Understanding Physical Mechanisms in Chromatin Folding

by

Maheshwaram Sumanth Kumar

A thesis submitted to Jawaharlal Nehru University for the award of the degree of Doctor of Philosophy

Declaration

I, hereby declare that this thesis is composed independently by me at the Raman Research Institute, Bangalore, India, under the supervision of **Dr. Gautam V. Soni**. The subject matter presented in this thesis has not previously formed the basis for the award of any degree, diploma, membership, associateship, fellowship or any other similar title of any other university or institution. I also declare that I have run it through the **Turnitin** plagiarism software.

Signature of the Supervisor
Dr. Gautam V. Soni
Raman Research Institute
Bengaluru 560080

Signature of the Student
Maheshwaram Sumanth Kumar

Thesis Certificate

This is to certify that the thesis entitled **Understanding Physical Mechanisms in Chromatin Folding**, submitted by **Maheshwaram Sumanth Kumar**, for the award of the degree of **Doctor of Philosophy** of Jawaharlal Nehru University, Delhi, India is his original work. This has not been published or submitted elsewhere for the award of any degree or diploma.

Signature of the Director
Prof. Tarun Souradeep
Director
Raman Research Institute
Bengaluru 560080

Signature of the supervisor
Dr. Gautam V. Soni
Raman Research Institute
Bengaluru 560080

To my family,

for always being there for me

Acknowledgements

It has been an incredible learning experience working on my thesis under Gautam's supervision. His encouragement and guidance throughout my scientific journey at RRI are highly invaluable. Regular discussions and debates with him have helped me keep my spirits high, which is rather challenging over such a long journey. I admire his consistent positive outlook on life and his firm belief in simpler ways to do things.

Meetings with my advisory committee (Pramod and Raghu) have helped me get a holistic view of my thesis. I thank Sayantan for becoming a member of my advisory committee upon Raghu's retirement. Sandeep, Lakshminarayana, Ranjini, and Arun Roy, and EEG group members were very kind in allowing me to use their lab facilities whenever required. I thank Yateendran for all the SEM imaging of nanopores and for helping me operate AFM and Confocal microscope. I also thank Dhason for his initial help with SEM imaging.

I thank Serene and Divya for working religiously on optimizing protocols for all the samples used in this thesis work. I also thank Lavanya, Jyoti and Surabhya for their timely support with samples for various experiments. I thank Mahesh for his time and effort in optimizing the protocols and performing AFM imaging of biomolecules. I thank Jayesh for introducing me to the lab work, particularly the nanopore experiments. I thank Saurabh and Sukanya for providing critical feedback about my results and manuscripts. I also appreciate Manohar, Koushik, Nagesh, Punit, Prabhakaran for being great labmates.

Away from the lab, I enjoyed the company of Deepshika, Irla, Saichand, Surya, Niranjan, Jaggu, Anindya, Sreyas, Atul, Subodh, Sagar, Sreeja, Sanjay, Dubey, Mari, Swamy, etc. in a wide variety of activities like bike rides, hiking, cooking, and many more apart from discussing science.

I thank the Admin, academics, SAAC including C.S.R. Murthy, Naresh, Radha, Vidya, Merissa, Harini, Gayatri, Shailaja for taking care of various aspects of my PhD program; KRK, Amudha, Chaitanya, Raja and Murali for helping us with lab management-related activities; Mechanical workshop (Ibrahim) for fabrication of various components required for my experiments like Faraday cage, fluid cells, etc.; Electrical dept (Muneeshwaran, GB) for helping me with various electrical issues in the lab and hostel; Sarvagopalan (RAL) for initial help with troubleshooting electrical problems in the nanopore system; IT department, Purchase, Accounts, Library (Manju, Nagaraj, etc.), and Canteen for helping with various other aspects.

I sincerely thank Prasanta Panigrahi, Alok Pan and Aditi Sen for preparing me for a research career by allowing me to work under their guidance for summer projects at IISER Kolkata, NIT Patna and HRI during my undergraduate studies. I also thank my undergraduate teachers, Raghu, Madhavan, Chandrasekhar and Sivaji at NIT Calicut for supporting me in doing my best.

Finally, but most importantly, I thank all my family members for their exceptional support and encouragement in letting me pursue my dreams of career and life in general.

Abstract

The mechanics of DNA, affecting its structures, control multiple biological processes such as growth, division and differentiation of a cell resulting in the development and survival of organisms. Packaging DNA inside a cell is a fundamental biophysical problem directly relevant to gene regulation. We used gel electrophoresis, AFM imaging and primarily Nanopore-based single-molecule technique to study these structural conformations. This thesis highlights our investigations on structural forms of DNA polymer and nucleosome arrays. This thesis is divided into six chapters, as summarized below.

Chapter 1 will discuss the two basic chromatin folding mechanisms that result in supercoiled and nucleosome structures and the critical parameters that need to be studied to understand them better.

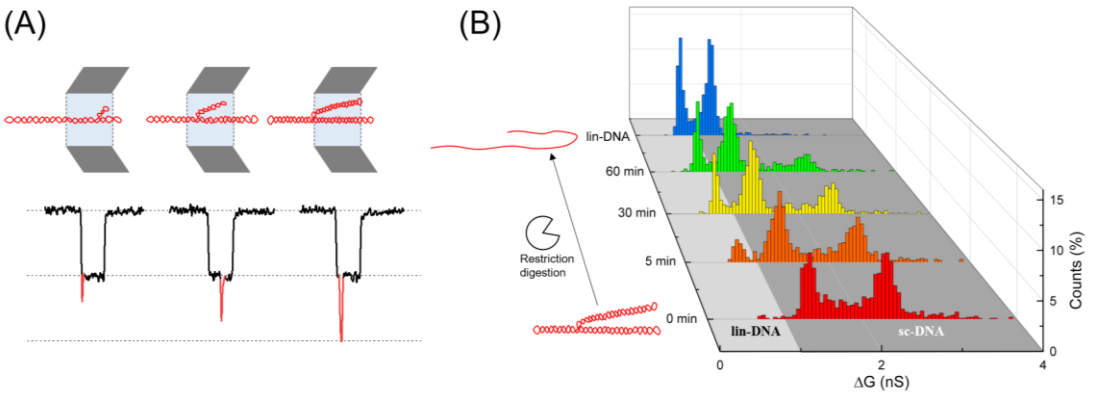


Figure 0.1: Study of plasmid conformations using nanopores

(A) It shows high-resolution detection of branches on supercoiled DNA. (B) Shows the nanopore-based quantification of enzyme-dependent conversion of supercoiled DNA to linear DNA as a function of reaction time.

Chapter 2 will demonstrate the methods to perform resistive pulse technique using custom-made nanopores for high-resolution studies. We present the results on characterization of nanopores and use the existing theoretical models to obtain a better understanding of the pore geometry. We also show the multi-fold events, a characteristic feature of λ-DNA translocation through nanopores. Later we also discuss about other techniques like atomic force microscopy (AFM) and gel electrophoresis for characterization of the samples prepared in-house.

Chapter 3 will report the study of various plasmid conformations – linear, circular and supercoiled DNA using nanopores. Then we present our results obtained using ~ 20 nm sized pores, on the heterogeneity in branches formed on supercoiled plasmid DNA and we explain the ECD-based analysis methods to characterize the branch lengths and positions on each molecule (Fig 0.1A). From these results, we also obtained the value of effective pore length that is crucial for volumetric quantification of biomolecules. Next, we show the nanopore-based quantification for the enzymatic conversion of supercoiled DNA to linear DNA (Fig 0.1B).

Chapter 4 will report the study on detection of mononucleosomes using nanopores. Here, we show that smaller sub-nucleosome structures are formed due to the breaking of mononucleosomes while they are translocating through the nanopore (Fig 0.2A). Further experiments showed that the smaller-structures are predominantly detected at higher voltages. We performed volumetric quantification of these structures using DNA for the nanopore calibration. We present a model for this electric-field induced breakdown of nucleosomes and show that the smaller structures are most likely the hexasomes and tetrasomes.

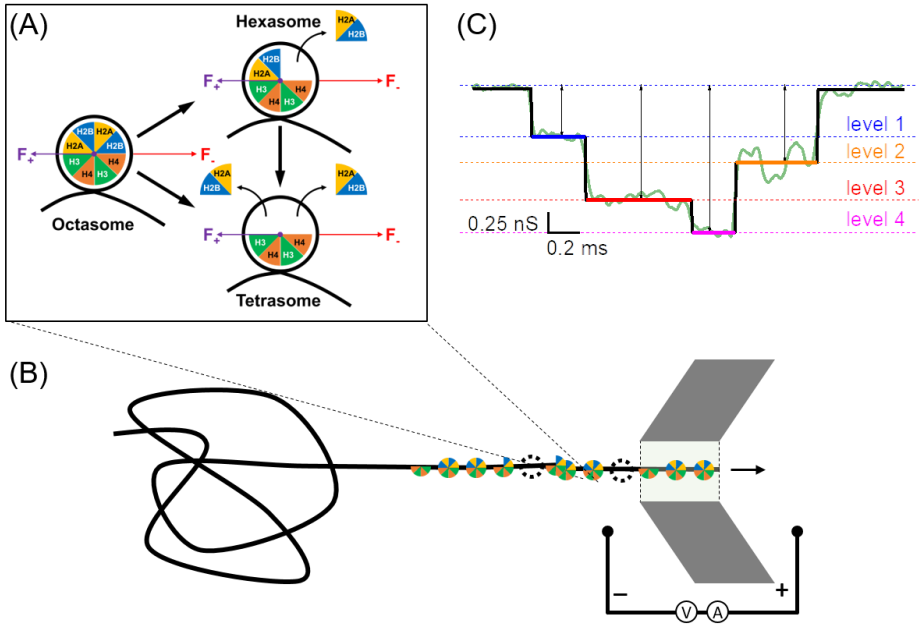


Figure 0.2: Probing structural stability of mononucleosomes and nucleosome arrays using nanopores

(A) Schematic depicts the breaking of mononucleosomes (octasomes) to hexasomes and tetrasomes. (B) Schematic shows the breaking of nucleosomes on the 12mer arrays during its translocation. (C) A multi-level generated during the translocation of 12mer nucleosome array.

Chapter 5 will report the study on characterization of the nucleosome states on the

nucleosome arrays (Fig 0.2B). The analysis methods and the interpretation of these structures from the multi-level electrical signatures are discussed (Fig 0.2C). The interpretation of various biologically relevant structures from the nanopore-generated electrical signatures would be discussed in detail. In summary, this thesis work informed us of the huge heterogeneity in the structures that is possible at the level of DNA and DNA-protein complexes.

Chapter 6 will conclude the results of the thesis and propose ways forward in applying single-molecule techniques to make advancements in the field of chromatin biology.

List of publications

1. S. K. Maheshwaram, K. Sreenivasa, and G. V. Soni, "Fingerprinting branches on supercoiled plasmid DNA using quartz nanocapillaries", *Nanoscale* 13, 320 (**2021**).
2. Serene R. David[#], S. K. Maheshwaram[#], Divya Shet, Mahesh B. Lakshminarayana and G. V. Soni, "Temperature dependent in vitro binding and release of target DNA by Cas9 enzyme", *Scientific Reports* 12, 15243 (**2022**) ([#]Equal contribution).
3. S. K. Maheshwaram, Divya Shet, Serene R. David, Mahesh B. Lakshminarayana and G. V. Soni, "Nanopore sensing of DNA-histone complexes on nucleosome arrays", Submitted.

Signature of the Supervisor
Dr. Gautam V. Soni
Raman Research Institute
Bengaluru 560080

M. Sumanth

Signature of the Student
Maheshwaram Sumanth Kumar

CONTENTS

List of figures

List of tables

CHAPTER 1

Introduction

A cell is the basic unit of life and carries out complex processes with great precision for its growth and division. This is accomplished by obeying the instructions coded in the inheritable genome in the form of Deoxyribonucleic acid (DNA). The building blocks of DNA are a set of four nucleotides made up of nucleobases (Adenine, Guanine, Thymine and Cytosine represented as A, G, T and C respectively) along with a deoxyribose and a phosphate group. These nucleotides covalently bind to each other in a directional manner to form a linear heteropolymer which code for the synthesis and functioning of various proteins and enzymes. Two such strands can hybridize together when their sequences are complementary to each other, say A and G on one strand binds only with T and C respectively on the other strand. This results in a double helical structure which is the

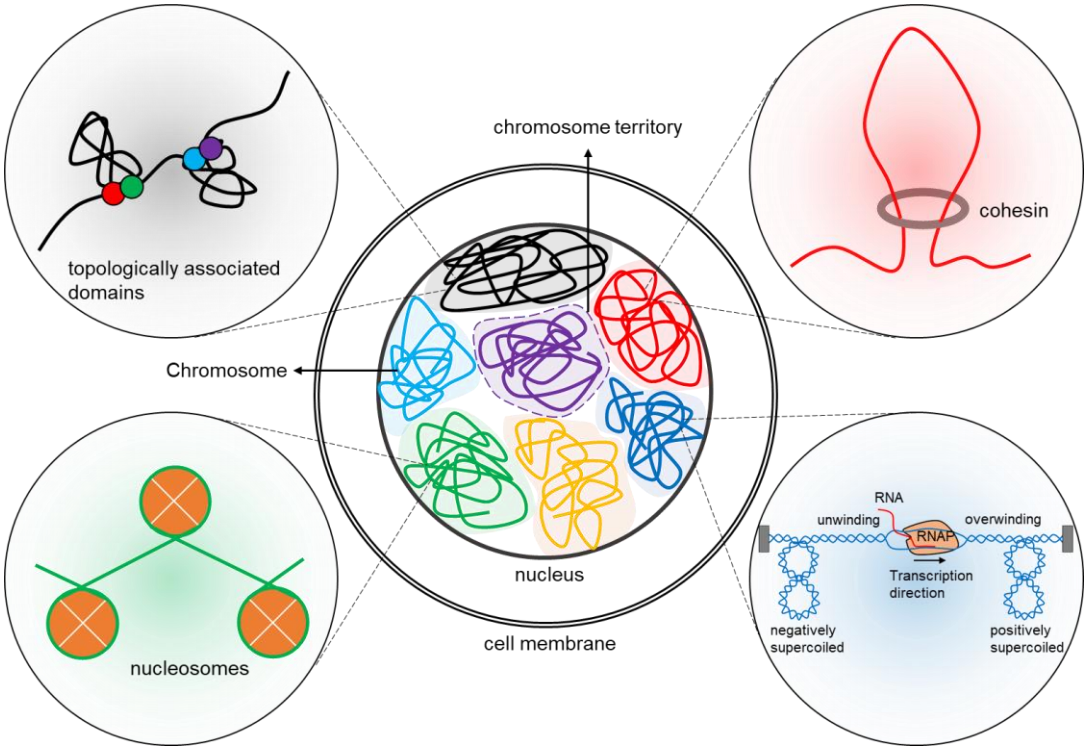


Figure 1.1: Schematic of a eukaryotic cell

It shows the various features such as supercoiling, nucleosomes, topologically associated domains and cohesion induced looping required for efficient gene packaging and regulation.

stable form of DNA found in all the living organisms. The DNA is negatively charged as each of the phosphate in its nucleotides contributes to a negative charge and hence electrostatics plays a key role in its organization. The repulsion among various portions of the polymer creates unfavourable conditions and poses challenges in packing, for example, a few meters long human genome into a cell nucleus which is only few microns in size. The random coiling of DNA to achieve such a compaction is an inefficient mechanism and necessitates the involvement of many other factors as shown in Fig 1.1. Bacteria takes advantage of the supercoiling mechanism where the twist is introduced in the DNA to bend it and form loops resulting in its compaction [1,2]. Eukaryotic cells on the other hand, forms nucleosomes by wrapping DNA around positively charged protein complexes called histone octamers to achieve a high level of organization. In this chapter, we elaborate on these mechanisms in detail which provides a basic understanding of the biomolecular structures investigated in this thesis.

