor, C. G. (2000) 'Biomolecular interactions measured by atomic force microscopy', *Biophys J,* 79(6), pp. 3267-81.

Winkler, F. K., Banner, D. W., Oefner, C., Tsernoglou, D., Brown, R. S., Heathman, S. P., Bryan, R. K., Martin, P. D., Petratos, K. and Wilson, K. S. (1993) 'The crystal structure of EcoRV endonuclease and of its complexes with cognate and non-cognate DNA fragments', *EMBO J,* 12(5), pp. 1781-95.

Xu, X. M. and Ikai, A. (1998) 'Retrieval and amplification of single-copy genomic DNA from a nanometre region of chromosomes: a new and potential application of atomic force microscopy in genomic research', *Biochem Biophys Res Commun*, 248(3), pp. 744-8.

Xu, Y., Sato, H., Sannohe, Y., Shinohara, K. and Sugiyama, H. (2008) 'Stable lariat formation based on a G-quadruplex scaffold', *J Am Chem Soc*, 130(49), pp. 16470-1.

Zadegan, R. M., Jepsen, M. D., Thomsen, K. E., Okholm, A. H., Schaffert, D. H., Andersen, E. S., Birkedal, V. and Kjems, J. (2012) 'Construction of a 4 zeptoliters switchable 3D DNA box origami', *ACS Nano*, 6(11), pp. 10050-3.

Zadegan, R. M. and Norton, M. L. (2012) 'Structural DNA nanotechnology: from design to applications', *Int J Mol Sci*, 13(6), pp. 7149-62.

Zahran, M., Daidone, I., Smith, J. C. and Imhof, P. (2010) 'Mechanism of DNA recognition by the restriction enzyme EcoRV', *J Mol Biol,* 401(3), pp. 415-32.

Zaremba, M., Owsicka, A., Tamulaitis, G., Sasnauskas, G., Shlyakhtenko, L. S., Lushnikov, A. Y., Lyubchenko, Y. L., Laurens, N., van den Broek, B., Wuite, G. J. and Siksnys, V. (2010) 'DNA synapsis through transient tetramerization triggers cleavage by Ecl18kI restriction enzyme', *Nucleic Acids Res*, 38(20), pp. 7142-54.

Zaremba, M., Sasnauskas, G., Urbanke, C. and Siksnys, V. (2005) 'Conversion of the tetrameric restriction endonuclease Bse634I into a dimer: oligomeric structure-stability-function correlations', *J Mol Biol*, 348(2), pp. 459-78.

Zhang, F., Wen, Y. and Guo, X. (2014) 'CRISPR/Cas9 for genome editing: progress, implications and challenges', *Human Molecular Genetics*, 23, pp. R40-R46.

Zhang, H., Yu, H., Ren, J. and Qu, X. (2006) 'Reversible B/Z-DNA transition under the low salt condition and non-B-form polydApolydT selectivity by a cubane-like europium-L-aspartic acid complex', *Biophys J,* 90(9), pp. 3203-7.

Zhong Q., Inniss D., Kjoller K., Elings V. (1993) 'Fractured polymer/silica fiber surface studied by tapping mode atomic force microscopy' *Surf. Sci. Lett*. 290