1.1 Supercoiling

Commonly occurring DNA is a right-handed DNA helix, making a complete revolution for every 10 base pairs with the inter-base pair distance being 0.34 nm. It resists any torsional deformation that would change the natural twist. The twist in a polymer is mathematically represented by the linking number (Lk). In the relaxed state, the linking number (Lk_0) of DNA is calculated as

$$\mathrm{Lk}_0(\mathrm{turns}) = \frac{\text{Total number of basepairs in DNA}}{(\text{Number of basepairs})/\text{turn}} \qquad (1.1)$$

For example, a 150 bp DNA has a linking number of 15 as the number of basepairs/turn is 10. Consider the situation shown in Fig 1.2A where the DNA is immobilized and twisted in left-handed direction by making a full rotation at one of its ends. This reduces the overall twist in the right-handed helix of the DNA by one turn and it is said to have acquired a twist (Tw) of -1. When the ends are set free to only translate along the axis of DNA, keeping the rotation fixed, the DNA converts from an under twisted helix to a naturally twisted state but accompanied by the formation of a right-handed superhelix

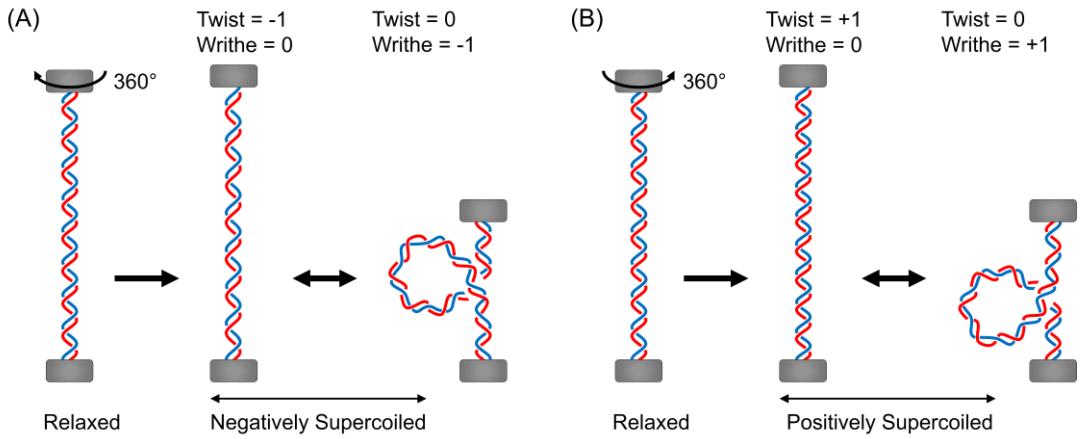


Figure 1.2: Schematic of supercoiled DNA

Mechanical twisting of DNA leads to torsional stress and results in conformations whose sum of twist and writhe is conserved. The reduction of twist in the DNA leads to negative supercoiling (A) and increase in twist leads to positive supercoiling (B).

with one self-intersection point. The resultant structure is called a plectoneme with the superhelicity represented by a quantity called Writhe (Wr), which in this case has a value of -1. The quantities, twist and writhe can change as they are interconvertible but their sum which is called the change in linking number (ΔLk) is a topologically invariant quantity [3]:

$$\Delta Lk = Tw + Wr \tag{1.2}$$

The value ΔLk, which has a value of -1 in this case, remains constant as long as the torsional stress is retained within the molecule. In the relaxed state, the values of both the Tw and Wr are 0 and hence ΔLk is also 0. When the DNA is overtwisted as shown in Fig 1.2B, it assumes a left-handed superhelix upon bringing its ends closer and the Tw, Wr and ΔLk have opposite signs but equal magnitudes. The interconvertibility of twist and writhe implies that a DNA loop can be formed either by bending or simply twisting it which is highly relevant for the nucleosome formation. The circular DNA in the bacteria is prone to supercoiling when it is cut open and ligated back upon acquiring torsional stress due to the cellular processes.

In order to compare the amount of torsional stress or degree of supercoiling in a DNA molecule that is independent of its length, a quantity called the supercoiling density is defined as

$$\sigma = \Delta Lk / Lk_0 \tag{1.3}$$

The bacterial DNA is found to naturally exist as a negatively supercoiled DNA [4,5] with a σ of -0.06. Positively supercoiled DNA is extremely rare and was observed only in the hyperthermophilic bacteria [6], as a mechanism to prevent its degradation at high temperatures. As the supercoiling density increases, the DNA transitions from relaxed circular form to plectoneme forms with increasing number of loops. Further increase in the supercoiling density results in formation of branched plectonemic structures. These structures have been characterized by electron microscopy [1] and AFM imaging [7–9].

1.2 Significance of supercoiling

The supercoiling of DNA is important in various biological processes. The stable helical structure makes it inevitable for the DNA to avoid torsional stress during any biological process that requires access to the genome. The double helix needs to be unwound for the genetic machinery to get access to the underlying sequence. For example, the RNA polymerase during transcription, melts the DNA and unwinds it before reading the sequence. Polymerase combined with the mRNA forms a heavy complex and leads to the rotation of DNA while zipping through the genes. The unwinding of DNA results in negative supercoiling behind polymerase and the over winding results in positive supercoiling ahead of it. The negative supercoiling favours the formation of nucleosomes whereas the positive supercoiling ahead results in the unwrapping of nucleosomes making it easier for the polymerase movement (Fig 1.3A left [10]). Similar situation arises during the replication process where the DNA has to be opened up according to the movement of the replication fork (Fig 1.3A right [10]). Supercoiling is also known to have consequences in the formation of alternate DNA structures, such as DNA breathing, cruciform structures, or Z-DNA structure where the DNA helix is in the opposite direction (Fig 1.3B [11]). Conversely, The presence of DNA bending structures such as cruciform and A-tract were observed to have an effect on the branching point leading to a change in the global structure of the plasmid [12]. This is also an efficient mechanism to bring distant sites on the DNA together for their communication required for processes such as enhancer-promoter interaction [13,14], genetic recombination (Fig 1.3C [15]), etc. The dynamics of the plectonemes on a linear DNA molecule have been studied using techniques like magnetic tweezers and fluorescence microscopy [16–20] whereas the

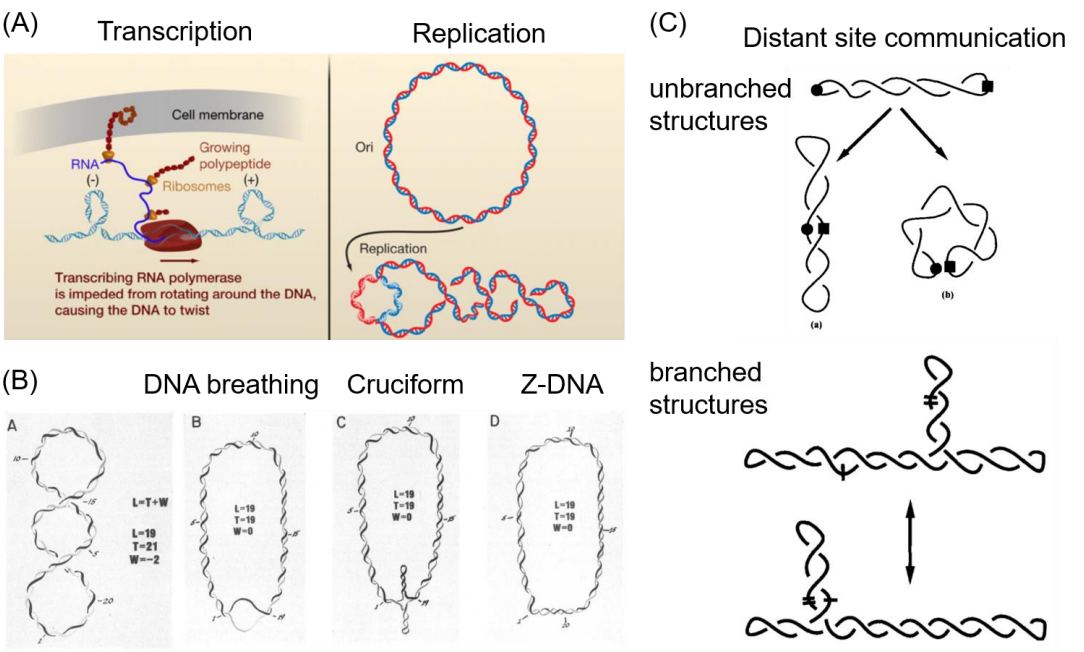


Figure 1.3: Significance of supercoiling in biology

(A) The processes such as transcription and replication are affected due to the formation of supercoiled loops ahead of the RNA polymerase and replication fork respectively. (B) Supercoiling favors the formation of alternate DNA structures such as DNA breathing, cruciform and Z-DNA. (C) The distant sites on DNA can come into contact by slithering or bending in case of unbranched supercoiled DNA and branch sliding in case of branched supercoiled DNA. The figures (A)-(D) are Reproduced with permission from Refs [10], [11] and [15] respectively.

reports available on the dynamics of branched supercoiled DNA is very limited [9]. In this thesis, we have studied the structures of branched supercoiled DNA.

1.3 Nucleosome structures

Chromatin was initially thought to exist as a 30 nm fiber inside the cell as observed from electron microscopy studies [21]. Further investigations revealed the bead on a string pattern with the beads spaced at equally separated from each other all over the length of the fibre [22]. It seemed as if there exists a basic unit of chromatin which later turned out to be nucleosome. The other proof that chromatin is quantized comes from the experiment where it was partially digested and analyzed using gel electrophoresis. A ladder like pattern was observed with a huge number of bands and equally distant from each other. This confirmed that the chromatin is made up of some repeated features [23]. In order to understand the structure of chromatin in detail, the basic units of chromatin were extracted by MNase digestion and the X-ray crystallography studies were

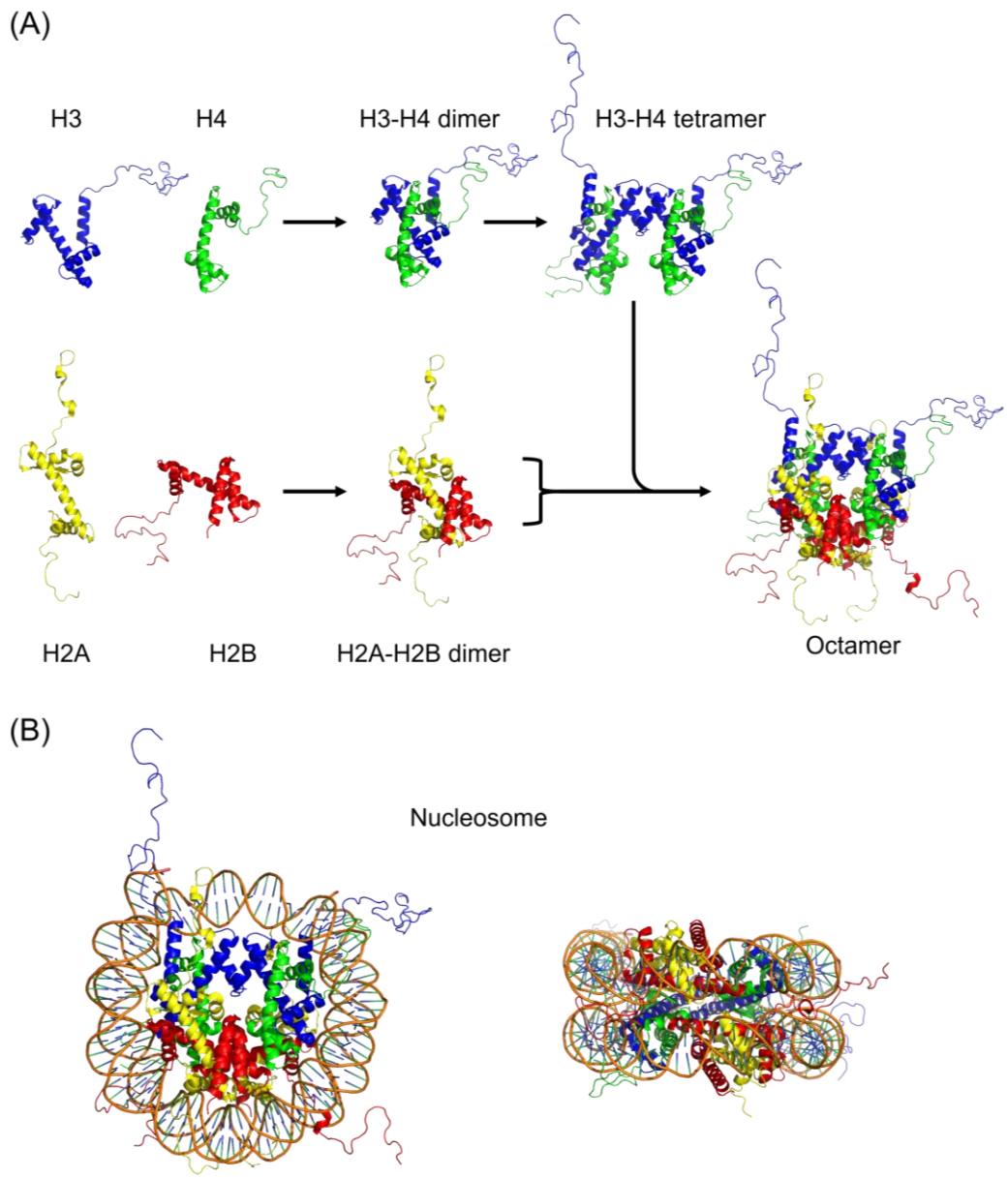


Figure 1.4: Schematic showing the assembly of histone octamer and nucleosome complex

(A) Histone octamer is formed by a H3-H4 tetramer and two copies of H2A-H2B dimers. (B) Nucleosome complex shown in face-on (top panel) and edge-on (bottom panel) views. Crystal structure 1KX5 (PDB ID) was used for the representation of histones and nucleosome complexes with the help of PyMOL software.

performed. Karolin Luger obtained the mononucleosome crystal structure at a high resolution of 2.8 Å that led to its detailed understanding [24]. Nucleosome consists of 147 bp DNA and a complex of eight histone octamers, a pair of each of the H2A, H2B, H3, H4 proteins as shown in Fig 1.4. It was found that there are about 14 contacts between DNA and histone octamer. The nucleosome is a highly stable structure and can withstand high temperatures. Extensive theoretical work was done in order to understand the physics involved in chromatin formation [25]. Nucleosome formation requires the

bending of DNA around the histone octamer but energy needs to be spent in bending DNA. The energy is provided by the energy obtained due to the electrostatic attraction between negatively charged DNA and positively charged histones. 147 bp DNA has a charge of 294 e^- and histone octamers carry a charge of 220 e^+ out of which 134 e^+ are buried inside the octamer. The nucleosomes are stabilized due to the two opposing mechanisms, electrostatic interaction between oppositely charged DNA and histone octamer and the bending of DNA around the octamer. From the crystal structure of nucleosome, it was known that DNA contacts histones at a total of 14 points. It was estimated that about -2 k_BT is the free energy (ΔG) for nucleosome formation per contact site which would be about 14 x (-2) = -28 k_BT per nucleosome [26,27]. The bending energy of DNA (E_b) can be found using

$$\frac{E_b}{k_B T} = \frac{L_p L_{wrap}}{2R_0^2} \tag{1.1}$$

Here, the persistence length (L_p) is 50 nm, the length of DNA wrapped around the octamer (L_{wrap}) is 127 bp x 0.34 nm/bp = 43.2 nm, the radius of DNA loop (R_0) is 4.3 nm. Using these values, E_b is estimated to be +58 k_BT. As we know that $\Delta G = E_{el} + E_b$, the pure electrostatic energy (E_{el}) is about -86 k_BT due to the interaction between DNA and histones which reveals that 67 % of the energy is used in bending the DNA. This shows that nucleosomes are quite stable complexes.

1.4 Chromatin folding

The genome in eukaryotes is tightly packed inside the nucleus in a hierarchical order as shown in Fig 1.5 [28,29]. The chromatin folds into higher order structures which introduces an additional physical barrier apart from the individual nucleosomes in the gene regulatory mechanism [25,30,31]. The global chromatin structure is dynamic and was observed to vary depending on the stage of the cell cycle, such as it assumes euchromatin form in the interphase and heterochromatin form in the mitotic phase. DNA packaging in the form of chromatin is important for biological function. Although the key players in chromatin folding are known, the underlying biophysical interactions and dynamic tunability mechanisms of these interaction forces are unknown. The basis for

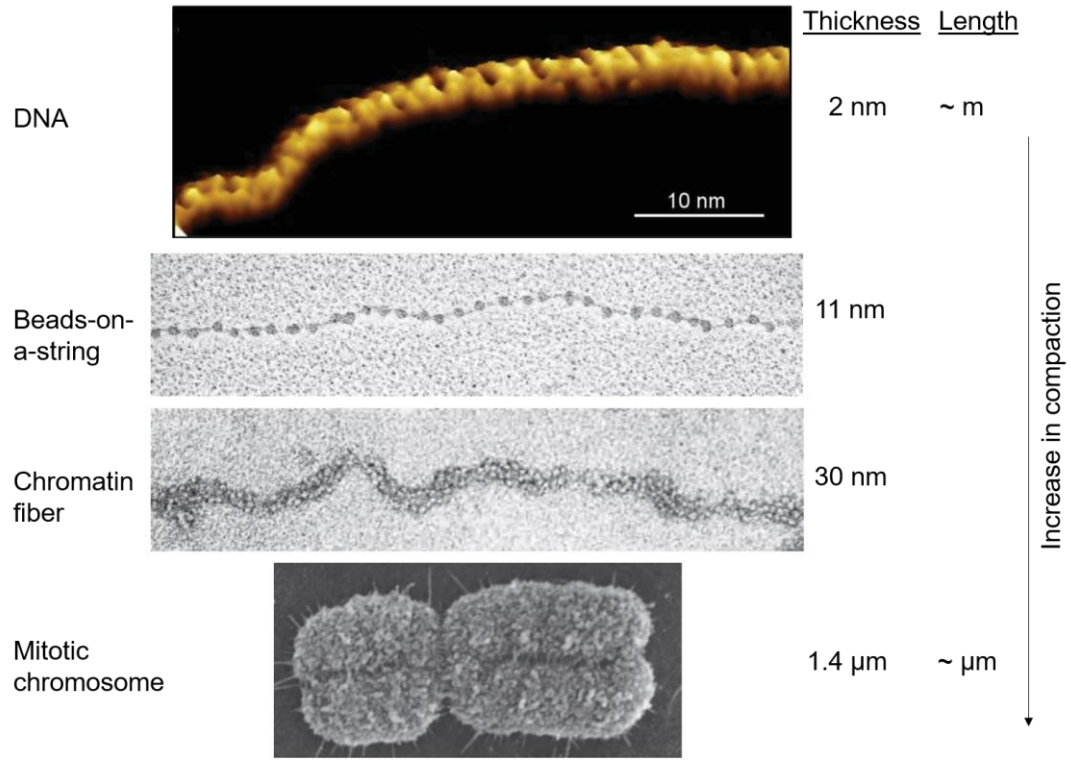


Figure 1.5: Chromatin packing

This model shows some of the many levels of chromatin packing postulated to give rise to the highly condensed mitotic chromosome. (A) and (B)-(D) are reproduced from Refs [29] and [28] respectively.

higher order structures is believed to be dominated by interactions between these nucleosomes [32]. In vitro experiments have shown that chromatin transitions from a 10 nm fiber to a 30 nm fiber followed by large globular structure as the salt concentration increases. There are various factors that affect the chromatin structure like, the salt concentration, nucleosome interactions, and its interactions with enzymes. The interest in measuring the physical quantities related to biomolecular systems is growing especially in the context of chromatin folding. Numerous *in vitro* experiments have provided great insights and made our understanding better. Reports on previous research work propose that a significant role in chromatin folding is played by inter-nucleosomal interactions via their histone tails [33]. It was conjectured that epigenetic modifications like acetylation, methylation, etc., on the nucleosome tails gives further control on the chromatin folding [34]. Modification on H3 and H4 tails were found to be more effective in changing the nucleosome-nucleosome interactions compared to other histones. The removal of tails has drastically reduced the attractive interaction leading to the loss of ability for chromatin compaction [35,36]. It was shown that chromatin exists in open

form when the divalent ions are low in concentration and becomes folded or compact at higher concentrations as shown by SAXS [37], ultracentrifugation [38], and electron microscopy [39]. Experiments with higher valent positive ions such as Mg^{2+}, Spd^{3+}, Sp^{4+}, $Co(NH_3)_6^{3+}$ and found that the chromatin compaction requires lower concentrations of ions when the valency is higher which works by bridging the negative surfaces of each of the nucleosomes [38]. These interactions are modulated when the nucleosomes are epigenetically modified as studied using bulk analysis techniques such as ultracentrifugation [32,36,38], and single-molecule techniques such as fluorescence microscopy [40], optical tweezers [41–43], magnetic tweezers [44–46], AFM [47,48] and DNA origami-based force spectrometer [49], etc. The force corresponding to the rupture of interactions between neighbouring nucleosomes was estimated to be about 3 - 14 k_BT [42,44,49]. The interaction strength between the nucleosomes was found to be reduced upon either acetylation or removal of N-terminal tails [49], thus modifying the state of chromatin.

The chromatin structure is also regulated by the interaction of nucleosomes with external architectural chromatin proteins (ACP) [50]. Currently we know the proteins involved in this process like PRC2, HP1, MeCP2, RCC1, etc., but the physical mechanisms involved are not well understood. It was observed in vivo that chromatin regions with like-epigenetic modifications are organized into separate domains [31]. The chromatin inside the cell is continuously modified chemically in order to have a better control on giving access to the genetic machinery. The state of the chromatin is classified as transcriptionally active, inactive or polycomb-repressed states when marked with H3K4me2, unmodified H3 or H3K27me3 respectively. In vivo, it was found that the nucleosomes are organized in the form of clutches [30]. The density of the nucleosomes in the clutches was found to be lower in case of stem cells compared to the differentiated cell. This is important to understand since a large number of diseases and disorders are directly correlated to chromatin folding and its structural failures. For example, defects in chromatin repression due to the mutations in ATRX gene leads to Alpha thalassemia mental retardation syndrome. Similarly, defects in chromatin activation leads to Rubenstein-Taybi syndrome which causes congenital malformations that include cardiac anomalies [51]. The reduced expression of Heterochromatin Protein1 (HP1) causes defects in chromatin organization which has implications in breast, brain, colon

cancers [52]. The individual contribution of various factors involved in the chromatin condensation process is yet to be understood.

1.5 Partial nucleosome structures

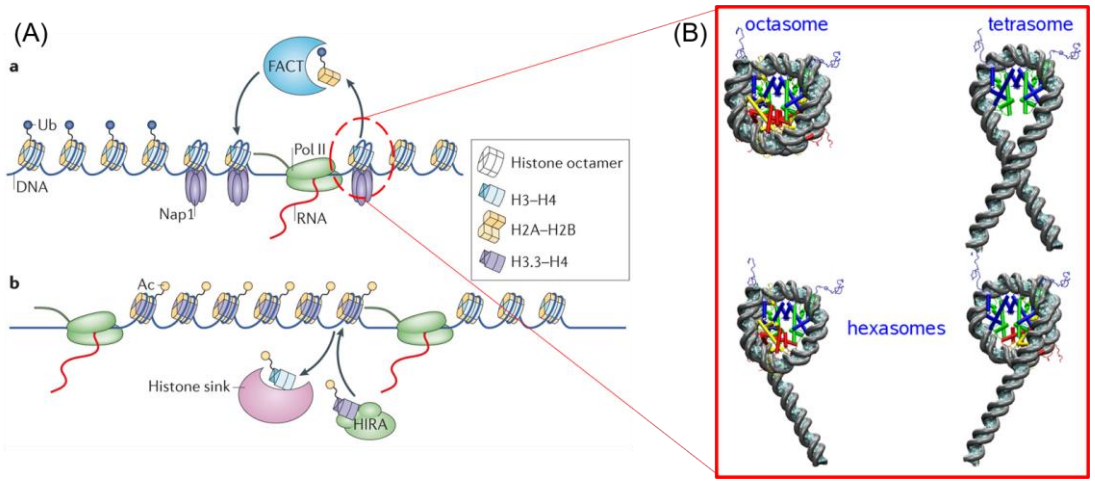


Figure 1.6: Octasomes, hexasomes and tetrasomes

(A) a. During the transcription process, the RNA polymerase movement requires the access to DNA which could be achieved by partial removal of histones from the nucleosome. The red dotted circle shows the nucleosome which can exist as octasomes, hexasomes or tetrasomes as shown in (B). In (B), the structure of nucleosomes is based on the crystal structure (PDB ID: 1KX5) but the hexasomes and tetrasomes are obtained from computer simulations based on nucleosome (octasome) structure. (A) and (B) are reproduced with permission from Refs [56] and [54] respectively.

Along with the packaging of DNA, there should also be mechanisms to open it up whenever required for the biological processes. The most important factor here is the stability of mononucleosomes which has been studied extensively as the modifications of the nucleosome structures are quite important during various biological processes. At much smaller length scales, nucleosomes being the basic units of chromatin regulate the gene expression by controlling the access to the underlying sequence with the help of certain factors such as chromatin remodellers, chemical modifications and histone variants. It was observed that the passage of RNA polymerase through the nucleosome results in the loss of a H2A-H2B dimer which suggests that partial nucleosomes could be seen as a foot-print of transcription process [53]. Using ChIP-seq techniques, it was found that the hexasomes and tetrasomes [54] (Fig 1.6B) from the nucleosomes is directly related to the gene transcription levels [55] due to the movement of RNA polymerase along the DNA [56] (Fig 1.6A). Another genome-wide study has suggested

the existence of sub-nucleosomal particles with a single copy of all four histones which were called half-nucleosomes [57]. Various partial nucleosome species were observed and quantified using AFM [58,59] and FRET [60] studies under varied salt conditions. The authors of the paper [59] suggested the existence of four different nucleosome species – disomes, tetrasomes, hexasomes and octasomes (intact nucleosomes) with the help of molecular dynamic simulations in combination with AFM imaging and FRET and SAXS data available from previously published reports. The octasome has one H3-H4 tetramer and two H2A-H2B dimers, a hexasome has one H3-H4 tetramer and one H2A-H2B dimer, a tetrasome has one H3-H4 tetramer and a disome has one H3-H4 dimer. The crystal structures of the hypothesized nucleosome structures other than the octasomes are currently not available. The mechanism behind nucleosome formation and its kinetics under in vivo like conditions was reported [61]. It was found that NAP1 plays a key role in proper assembly of nucleosomes and it takes place through tetrasomes and hexasomes as the intermediates and their reaction kinetics is reported. They could also be the intermediate structures required for the histone exchange process. The formation and stability of partial structures could also play a key role in processes where access to genetic information is required, as they are less stable compared to intact nucleosomes and hence the sequence of the nucleosome protected DNA will be easily accessible to the enzymes as they need lower energy to break the histone-DNA interactions. In this thesis, we have studied these nucleosome sub-structures in detail.

CHAPTER 2

Experimental techniques

Abstract

This chapter explains various techniques in the thesis such as Glass nanocapillaries, AFM and gel electrophoresis. All the steps involved in setting up the nanopore experiment beginning with the fabrication of glass nanocapillaries, mounting them in fluid cells, preparation of electrodes and the instrumentation required to record the measurements are demonstrated. The nanopores in a vast size range are characterized and the results are explained with the existing model. The resolution of a nanopore device is demonstrated by distinguishing the number of DNA folds assumed by DNA during its translocation. Then the gel electrophoresis which is a fundamental and important technique for molecular biology studies is explained. We also explain the working principle of atomic force microscopy imaging as it will be used for various studies presented in this thesis. We end this chapter by demonstrating the application of gel electrophoresis and AFM for a quantitative study of Cas9-DNA kinetics. The rest of the thesis builds upon the experimental techniques presented in this chapter.

2.1 Nanopore technology

The most common techniques used for studying nanoscale objects are atomic force microscope (AFM) and electron microscope (EM). The drawback of AFM is that the molecules must be attached to the substrate and in case of EM either the fixation or freezing of molecules is necessary. The optical microscopy techniques have the limitation on spatial resolution which is only up to the order of the wavelength of light due to diffraction. Nanopore platform has recently emerged as an excellent platform for detecting various nano-scale objects (biological or synthetic) in a label-free way. The nanopore device works on the principle of resistive pulse technique developed by Coulter [62] for detecting micron sized cells. In this technique the nanopore acts as a channel between two reservoirs filled with electrolyte. When a bias voltage is applied across the nanopore, open pore current is measured due to the movement of electrolyte ions. For the duration an analyte particle translocates through the nanopore, it displaces ions in the nanopore resulting in conductance drop. These electrical events (conductance drops) not only record presence of analyte particles, they also contain details about the size and shape of the particle. Various types of nanopores are used which are made from silicon oxide [63], silicon nitride [64], carbon nanotubes [65], graphene [66], and glass capillaries [67–69]. They have been used to detect molecules like DNA, RNA, proteins and protein-DNA complexes [70–77]. The application of nanopores in the detection of conformational changes in proteins is also demonstrated [78–80]. They have been used for measuring the kinetics of biomolecular reactions [80–85] and to study conformations of DNA in the distribution of knot conformations [86,87], G-quadruplexes [88] and supercoiled DNA [89–91]. There are recent reports on fingerprinting proteins on DNA or RNA such as dCas9 [75,92,93], RNA polymerase [94] and ribosomes [95] and the sub-structures on the DNA such as knots [86,87] and G-quadruplexes [88] using nanopores. The binding modes of a transcription factor when complexed with DNA were also well-resolved using nanopores [96]. Previous studies also demonstrated the measurement of interactions involved in stabilizing STV-biotin complex [97], EcoRI-DNA complex [98], and nucleosomes [99,100]. A quantitative analysis of nucleosomes and its sub-components such as histone octamers, tetramers and monomers was earlier reported using nanopores [101].

Nanopores are unique in the sense that it allows us to detect the conformations as well as apply forces on the molecules thereby providing us a possibility to investigate both the aspects at the same time. We believe nanopore is an ideal technique to study conformations of supercoiled and nucleosome structures with high-throughput. The quantitative studies performed during this thesis work would help us understand the general principles governing the chromatin organization in the nucleus.

The following section details the developments made in this technique since its invention by Coulter in 1953.

2.1.1 Evolution of resistive pulse technique

1. Micron sized pores to detect micron sized particles:

The resistive pulse technique was invented by Wallace H. Coulter in 1953 in an attempt to simplify the procedure for blood count analysis [62]. In this technique, a membrane separates the two vessels or reservoirs filled with an electrolyte solution and the channel is made part of an electrical circuit by placing electrodes in either of the vessels as shown in Fig 2.1 (left). The suspension of blood samples is added in the left vessel and the fluid level is maintained to be higher than the right vessel. This creates the pressure difference and the fluid flows also taking the suspended cells from left vessel to the right side through the channel. In Fig 2.1 (right-top), the schematic shows the reference current (i_0) measured when there is no particle in the channel. When a particle of conductivity higher than the medium conductivity moves through the channel, this leads to an increase in the current (i_1) as shown in Fig 2.1 (right-middle) and goes back to its reference value once the particle moves out of the channel. Similarly, when a particle of conductivity lower than the solution conductivity passes through the channel, it results in a momentary decrease in the current (i_2) as shown in Fig 2.1 (right-bottom). By counting the number of spikes detected due to current changes in a particular duration and estimating the volume of fluid moved through the channel, the concentration of cells in the suspension is estimated. From the analysis of magnitude of current changes, the cellular volumes could be estimated and this helps in distinguishing different blood types of blood cells and gives a complete blood count. This technique is now routinely used in diagnostics setting and these instruments are called Coulter counters.

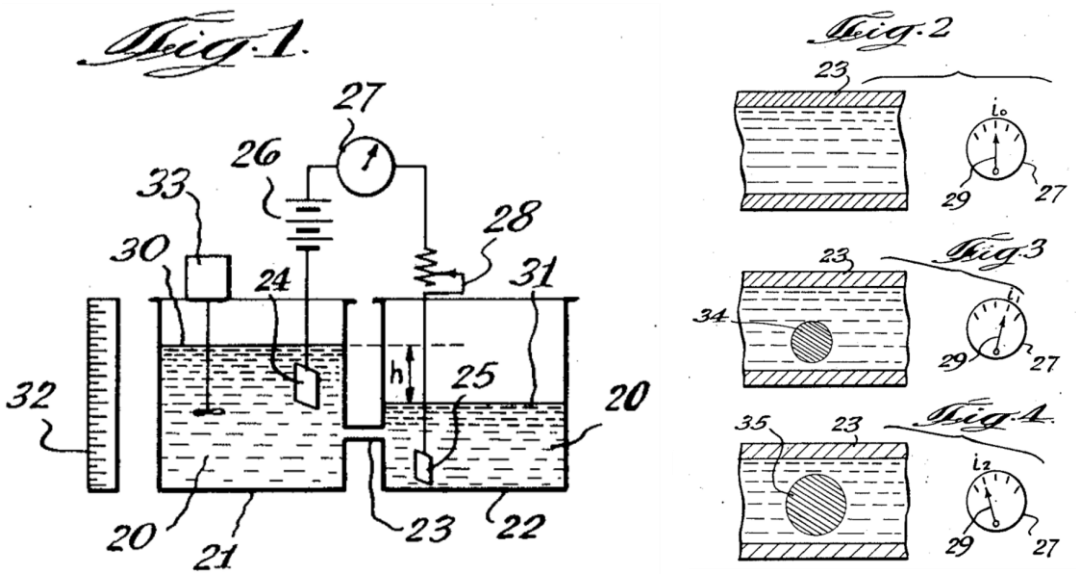


Figure 2.1: Resistive Pulse Technique as designed by Coulter

Schematic on the left (labelled as Fig. 1) shows the experimental setup with two vessels filled with conducting solution connected by a channel and a height difference in the fluid levels is maintained to let the fluid flow from left vessel to the right one. Particles to be counted are added in the left vessel and the move through the channel under flow. Schematic on the right (labelled as Fig. 2) shows the different scenarios related to the experiment. There are electrodes placed in either of the vessels that form a part of the electrical circuit. When there is no particle in the channel, the current due to ionic flow through the channel is read in the ammeter as i_0. When a particle of higher conductivity compared to medium conductivity moves through the channel, there is a momentary increase in the current shown as i_1. Similarly, the movement of a particle of lower conductivity compare to medium conductivity results in a decrease in current shown as i_2. Figure reproduced from Ref [62].

Grover *et.al.* demonstrated this technique using synthetic rubies with cylindrical holes to accurately quantify the volumes of polystyrene latex particles [102,103]. Based on the studies done by Maxwell [104] on conductivity of particle suspension, the following equation is obtained that relates the relative current change in the pore to the volume of the particle [102,103]:

$$\frac{\Delta I}{I} = 1.5\left(\frac{\delta}{1-\delta}\right) \tag{2.1}$$

for non-conducting particles when $\delta << 1$. Here, δ is the ratio of the particle volume (v) to that of the orifice (pore) volume (V). In order to measure the volume of the pore, firstly the effective pore length (L_{eff}) was measured using the following equation [103]:

$$L_{eff} = \frac{\pi d^2}{\rho G} \tag{2.2}$$

where D_P is the orifice diameter, ρ is the resistivity of the solution, G is the conductance measured across the orifice. The volume of the particle is obtained from the equation $\delta = v/V = 4v/(\pi D_P^2 L_{eff})$. This is shown in Fig 2.2 which compares the volumes obtained by resistive pulse technique to that of electron microscope.

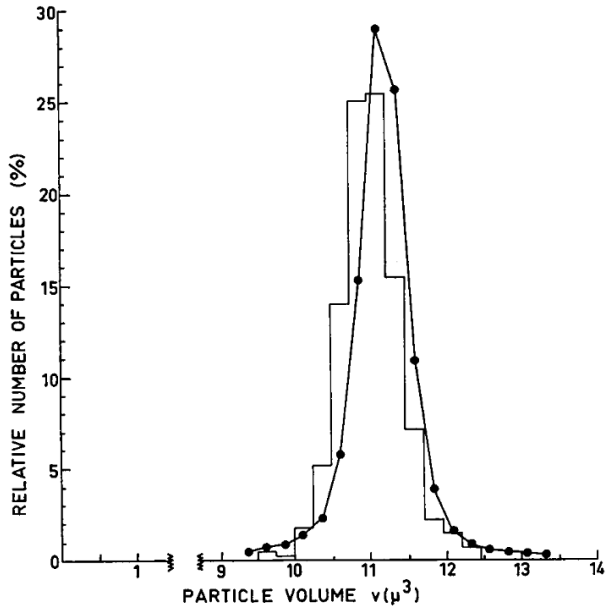


Figure 2.2: Volume quantification using resistive pulse technique.

Volume distribution of polystyrene latex particles as determined electrically (broken line) and with electron microscope (histogram). Reproduced with permission from Ref [103].

Later when these experiments were performed on RBC suspensions, it was found that there are two types of spikes – type A and type B as shown in Fig 2.3 [105]. Type A spikes remain flat throughout and type B are peaked at some point during their passage through the orifice. The amplitudes of peaked spikes are higher than the flat spikes and this can be explained by the dependence of shape factor on the orientation of cells while passing through the orifice. It was also shown by theoretical studies that the change in current depended upon the geometry of the particle such as their dimensions along their principal axis in all the three directions. The relative current change is given by the equation 2.3 [102,105]:

$$\frac{\Delta I}{I} = \gamma\left(\frac{\delta}{1-\delta}\right) \tag{2.3}$$

when $\rho_1 \gg \gamma\rho_2$.

For an oblate spheroid with principal axes a, b, b and m = a/b < 1, γ can be calculated from the following expression:

$$\frac{1}{\gamma} = \frac{m\cos^{-1} m}{(1-m^2)^{3/2}} - \frac{m^2}{1-m^2} \tag{2.4}$$

For prolate spheroid (m > 1):

$$\frac{1}{\gamma} = \frac{m^2}{m^2 - 1} - \frac{m \cosh^{-1} m}{(m^2 - 1)^{3/2}} \tag{2.5}$$

The above equations are valid when the a-axis is along the direction of electric field, otherwise, γ must be replaced by

$$\gamma' = \frac{2\gamma}{2\gamma - 1} \tag{2.6}$$

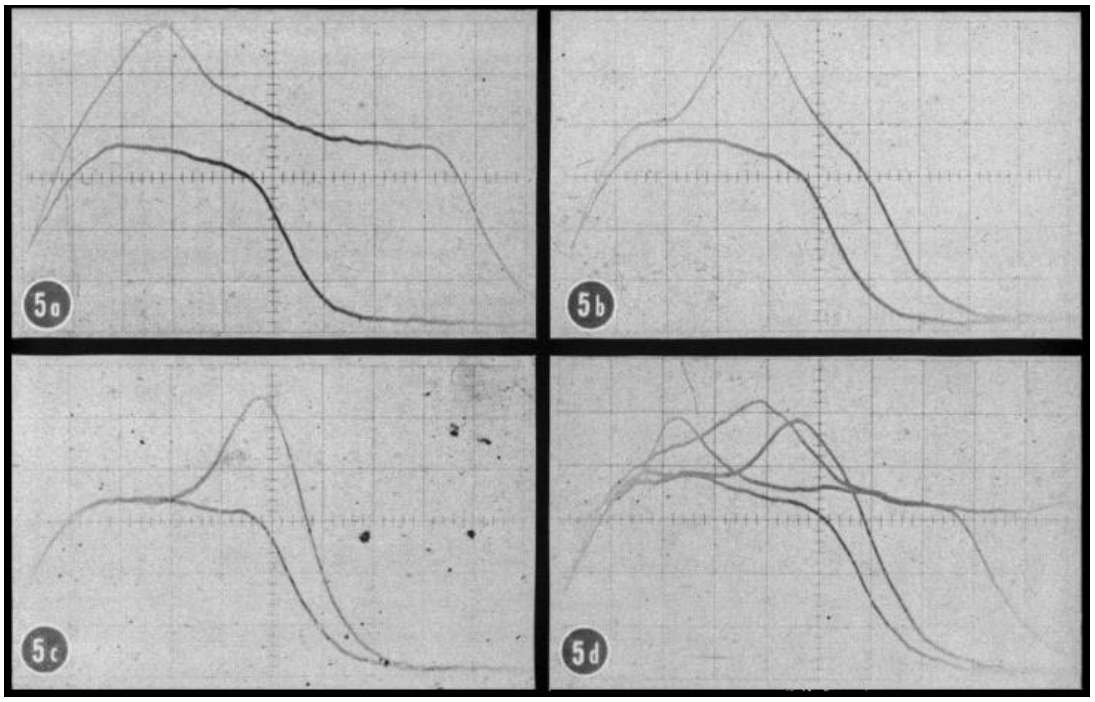


Figure 2.3: Resistive pulse analysis of RBC (spheroidal particles)

Oscillograms of pulse shapes of human red blood cells in an orifice 50 µm x 150 µm. Flow rate, 450 cm/sec; current, 170 µamps; sweep 50 µsec/division. Reproduced with permission from Ref [105].

2. Sub-micron sized pores to detect nanoparticles:

These studies which were initially applied to study micron sized particles was later extended to characterize polystyrene nanospheres of 90 nm in size using pores etched in irradiated plastic sheet by Deblois-Bean [106]. Here, the particles were driven only the electric field and not by pressure difference. Consider a tube of effective length L_{eff} and diameter D_P filled with solution of resistivity ρ, the change in resistance of the solution in tube due to the presence of a spherical particle of diameter d when $d << D_P << L_{eff}$ which also implies $\delta << 1$ is calculated as follows from eqn (2.1):

$$\frac{\Delta I}{I} = 1.5\left(\frac{\delta}{1 - \delta}\right) \approx 1.5\delta \tag{2.7}$$

$$R = \frac{V}{I} \Rightarrow \Delta R = -\frac{V}{I^2}\Delta I = -\left(\frac{\Delta I}{I}\right) R \tag{2.8}$$

$$\Delta R = -(1.5\delta)R = -\frac{3}{2}\left(\frac{\frac{\pi d^3}{6}}{\frac{\pi D_P^2 L_{eff}}{4}}\right)\left(\frac{4\rho L_{eff}}{\pi D_P^2}\right) = -\frac{4\rho d^3}{\pi D_P^4} \quad (2.9)$$

$$|\Delta R| = \frac{4\rho d^3}{\pi D_P^4} \quad (2.10)$$

The decrease in current is seen as increase in resistance while a particle moves through the pore. As the eqn (2.10) is based on Maxwell's results for the case of a suspension of small spheres in a medium with spherical boundaries of diameter much bigger diameter compared to the suspended spheres, a correction term was obtained to make the equation valid for a large range of d/D_P values, giving the equation

$$|\Delta R| = \frac{4\rho d^3}{\pi D_P^4} F\left(\frac{d^3}{D_P^3}\right) \quad (2.11)$$

Here, $F(d^3/D_P^3)$ is close to 1 for $d << D_P$.

3. Nanometer sized pores to detect biomolecules:

The fabrication of nanometer size pores is done by drilling silicon oxide or silicon nitride membranes using TEM [107] or dielectric breakdown [108]. The thickness of the membrane decides the pore length (L) and the pore diameter (d) is controlled during the fabrication process. Considering the nanopore to be a cylinder, its conductance in an electrolyte solution with conductivity (σ) or inverse of resistivity (ρ) is given by:

$$G_{pore} = \frac{\sigma \pi D_P^2}{4L} \quad (2.12)$$

The problem of edge effects was previously theoretically treated by Hall and termed the resistance offered by a flat disc or pore mouth to the flow of ions from the reservoirs into the pore to be access resistance [109]

$$R_{access} = \frac{1}{2\sigma D_P} \quad (2.13)$$

The access resistance is in series with the bulk resistance of the pore, and these resistances add up (conversely, the reciprocals of the conductances add up) as shown below [110]:

$$G(D_P) = \left[\frac{1}{G_{pore}} + \frac{1}{G_{access}}\right]^{-1} = \sigma\left[\frac{4L}{\pi D_P^2} + \frac{1}{D_P}\right]^{-1} \tag{2.14}$$

The importance of access resistance in the explanation of results obtained using nanopores was demonstrated as shown in Fig 2.4A where the conductance of the pore as a function of its diameter is plotted and fitted with eqn (2.14) to determine its effective length.

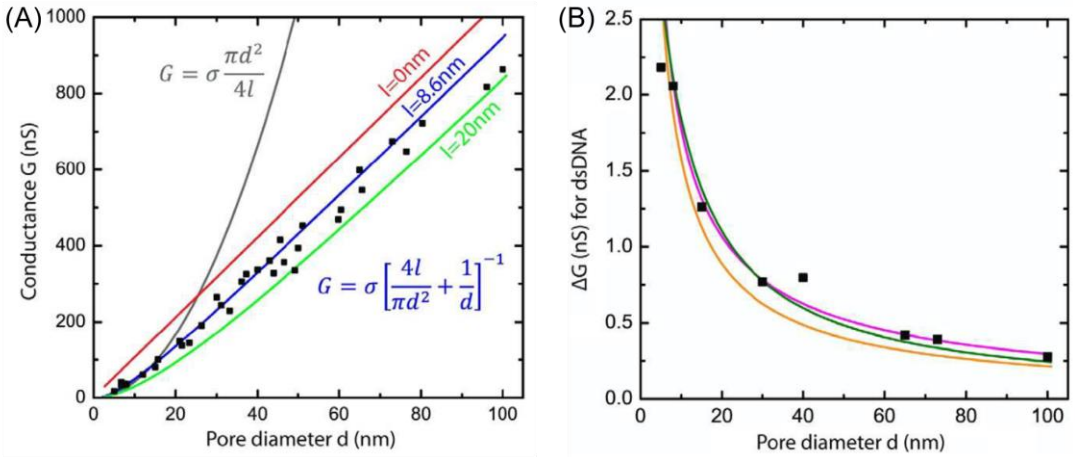


Figure 2.4: Plot of Conductance vs pore diameter for planar nanopores.

(A) Experimental data for conductance versus pore diameter of pores drilled into a 20 nm thin silicon nitride (SiN) membrane. The solid lines are fits to equation (1) (gray line, l = 20 nm) and equation (2) (red, l = 0; blue, l = 8.6 nm; green, l = 20 nm). (B) Experimental data for the change in pore conductance ΔG upon insertion of a dsDNA molecule, as a function of pore diameter. The solid limes are for a simple geometric model with cylindrical DNA in a cylindrical pore (green), and in a hyperboloid-shaped pore in two limits (purple and orange). Reproduced with permission from Ref [110].

The translocation of DNA is performed in these nanopore and it was observed that the conductance drop (ΔG) decreases with the pore diameter (Fig 2.4B). This observation is explained by the following model which takes the access resistance into account [110]:

$$\Delta G = G - G_{with\ DNA} = G(D_P) - G(D_{with\ DNA}) \tag{2.15}$$

where

$$D_{with\ DNA} = \sqrt{D_P^2 - D_{DNA}^2} \tag{2.16}$$

Here D_{DNA} is the diameter of DNA which is considered to be 2.2 nm.

4. Characterization of proteins using nanopores:

It was demonstrated that nanopores could be used to determine the properties of proteins, such as, their shapes and sizes. The following equations were used to quantify the volumes of proteins [101,111–114]:

$$\frac{\Delta I}{I} = \gamma\left(\frac{\delta}{1-\delta}\right) \approx \gamma\delta \ (when\ \delta \ll 1) \tag{2.17}$$

This equation takes the form [101,111–114]

$$\Delta I = \frac{\gamma V_A V}{\rho l_{eff}^2} \text{ for } l_M < l_{eff} \tag{2.18}$$

$$\Delta I = \frac{\gamma V_A A_X}{\rho l_{eff}} \text{ for } l_M > l_{eff} \tag{2.19}$$

Here, V is the excluded volume, A_X is the cross-sectional area of aggregates with lengths longer than the pore effective length, γ is the shape factor, V_A is the applied voltage, ρ is the resistivity of the electrolyte, l_P is the pore length, r_P is the radius of nanopore, l_M is the length of the aggregate, effective length of the pore is $l_{eff} = l_P + 1.6r_P$. The volumes of a number of proteins were measured and the values agree well with the estimations made from their crystal structures.

5. Glass nanocapillaries for nanoparticle detection:

Resistive pulse technique can also be implemented using conical nanocapillaries fabricated from glass capillaries which are typically of millimeter diameter and tapered down to a few tens of nanometers [68,69,115,116]. This is achieved using Sutter puller which melts the center of the capillary immobilized at two ends followed by pulling it apart. This generates two nanocapillaries and the pore size can be controlled by tuning the parameters of heat and pull. The capillaries can be further fine-tuned to desired tip sizes by shrinking them using SEM (Fig 2.5) [67,69,115].

In its simplest form, glass nanopores are modelled as pores of simple cylindrical geometry with an effective length. In last decade, multiple models have emerged to take into account the shape of nanopores in the glass nanocapillaries. The nanopore can be modelled as a single cone as shown in Fig 2.6A with the pore diameter D_P, taper length

L_T and the diameter of the capillary at the end of the taper as D_T. The conductance of the nanopore is then written as [67]:

$$G = \sigma\left(\frac{4L_T}{\pi D_p D_T} + \frac{1}{2D_p} + \frac{1}{2D_T}\right)^{-1} \tag{2.20}$$

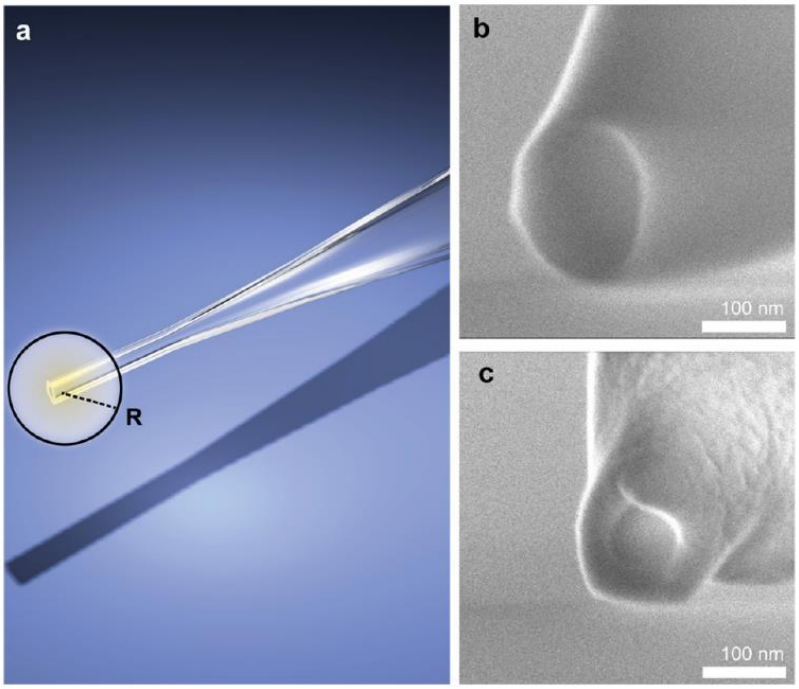


Figure 2.5: Shrinking and characterization of glass nanopores.

(a) Scheme of the conical end of the nanocapillary. The shaded area depicts the region, which is imaged by the SEM beam. The radius R is the penetration depth of the electron beam calculated by the theoretical penetration formula given by Kanaya−Okayama. (b) SEM in-lens image of a quartz nanocapillary magnified 196k times at a stage angle of 60 degrees to increase the three-dimensional perception. The electron high tension was at 3.0 kV, the beam current was at 171 pA, and the working distance was 3.3 mm. (c) Shrunken nanocapillary after 14 min of irradiation under constant angle and beam parameters. Clearly a reshaping of the nanocapillary is visible, precluding the possibility of significant carbon deposition. Reproduced with permission from Ref [115].

This model works well for unshrunk capillaries as shown in dotted black line without any fitting parameters (Fig 2.7A). As the nanopore is gradually tapered like a hour shaped glass, it would be more accurate to model it as two cones instead of a single cone. This requires extension of the single cone model to double cone model (Fig 2.6B) with two taper lengths (L_{cond} and L_T) and two 'inner' diameters (D_P and D_{cond}):

$$G(D_p) = \sigma\left(\left(\frac{4L_{cond}}{\pi D_P D_{cond}} + \frac{1}{2D_P}\right) + \left(\frac{4L_T}{\pi D_{cond} D_T} + \frac{1}{2D_T}\right)\right)^{-1} \tag{2.21}$$

By fitting this equation to the data, the parameters L_{cond} and D_{cond} were obtained as 543 nm and 514 nm respectively (Fig 2.7B). For the case of DNA translocation, it is assumed

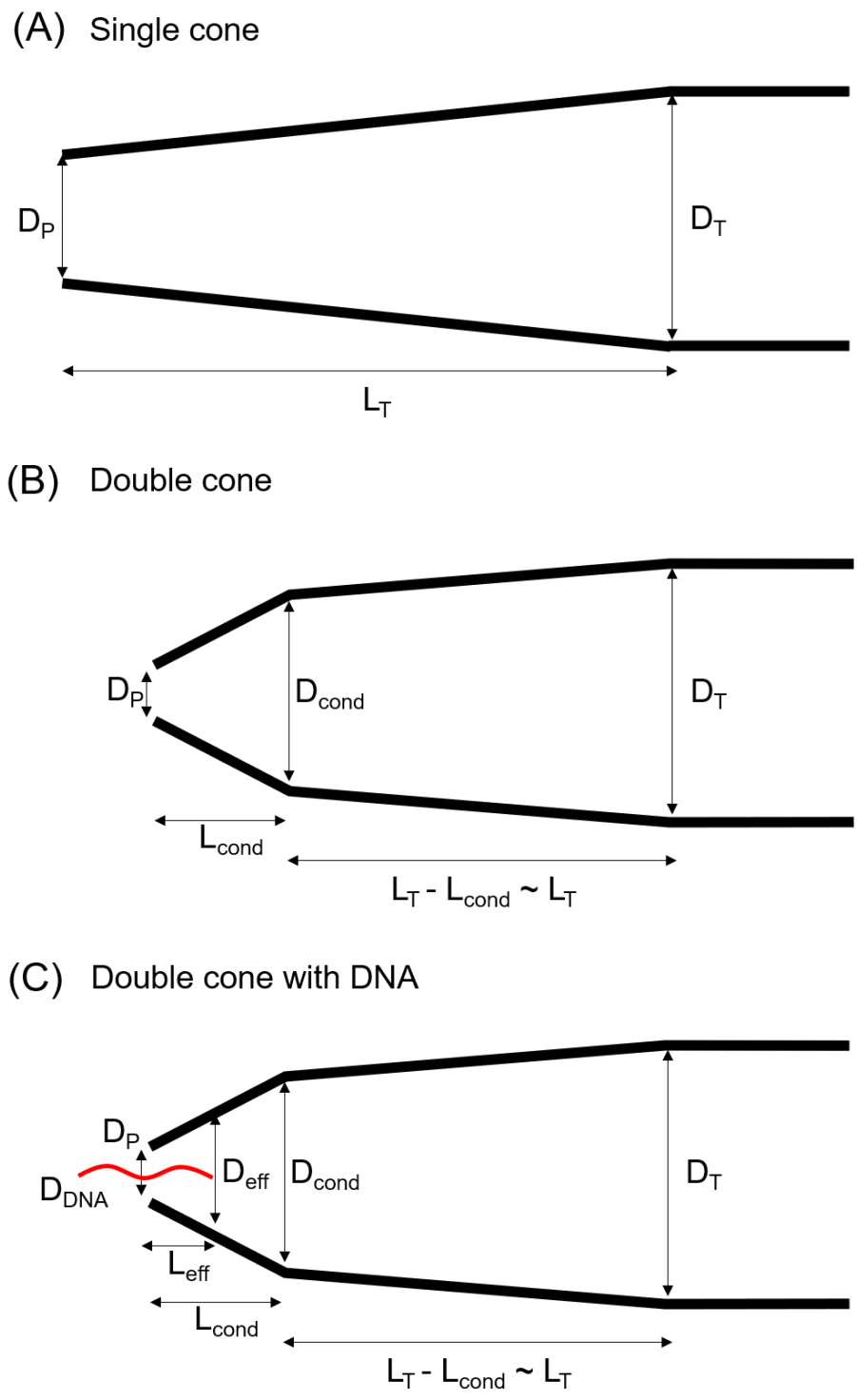


Figure 2.6: Geometry of glass nanopores

Glass nanopore is modelled as a single cone (A) or a double cone (B). (C) The DNA is assumed to be detectable only within a small region of the cone near the pore upto a length L_{eff}.

that DNA is detected only in a region L_{eff} of the first cone (Fig 2.6C) and the following equations are obtained:

$$G = \sigma\left(\left(\frac{4L_{eff}}{\pi D_P D_{eff}} + \frac{1}{2D_P}\right) + \left(\frac{4(L_{cond} - L_{eff})}{\pi D_{eff} D_{cond}}\right) + \frac{1}{B}\right)^{-1} \quad (2.22)$$

$$G_{with\ DNA} = \sigma\left(\left(\frac{4L_{eff}}{\pi D_{with\ DNA} D_{eff}} + \frac{1}{2D_{with\ DNA}}\right) + \left(\frac{4(L_{cond} - L_{eff})}{\pi D_{eff} D_{cond}}\right) + \frac{1}{B}\right)^{-1} \quad (2.23)$$

where $D_a = D_P + \frac{L_{eff}(D_{cond} - D_P)}{L_{cond}}$ and $\frac{1}{B} = \left(\frac{4L_T}{\pi D_{cond} D_T} + \frac{1}{2D_T}\right)$.

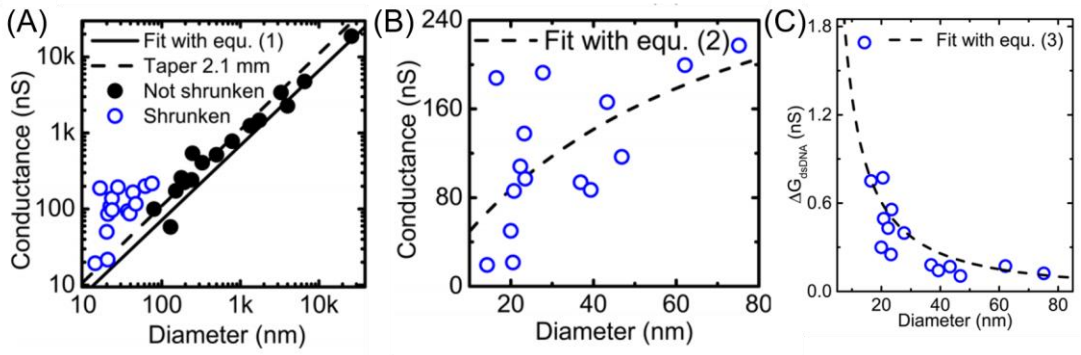


Figure 2.7: Plot of conductance vs pore diameter for glass nanocapillaries.

(A) Conductance as a function of the diameter of shrunken (hollow circles) and unmodified (filled circles) nanocapillaries. The continuous line is a fit with eq 2.20, leaving the taper length t as the fitting parameter. The dashed line is the result of eq 2.20 using the actual taper length of 2.1 mm. (B) Zoom of (A), focusing on the conductance dependence from the diameter of the shrunken nanocapillaries. The dashed line is a fit with eq 2.21. (C) Conductance drop due to DNA as a function of pore diameter is shown with the data fit to eqn (2.15) written for double cone model.Reproduced with permissison from Ref [67].

The conductance drop due to DNA translocation is given by eqns (2.15) and (2.16) along with eqns (2.22) and (2.23) [67]. By fitting this equation to the data as shown in Fig 2.7C, the value of sensing length (L_{eff}) was obtained to be 32 nm. This gives an idea about the geometry of the sensing region.

Here, we present the summary of various models that are used for explaining the conductance and conductance blockade in the nanopore experiments.

Grover (ellipsoidal particles in cylindrical pores) [102,105]	$\frac{\Delta I}{I} = \gamma\left(\frac{\delta}{1-\delta}\right)$
Deblois-Bean (spherical particles in cylindrical pores) [106]	$\lvert\Delta R\rvert = \frac{4\rho d^3}{\pi D_P^4}$
Yusko, et.al.(ellipsoidal particles in cylindrical pores) [101,111–114]	$\Delta I = \frac{\gamma V_A \mathrm{V}}{\rho l_{eff}^2}\ for\ l_M < l_{eff}$ $\Delta I = \frac{\gamma V_A \mathrm{A_X}}{\rho l_{eff}}\ for\ l_M > l_{eff}$

Cees Dekker (DNA in cylindrical pore) [110]	$G(D_P) = \sigma\left[\frac{4L}{\pi D_P^2} + \frac{1}{D_P}\right]^{-1}$ $\Delta G = G(D_P) - G(D_{with\ DNA})$ $D_{with\ DNA} = \sqrt{D_P^2 - D_{DNA}^2}$
Radenovic (DNA in double conical pore) [67]	$G(D_p) = \sigma\left(\left(\frac{4L_{cond}}{\pi D_P D_{cond}} + \frac{1}{2D_P}\right) + \left(\frac{4L_T}{\pi D_{cond} D_T} + \frac{1}{2D_T}\right)\right)^{-1}$ $\Delta G = G(D_P) - G(D_{with\ DNA})$

Table 2.1: Summary of nanopore models for conductance blockade measurements.

2.1.2 Numerical modelling of nanopores

We show single cone model (Fig 2.8A) for the characterization of nanopores, this provides clarity on the regions of nanopores that are crucial in modelling of conductance and conductance blockade. Figure 2.8B plotted for a 20 nm pore and 4 M LiCl with varying entry cone angles. Here, the diameter of the other end of the cone D_T is a function of L_T given by $D_T = D_P + 2L_T\tan\theta$. The nanopore with the parameters defined are shown in a schematic in Fig 2.8A. We plot the resistance as a function of the capillary taper length using eqn (2.20) in Fig 2.8B. The resistance increases as we go along the taper length initially but saturates after a certain point. This is because all the electric potential is dropped across the nanopore within this length. We can define the effective length (L_{eff}) as the point along the taper where the resistance is about 90% of the total capillary resistance. We consider the total capillary length to be 10 μm for the calculations. The dependence of L_{eff} on the parameters such as half-cone angle (θ) and pore diameter can be obtained as follows.

$$R\left(L_{eff}\right) = 0.90R(L_T) \tag{2.24}$$

$$\frac{4\left(L_{eff} + \frac{\pi}{8}\left(D_P + D_{eff}\right)\right)}{\sigma D_P D_{eff}} = 0.90 \times \frac{4\left(L_T + \frac{\pi}{8}(D_P + D_T)\right)}{\sigma D_P D_T} \tag{2.25}$$

At large value of L_T, $D_P + D_T \approx D_T = 2L_T tan\theta$. This gives us

$$L_{eff} = \left(\frac{C - \frac{\pi}{4}}{1 + \frac{\pi}{4} tan\theta - 2C tan\theta} \right) D_P = k(\theta) D_P \tag{2.26}$$

where $C = 0.45(1 + \pi/4(tan\theta))/tan\theta$.

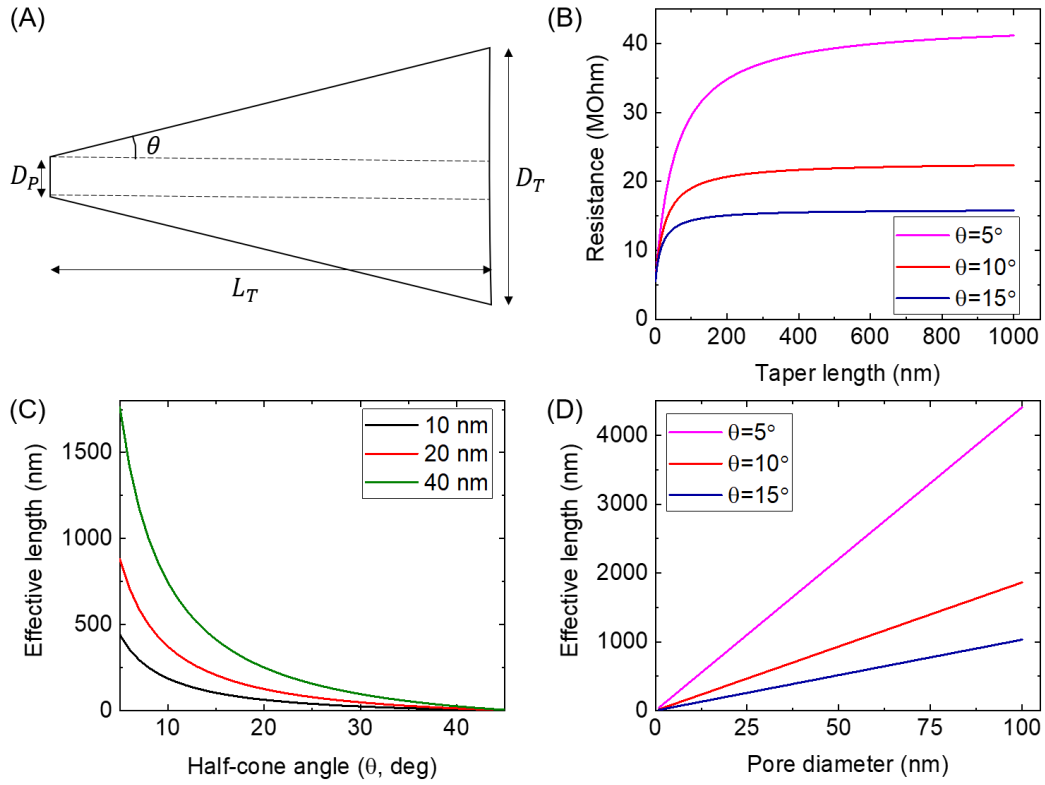


Figure 2.8: Numerical analysis of sensing length.

(A) Schematic of a conical nanopore with pore diameter (D_P), taper length (L_T) and diameter at the end of the taper (D_T) with a half-cone angle of θ. (B) shows pore resistance as a function of the taper length considered for its estimation for various cone angles. (C) shows the dependence of effective pore length on the cone angles for various pore diameters. (D) shows the effective length of the pore as a function of pore diameter for various cone angles.

In Fig 2.8B, for a half-cone angle of 15°, we find that the sensing length is 206 nm. Similar estimates are performed for various cone angles. In Fig 2.8C, sensing length is plotted as a function of half- cone angle. The sensing length sharply decreases from 206 nm to 48 nm as the cone angle increases from 15° to 30° and it reaches 2 nm at 45°. We also analytically find that the effective pore length is linearly dependent on the pore diameter with the constant of proportionality being a function of half-cone angle of the capillary. We plotted the effective length as a function of pore diameter for various cone angles in Fig 2.8D. This concludes that the length of whole taper which is around 3 mm clearly is not required for modelling the nanopore observations as the nanopore senses the changes in the ion flow only with-in the effective length of the pore. The cone angle

is one of the important parameters for modelling resistance of nanopore. The effective length would provide us the limitation on the spatial resolution of our system. The Fig 2.8C shows the possibility of obtaining short sensing lengths that are comparable to planar nanopores. In this thesis, we mostly used the shrunk nanopores which are difficult to accurately model because the internal profiles of those capillaries are unknown. Later part of the thesis implements an ECD-based technique to accurately profile the capillary geometry allowing us to model the DNA blockade in the nanopore.

2.1.3 Fabrication of glass nanopores

To carry out the experiments, the nanopores were fabricated using quartz capillaries with filament (OD = 1 mm and ID = 0.5 mm: QF-100-50-7.5, Sutter Instruments). We chose filamented glass capillaries over unfilamented ones as that would increase the capillary action to let the capillary fill much faster and without air bubbles. Filament is a glass rod of 0.16 mm diameter attached to the internal wall of the capillary. After cleaning with water (MilliQ, Millipore) and ethanol and then nitrogen dried, capillaries were pulled using Sutter Puller P-2000G with the following two-line program. Line 1: Heat (810), Filament (3), Velocity (35), Delay (145), Pull (75), Line 2: Heat (660), Filament (0), Velocity (15), Delay (128), Pull (200). We have consistently achieved pore diameters in the range of 40-80 nm using the above program. Note that the pore diameters are affected by local conditions such as puller condition, cleanliness of capillaries, humidity, temperature, room vibrations and laser stability. These nanopores on capillary tips were further shrunk to the desired size using a 3 keV beam of a scanning electron microscope, SEM (Carl Zeiss Ultraplus FESEM) as described previously [69]. Pulled and imaged nanopores are stored in desiccated air-tight dust-free containers until further use. We found that nanopores stored up to two months give good low-noise data.

2.1.4 Experimental preparation

The nanopores were cleaned in oxygen plasma for 1 min at 50 W of RF power and 250 mtorr of Oxygen pressure using PlasmaPrep-III system (SPI Supplies). The capillary was then glued into the Teflon fluid cell as shown in Fig 2.9A (right) and filled with the

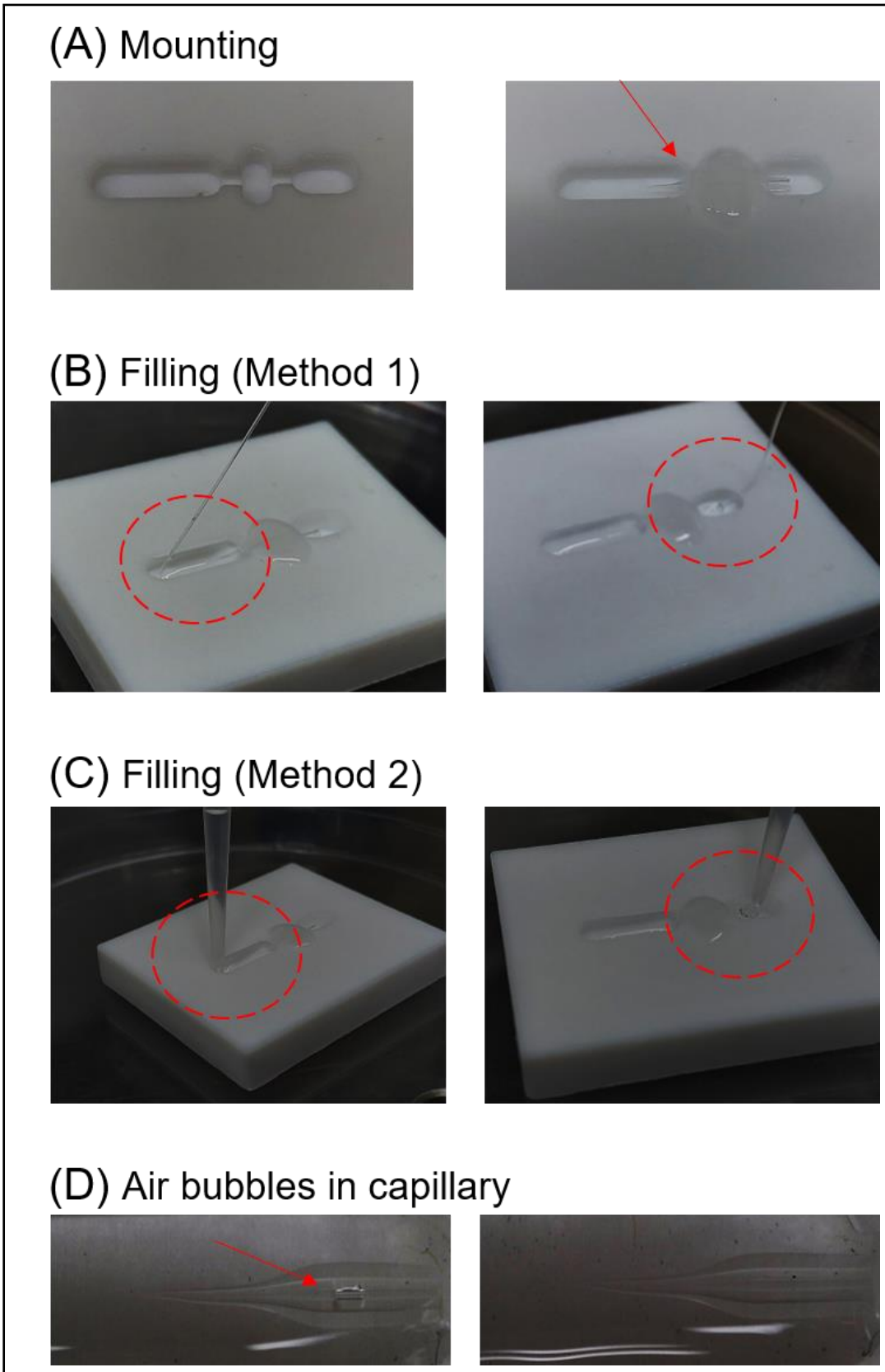


Figure 2.9: Procedure to prepare the capillary for experiments

The snapshots of the mounting (A), filling methods (B & C) and capillary with air bubbles (D) are shown. More details are given in the text.

Nanopore Buffer (NPB) (4M LiCl in 10 mM Tris-HCl, 1 mM EDTA at pH 8). The filling is done either by using the tapered plastic tip made out of Eppendorf pipette tips (Fig 2.9B) or by using the pipette tip as it is (Fig 2.9C). Firstly, the well where the nanopore is located is filled with buffer and wait for 5 min and let the buffer fill into the nanopore region by capillary action. Then add buffer at the other end of the capillary. Observe if any air bubbles are visible as pointed by red arrow in Fig 2.9D (left). To remove these air bubbles present inside the capillary, the fluid cell was kept in a vacuum desiccator for 15 min. Now the air bubbles should disappear as shown in Fig 2.9D (right). A small piece (1 inch) of pure silver wire is taken and one end of it is dipped into Clorox bleach solution for about 12 hrs to chlorinate it to become AgCl. Another method is to connect it to positive terminal and the other terminal is connected to Platinum wire and a voltage of 1.5 V is applied for 1.5 min. This would also coat the one end of the silver wire with AgCl. These wires could be used as electrodes for nanopore current measurements.

2.1.5 Recording measurements

The Ag/AgCl electrodes and Axopatch 200B amplifier (Molecular Devices) were used to apply a voltage and measure the current through the nanopore. The complete experimental setup is shown in Fig 2.10. In order to reduce the noise in electrical measurements, the nanopore setup was isolated from the external environment by placing it in a copper Faraday cage. As the amplifier has a saturation limit of ± 1000 mV and a current of ± 200 nA, we chose the experimental conditions in such a way that the applied voltage and measured current are well within those limits. The total current gain due to the amplifier is given by the product of two gain factors, voltage gain (alpha of 0.5 mV/mV) and headstage gain (beta of 0.1 mV/pA). The amplifier also provides an output to measure the applied voltage across the resistor.

Data acquisition:

The data was acquired using NI DAQ PCI-6251 (National Instruments) at a sampling frequency of 200 kHz using a custom-written LabVIEW code (recordtrace.vi) (screenshot of the software is shown in Fig 2.11). In order to optimize the performance

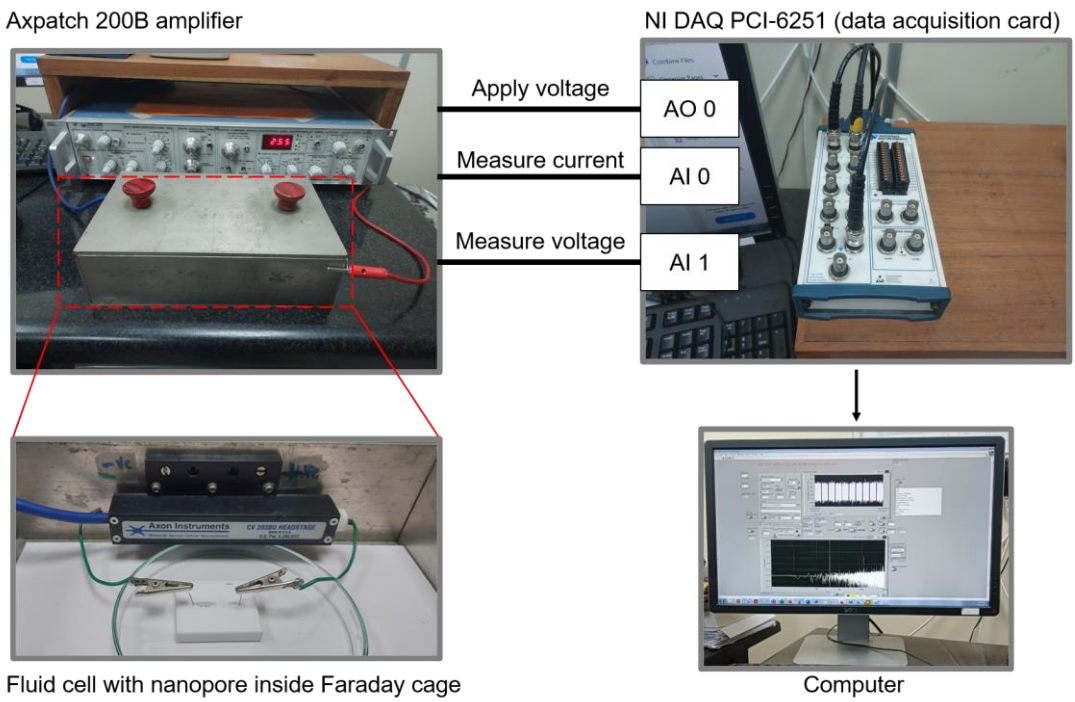


Figure 2.10: Experimental setup to record nanopore measurements

Various components of the experimental setup such as nanopore with flow cell, faraday cage, amplifier, DAQ card and computer required for the nanopore measurements are shown. The labels AO0. AI 0 and AI 1 represent the channels on DAQ for applying voltage, measuring current and voltage across the resistor respectively.

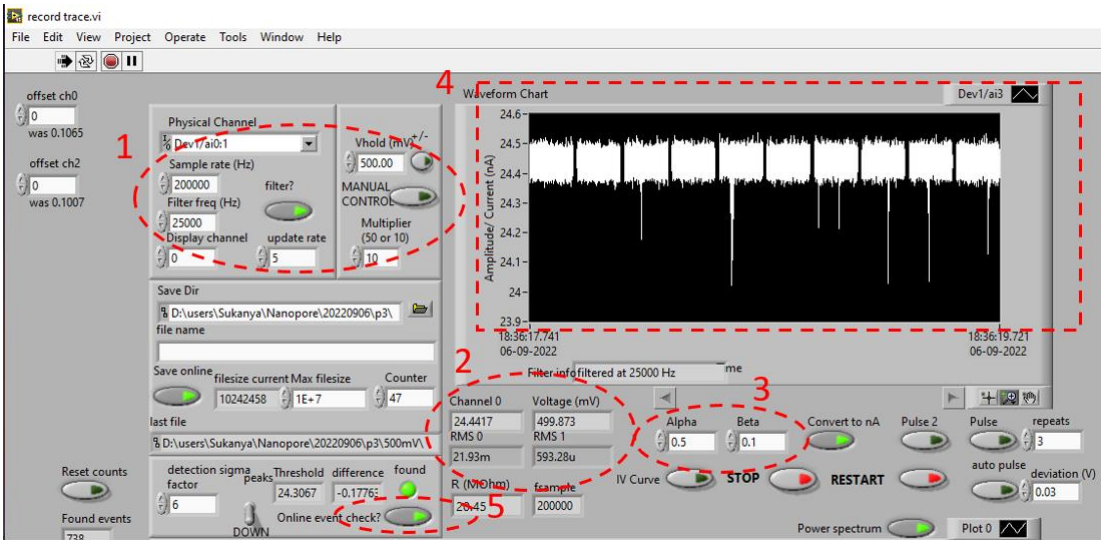


Figure 2.11: LabVIEW interface for recording current measurements.

The red regions in the picture are showing important information in the interface. Region 1 shows the parameters such as the Sample rate (Hz), filter frequency (Hz), applied voltage, etc. Region 2 shows the mean and standard deviation values corresponding to the measured current (nA) and the measured voltage (mV) across the resistor. Region 3 shows the amplifier gain values Alpha (α) and Beta (β). Region 4 shows the live recording of the current trace. Region 5 shows the button for detecting the events as they are recorded.

of the acquisition process, the data is recorded as buffers of 40 k points per every second.

The IV curves are recorded by a sub vi (get IV curve.vi) in the same code. For this purpose, the current is recorded at a sampling frequency of 45 kHz (DAQ Scanrate/s). A total of 5 k datapoints are recorded at every voltage. Once the voltage is changed, there is a waiting time of 0.5 sec before recording the current. This is done to ensure that the capacitive spikes that follow a sudden change in voltage do not interfere with the resistance measurements.

The data is saved in the binary file format with a maximum file size of 10 MB which contains 640 k points in current trace. To minimize the disk space usage, there is a button called 'Online event check?' integrated to detect the translocation events. This button can be turned ON to write only the buffers containing the events and to discard those without any events. The data can be low-pass filtered to see the events clearly by increasing the signal to noise ratio. Blockage of nanopores by dirt or junk can be seen as step-like or gradual drop in the open pore baseline current. Whenever there is a blockage of nanopore with some particles, the voltage is switched rapidly between opposite polarities or the ZAP button on the amplifier is pressed repeatedly so that the particle is removed from the nanopore walls and the experiment can be continued. The sticking of analyte molecules to the charged glass surface during translocation experiments is quite common and becomes an obstacle in performing long-term and fully automated experiments.

2.1.6 Characterization of nanopores

In Fig 2.12 we show characterization of our quartz nanopores for single molecule detection experiments. All the custom-made nanopores are characterized by optical microscope and SEM imaging to measure the taper length and tip geometry of the pore respectively (Fig 2.12A). Fig 2.12B shows representative SEM images of nanopores ranging from 13 to 93 nm diameter. The quartz nanopore is mounted on the fluid cell and filled with nanopore buffer (Fig 2.12C inset on top-left) and checked for linearity in its I-V characteristics in the range of ± 500 mV. We observed linear IV curves for all our nanopores as shown in Fig 2.12C. The conductance G (nS) for each pore is calculated from the linear fits to their respective IV curves and is plotted for 37 different nanopores of varying diameter in Fig 2.12C (inset on bottom-right). The open pore conductance of nanopores is modelled very well with the previously established double-cone model.

Here, the first cone is the tapered portion of the capillary tube with a cone length as the taper length, L_T, and base diameter, D_T, as the I.D. of the quartz capillary and the second cone is the conductance cone near the tip of the capillary with a cone length and base diameter of L_{cond} and D_{cond} respectively. Nanopore conductance, G (nS), for a pore of

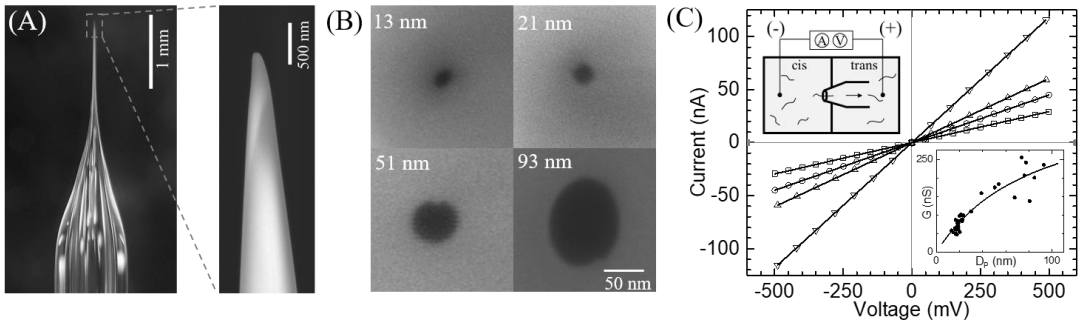


Figure 2.12: Nanopore characterization

(A) Left: The side-on optical image of a typical quartz capillary pulled with the laser puller. The average taper length of these capillaries was 2.8 ± 0.3 mm. Right: Zoomed image of the tip of the nanocapillary using an SEM microscope. (B) The representative SEM images of nanopores at the tip of the nanocapillary, with sizes ranging from 13 nm to 93 nm, used in this study. (C) I-V curves of representative nanopores (13 nm – squares, 21 nm – circles, 51 nm – up triangles and 93 nm – down triangles) showing linear characteristics are plotted. The top-left inset shows a schematic of the Teflon fluid cell mounted with a nanocapillary. The bottom-right inset shows pore conductance as a function of pore diameter plotted for N = 37 different nanopores. The solid line is the fit to eqn (2.21).

diameter D_P, in the double-cone model is given by eqn (2.21) [67]. Here, σ is the conductivity of buffer (NPB (4 M LiCl in 10 mM Tris, 1 mM EDTA pH 8), σ = 18.0 S/m), as measured by a conductivity meter. The experimental data was fitted by the equation (2.21) giving fit parameters L_{cond} and D_{cond} as 418 ± 99 nm and 176 ± 25 nm respectively. This signifies the region where the major fraction of resistance of the conical nanopore falls. Given the spread in the data around the fit curve, we note that the internal geometry of each custom-made pore is slightly different and plays a crucial role in determining the true value of nanopore conductance.

2.1.7 Translocation event analysis

In order to analyze a sample, it is added into the cis-chamber of the fluid cell. The translocation of the analyte molecules through the nanopore results in the reduction of current seen as downward spikes in current trace (Fig 2.13). We record about a few hundreds to thousands of events in a typical experiment. These events are then analysed using a custom-written LabVIEW code to extract the features about the analyte molecules as shown in Fig 2.14.

The portions of current trace which are due to pore blockage are excluded from the analysis. When the low-pass filter is applied, a total of sampling frequency/10 number of points are deleted from the beginning part of the file as it has signal processing artifacts.

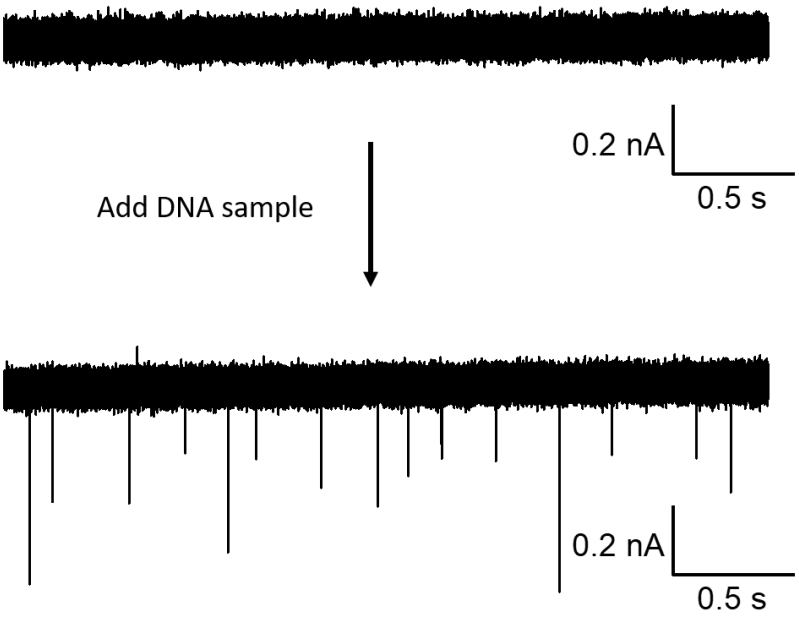


Figure 2.13: Detection of events upon addition of sample.

There are downward spikes (events) in the current trace due to the translocation of analyte molecules through the pore.

The current trace is divided into smaller portions whose length is defined by the patch size. The baseline or mean (μ, blue) and standard deviation (σ) of the patch is calculated for all the patches and a detection threshold (green) of σ times the peak detection factor (PDF, given as a user-input) is defined. In step 1, in each patch, the locations where the trace crosses the peak detection threshold are identified. In step 2, The patch is separated into two traces at the first location index. In the left portion, the trace is reversed and the location where the event crosses the rise/fall detection threshold (μ - 0.5σ) is found and it is called the start point of the event. In the right portion, the location where the event crosses the rise/fall detection threshold is found and it is called the end point of the event. The datapoints beginning from the start point till the end point comprises the event trace. In step 3, we calculate the event characteristics. The distance between the start and end points of the event is called the dwell time (Δt). The vertical difference between the baseline and the minimum value of the event trace is called as the amplitude drop (current drop (ΔI) or conductance drop (ΔG)). A set of points (padding) which are before and after the event trace are included to visualize the event relative to the baseline. In step 4,

the events are identified in subsequent patches in the file and the process is repeated across all the files. Once all the events are detected in a given dataset, a summary file

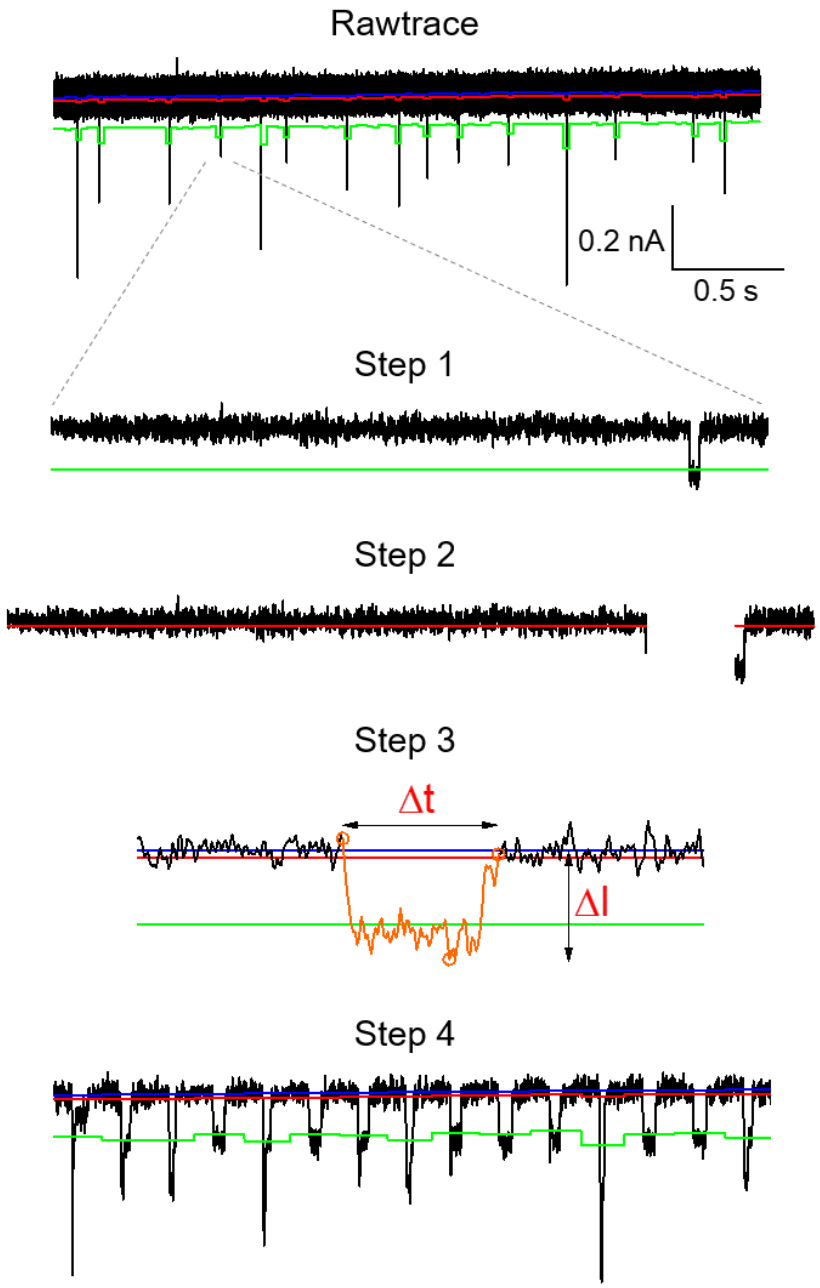


Figure 2.14: Algorithm for event analysis.

The current rawtrace (black) from a file is shown with its baseline (blue, μ), peak detection threshold (green, μ - 5σ) and rise/fall detection threshold (red, μ - 0.5σ). In step 1, the file is divided into smaller patches and the event location is detected. In step 2, the event is split into two arrays at the point where it crossed the peak detection threshold. The start and end points of the event are detected. In step 3, the event is extracted and its length (Δt) and height (ΔI) are measured. In step 4, all the detected events are concatenated.

with event characteristics is generated. The LabVIEW interface for the implementation of this procedure is shown in Fig 2.15.

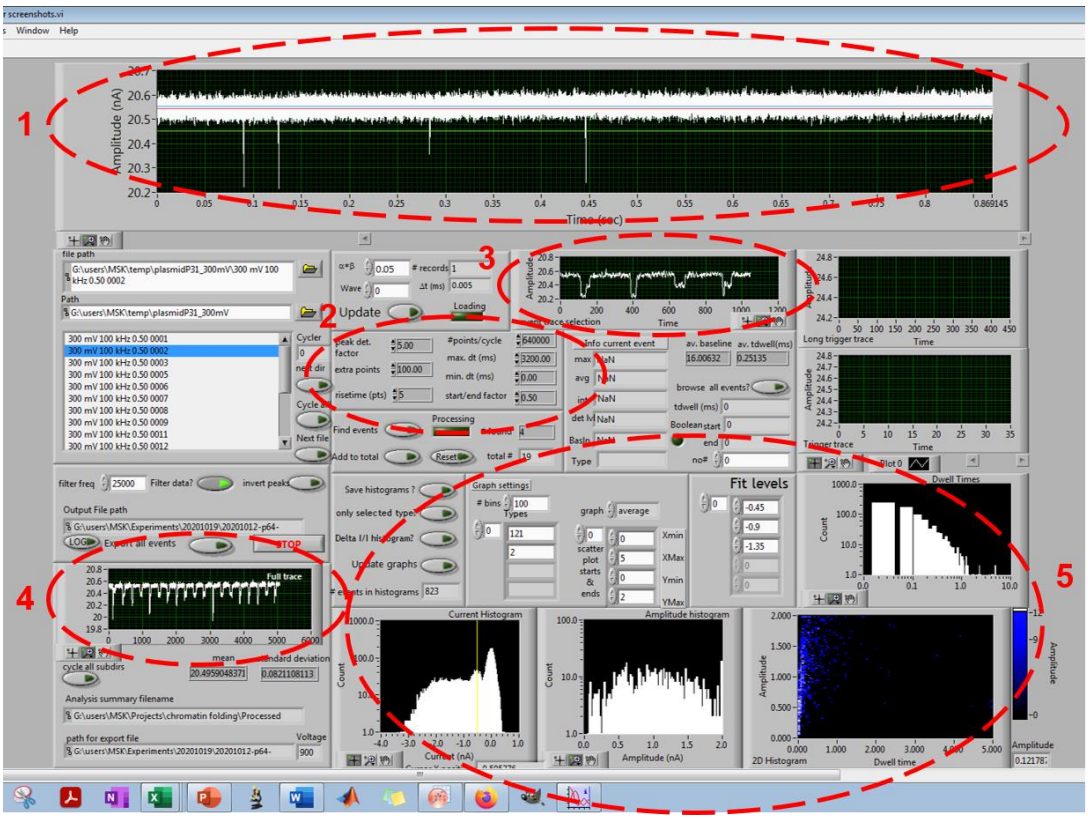


Figure 2.15: LabVIEW interface for translocation event analysis.

The important regions of the interface are marked by red circles. Region 1 shows the current trace saved in a file. Region 2 shows the important input parameters required for event detection. Region 3 shows the concatenation of events extracted from the file. Region 4 shows the concatenation of events extracted from all the files in the dataset. Region 5 shows the plots representing the statistics of the event characteristics.

2.1.8 Translocation of λ-DNA

We demonstrate the capability of our quartz nanopores in detecting multiple DNA conformations using a linearized 48.5 kb long λ-DNA translocating through a nanopore of diameter 20 nm at 500 mV bias voltage. In Fig 2.16A representative events corresponding to different levels of folding in the DNA are shown where ΔG and Δt represent the conductance drop and dwell time of the event, respectively. The conductance drop increases in well-defined steps (dotted lines) and the dwell time decreases with the increase in the number of folds [64,117]. These multi-level events correspond to multiply-folded DNA conformations as they translocate through the nanopore. Possible conformations that can result in the events shown are drawn schematically above the corresponding event. In Fig 2.16B, we plot the baseline

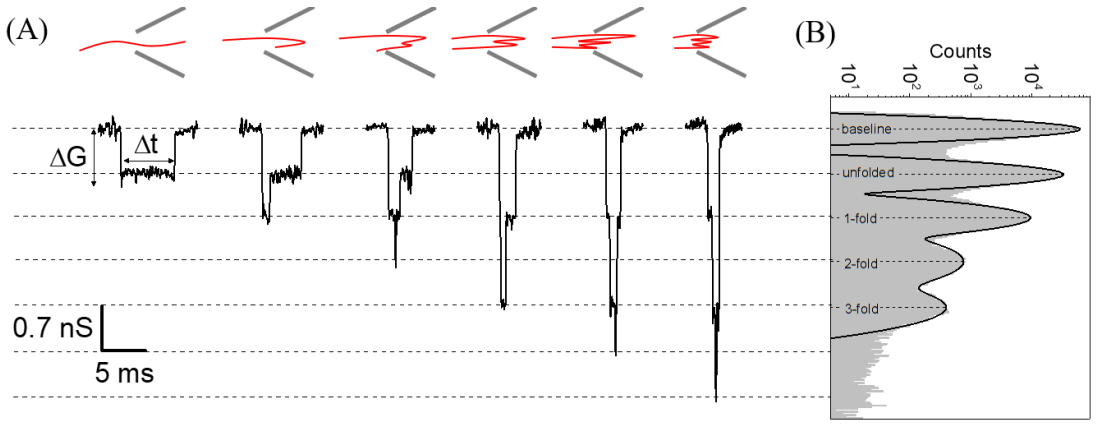


Figure 2.16: Observation of multiple conformations of λ-DNA in a glass nanopore.

(A) Representative events showing different levels of folding during λ-DNA translocation through a 20 nm pore (500 mV). Four different blockade levels, corresponding to different folded DNA conformations, are seen along with some rare events beyond the fourth level. Schematic of possible conformations that could generate the corresponding events is shown above them. (B) Histogram of the nanopore conductance, G (nS) during sample translocation. The histogram shows peaks (dotted lines) at integral levels corresponding to folds of λ-DNA. Data is lowpass filtered at 5 kHz.

subtracted conductance histograms (G histograms) for the translocation data and show the familiar equally-spaced multi-peak histogram showing conductance drop levels corresponding to different folded conformations of linear DNA [68]. We have fitted the conductance histogram with a multi-peak Gaussian function. Here, the first peak is from the open pore baseline and serves as a reference for the measurement of ΔG values ($\Delta G = G_{peak} - G_{baseline}$) for the different current levels. We find that the peaks corresponding to the events are located at 0.70 ± 0.07, 1.4 ± 0.1, 2.1 ± 0.2, and 2.8 ± 0.2 nS relative to the baseline. These ΔG values are integer multiples of 0.70 nS (unfolded configuration) and represent populations of multiply-folded conformations of λ-DNA. There are events beyond 3-fold level but they constitute only 9.3 % of the total number of events and hence didn't result in a clear peak in the Fig 2.16B. We have presented the events with higher-folds for completion.

2.1.9 Collision peaks

In a typical nanopore experiment, there are event-like spikes, called collision peaks detected which are quite faster compared to the translocation events. These are caused by the analyte molecules colliding with the pore mouth and goes back into the well resulting in unsuccessful attempt to translocate through the nanopore. It must be noted

that collision peaks are seen in all nanopore experiments and can interfere with the interpretation of results. It is important to identify and eliminate them.

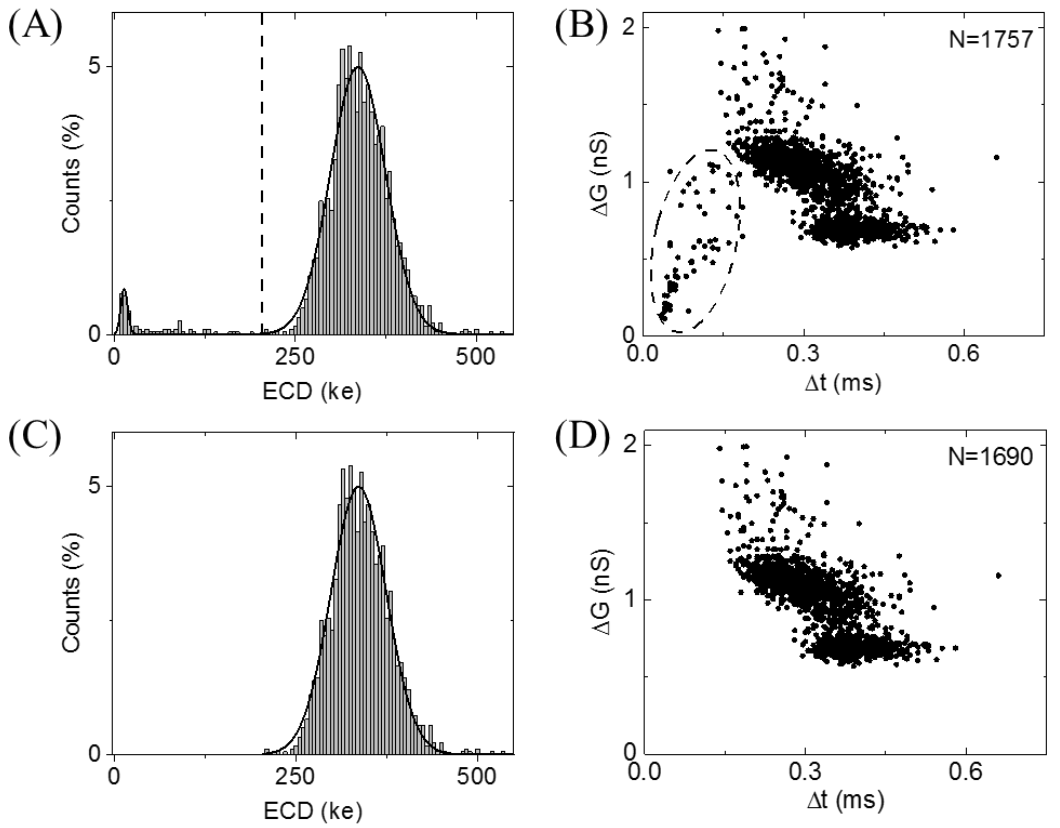


Figure 2.17: Identification and elimination of collision peaks from nanopore data.

Data presented here is for nanopore P31 of 19 nm, measuring translocation events of 3kb linear plasmid DNA at 300 mV. (A) & (B) shows the ECD histogram and ΔG-Δt scatter plot for lin-DNA without any ECD filter. We can see the presence of collision events. (C) & (D) shows the ECD histogram and scatter plot for the same data after filtering events only above the ECD threshold.

We estimate the event charge deficit (ECD), defined as the integral of the current drop over the dwell time duration, for every translocation event [118]. It is computed according to the equation,

$$ECD = \int_{0}^{\Delta t} \Delta I \, dt \tag{2.27}$$

for each event in a given dataset. The calculation of ECD also allowed us to eliminate the occasionally detected collision peaks as shown in Figure 2.17. We find that ECD shows two peaks at 14 ± 5 ke^- and 336 ± 38 ke^-. The events corresponding to the higher value are due to translocation and the events with lower value correspond to collision peaks. We filter out the collision events from the dataset by choosing only the events

with ECD > μ - 3.5 x σ. Here, μ and σ are the mean and standard deviation of the ECD peak for our sample.

2.2 Atomic force microscopy

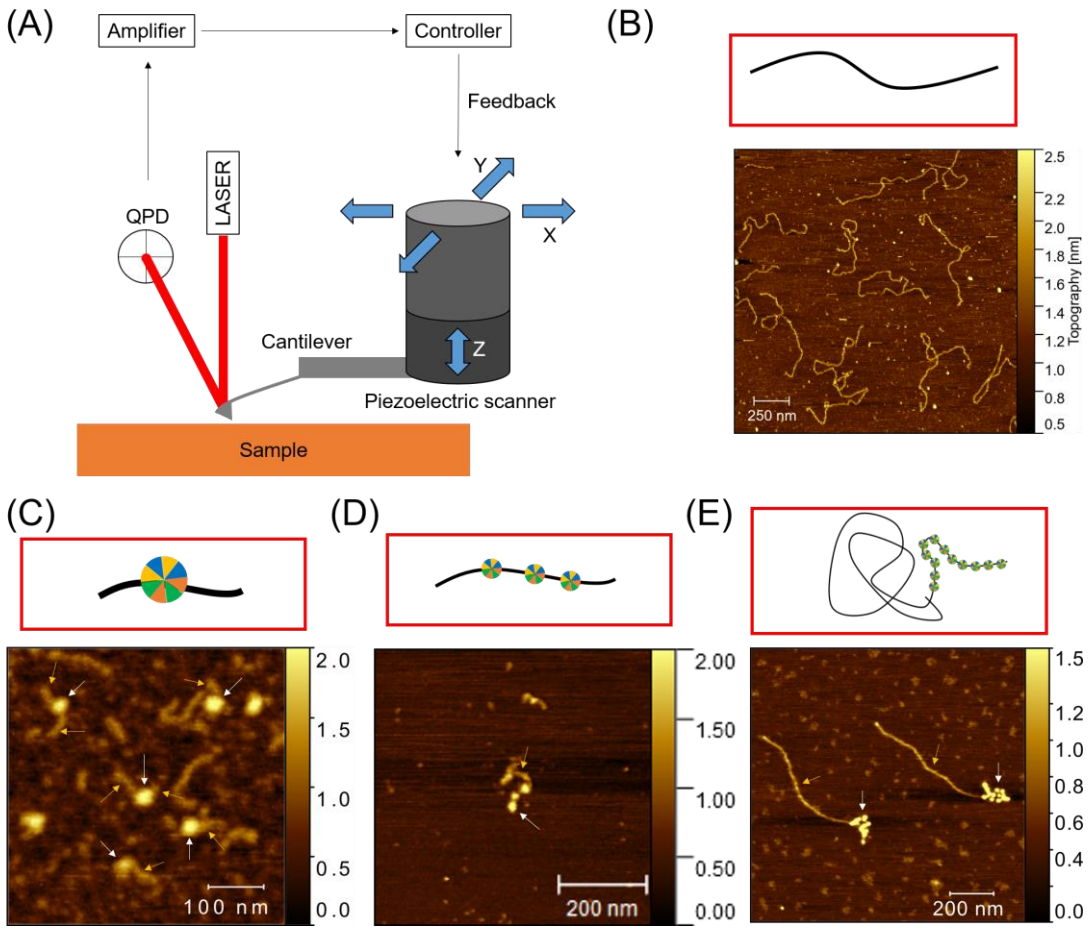


Figure 2.18: Basics of Atomic force microscope.

(A) A schematic of the AFM is shown with important components such as piezoelectric scanner, cantilever, sample, laser, QPD, amplifier and controller. (B)-(E) show the AFM images of 3 kb linear DNA, 344 bp nucleosomes, 3-5mer nucleosome array (Cat# 16-0022, Epicypher) and 12mer nucleosome array respectively.

AFM employs a sharp tip attached to a cantilever that moves close the surface of interest in a constant force mode and adjusts its distance from the surface in order to maintain the constant force (Fig 2.18A). The vertical movements (z-axis) made by the piezo which is then mapped onto the xy-positions thereby provide topographical image of the surface that was imaged. This technique relies on the change in oscillation amplitude depending on its distance from the surface. The parameters such as integral and proportional gains, tip speed and the setpoint applied during the imaging are of utmost importance for

accurate imaging. During tapping mode, the frequency sweep is done where the frequency vs amplitude for the cantilever plot is obtained. For imaging, a drive frequency which is slightly below the resonant frequency is chosen, where the amplitude is about 90% of the peak value. The amplitude setpoint decides the magnitude of force that is exerted on the surface during the imaging. The cantilever is raster scanned over the region of interest and the image is produced. Some of the images captured for the thesis work are shown in Fig 2.18B-2.18E. AFM is also a great tool to study the mechanical properties of materials as well as molecular interactions (see Appendix 1 and Appendix 2 for more details).

2.3 Gel electrophoresis

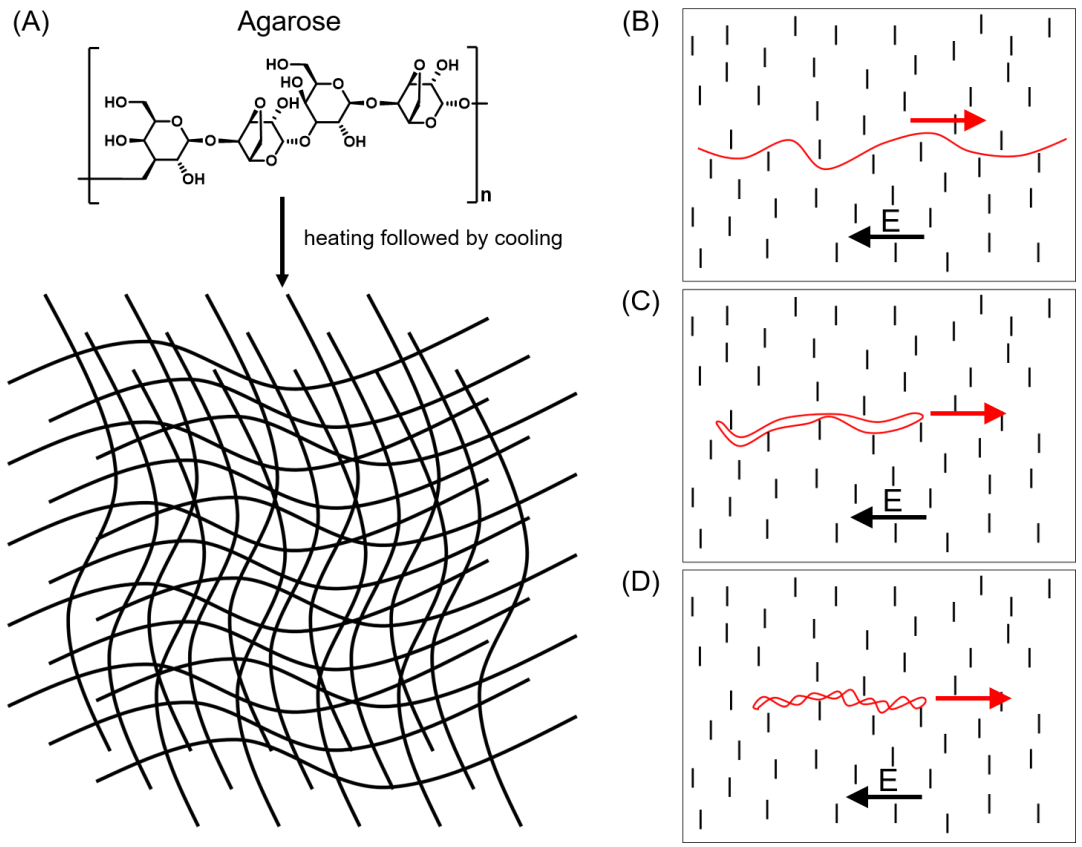


Figure 2.19:: Principle of gel electrophoresis.

(A) Agarose gel is formed by the cross-linking of agarose polymers (image taken from Sigma) with itself due to hydrogen bonding and it forms a porous mesh type of material. (B)-(D) depicts the movement of linear, circular and supercoiled DNA respectively through the gel under the influence of electric field (E). The red and black arrows denote the directions of DNA movement and electric field respectively. Here, the cross-sectional view of the gel network in the plane of DNA is shown.

The principle of this technique is to estimate the length of the DNA fragment from its

electrophoretic mobility (Fig 2.19). The samples are loaded at one end of the mesh made by gel materials such as agarose (Fig 2.19A) or polyacrylamide and they are dipped in the electrolyte solution. A voltage difference is applied across the length of the gel. DNA sample is added at the end of the gel where there is negative terminal. As the DNA is negatively charged, it moves through the gel towards the positive terminal. The gel material is porous mesh in 3D and the DNA has to find its way through this mesh. DNA finds resistance due to the porous material and slows down significantly. The resistance increases as the DNA length increases. This helps in resolving the DNA of different lengths. Generally, a calibrated sample of a mixture of different known DNA lengths, called a Ladder is run parallelly to the samples of our interest as they experience the same magnitude of electric field which means the sample of our interest runs at the same position and speed as the one of the samples in the ladder mixture and this helps in estimation of the unknown DNA length.

Apart from resolving different lengths of the DNA, the gel electrophoresis technique can also resolve the different conformations of DNA, such as linear (Fig 2.19B), circular (Fig 2.19C) and supercoiled forms (Fig 2.19D). As the supercoiled DNA is in a compact form, it experiences lowest amount of frictional force compared to the other forms and is the fastest. The linear DNA exists in coiled globular form in solution and needs to extend in order to reptate through the confinements in gel mesh and hence experiences friction during its movement. Circular DNA on the other hand, must additionally bend in order to move through the mesh apart from the reptation required similar to the linear DNA. This makes circular DNA move slower compared to linear DNA. Though the total charge on the DNA in three forms is the same, they have different effective mobilities. This indicates the differences in the motion of these forms through the mesh structure. The friction experienced by supercoiled < linear < circular DNA. Circular DNA experiences highest frictional drag due to the repulsive electrostatic interactions between the two fragments of the DNA on opposite sides as they have to move towards each other due to their movement through the narrow confinements in the mesh network. The pore sizes were estimated to be in range of 200 – 500 nm for gel containing 1% agarose.

2.4 Binding kinetics of Cas9-DNA complex

In this section, we demonstrate the application of AFM and gel electrophoresis techniques to perform quantitative binding studies on Cas9 and DNA. The discussion presented here will help us in understanding the protein-DNA interactions that are presented in the later chapters of this thesis.

2.4.1 Cas9 binding to target DNA: Role of DNA conformation

We first test the changes in electrophoretic mobility upon Cas9 binding to different DNA conformations. In Fig 2.20A, we show binding of RNA loaded Cas9 enzyme to target DNA plasmid in three different DNA conformations: supercoiled (scD) (lanes 2-6), nicked circular (cirD) (lanes 7-11) and linear (linD) DNA (lanes 12-16). Refer to Appendix 3 for details on materials and methods. Note that the Cas9 enzyme without the targeting guide RNA (*apo*-Cas9) does not bind to the target DNA (lanes 3, 8 and 13). For all three DNA conformations, the Cas9 binds to the DNA and shifts the DNA band only in the presence of the sequence specific targeting guide RNA (lanes 5, 10 and 15). Interestingly, we see the maximum gel shift when the Cas9 binds to the supercoiled (scD) conformation of the DNA (lane 5) when compared to the circular and linear conformations of the plasmid (lanes 10 and 15), where the enzyme bound conformation resulted in only a marginal shift in band position. This would indicate that for the more open DNA targets (linear and circular), Cas9 binding only locally affects the DNA conformation without structurally affecting the rest of non-target DNA.

For the much compact conformation of supercoiled DNA, on the other hand, Cas9 binding dramatically affects the electrophoretic mobility. Importantly, we find that the Cas9 enzyme stays bound to the DNA target after cleavage. This Cas9-DNA complex is stable (for all three conformations) and the cleaved products were released only upon disrupting the complex by heating the reaction mixture to 90 °C. This stable complex suggests that, post cleavage, the Cas9 enzyme keeps the cleaved ends of the DNA in a strongly bound complex. As expected, products of Cas9 cleavage of both scD and cirD resulted in the linear conformation (lane 6 and lane 11) and the linear DNA target was cleaved and the products were released as two shorter linear DNA fragments of lengths

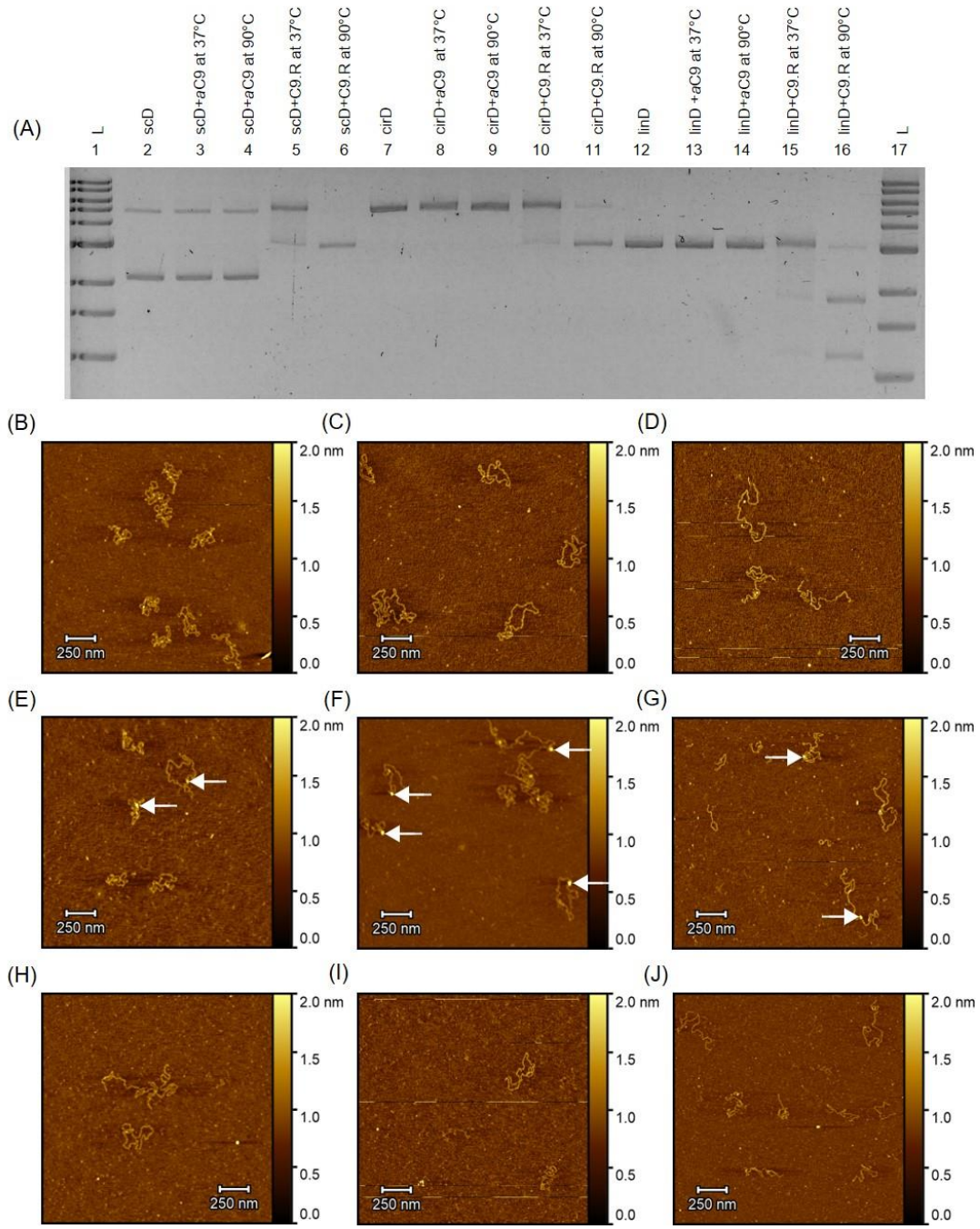


Figure 2.20: Cas9 binding to different conformations of DNA

(A) Gel image showing the Cas9 bound to supercoiled, nicked circular and linear DNA. Lane 1 and 17 show 1kb ladder. Lanes 2, 7 and 12 show control DNA. Lanes 5, 10 and 15 show the gel shift due to Cas9 bound DNA. Lanes 6, 11 and 16 show the release of cleaved products upon heating. (B)-(J) shows the AFM images of samples scanned using tapping mode in air. The three columns show supercoiled, circular and linearized plasmids respectively. (B)-(D) shows AFM images of bare supercoiled, nicked, and, linearized form of the plasmid, respectively. (E)-(G) shows Cas9-bound (white arrows) states of the three plasmid conformations. (H)-(J) show heat-released products of Cas9 cleavage of the three plasmid conformations.

1.2 kb and 1.8 kb respectively (lane 16). Note the faint band corresponding to a small amount of circular DNA in supercoiled DNA sample (see lanes 2-4). Fig 2.20B-D shows

the AFM images of all the three conformations: supercoiled (scD), nicked circular (cirD) and linear (linD). Fig 2.20E-G shows the Cas9 bound state of the three conformations before the heat release. We see here the Cas9 protein (white arrows) bound to the DNA molecules post the 1-hour incubation. We note that these molecules were not cleaved. Fig 2.20H-J shows AFM images of the heat released products of Cas9 cleavage for the three DNA conformations.

In Fig 2.20H&I, we see linearized products of a single Cas9 cleavage of the supercoiled and circular plasmids, respectively. Fig 2.20J shows the smaller length DNA products from Cas9 cleavage of linear DNA.

2.4.2 Cas9 binding to target DNA: Time dependence

To measure the binding/release kinetics, supercoiled form of target DNA was used for all further experiments. Target DNA was added to the reaction-mixture with RNA loaded Cas9 and incubated at 37 °C. In order to find the fraction of unbound DNA, 12.5 ul of the sample was collected from the reaction mix at different time points t = 0 s, 4 s,10 s, 30 s, 40 s, 50 s, 1 min, 2 min, 3 min, 10 min, 30 min and 60 min into tubes containing 25 mM EDTA and incubated on ice for 30 min to stop the reaction. One half of the t = 0 s time point sample was used for testing the release of the cleaved product by denaturant (adding 0.29 M SDS for 10 min at 37 °C). The other half was used as control to test for binding at 37 °C for simultaneous visualization on the gel. As seen from Lane 17 in Figure 2.21A, absence of any linear DNA confirms that the reaction makes no progress after addition of EDTA to the mixture. Time dependent binding of Cas9 enzyme to target DNA (scD) is shown in lanes 5 to 16 for the mentioned incubation times respectively.

The percentage of unbound scDNA can be estimated according to the following equation:

$$\%\ \text{Unbound scDNA} = \frac{\text{Intensity of unbound scDNA}}{\text{Total intensity of all bands}} \times 100 \tag{2.28}$$

We found that upon Cas9 binding, the amount of unbound scDNA decreased from 80 ± 4 % at 0 s to 7 ± 2 % within 60 min (Fig 2.21B). We quantify the binding timescales from the gel band intensities in Fig 2.21B and calculate the rate constant of the reaction (k_{bind}) by fitting data to the following rate equation:

$$y(t) = y_0 e^{-k_{bind} t} + c \tag{2.29}$$

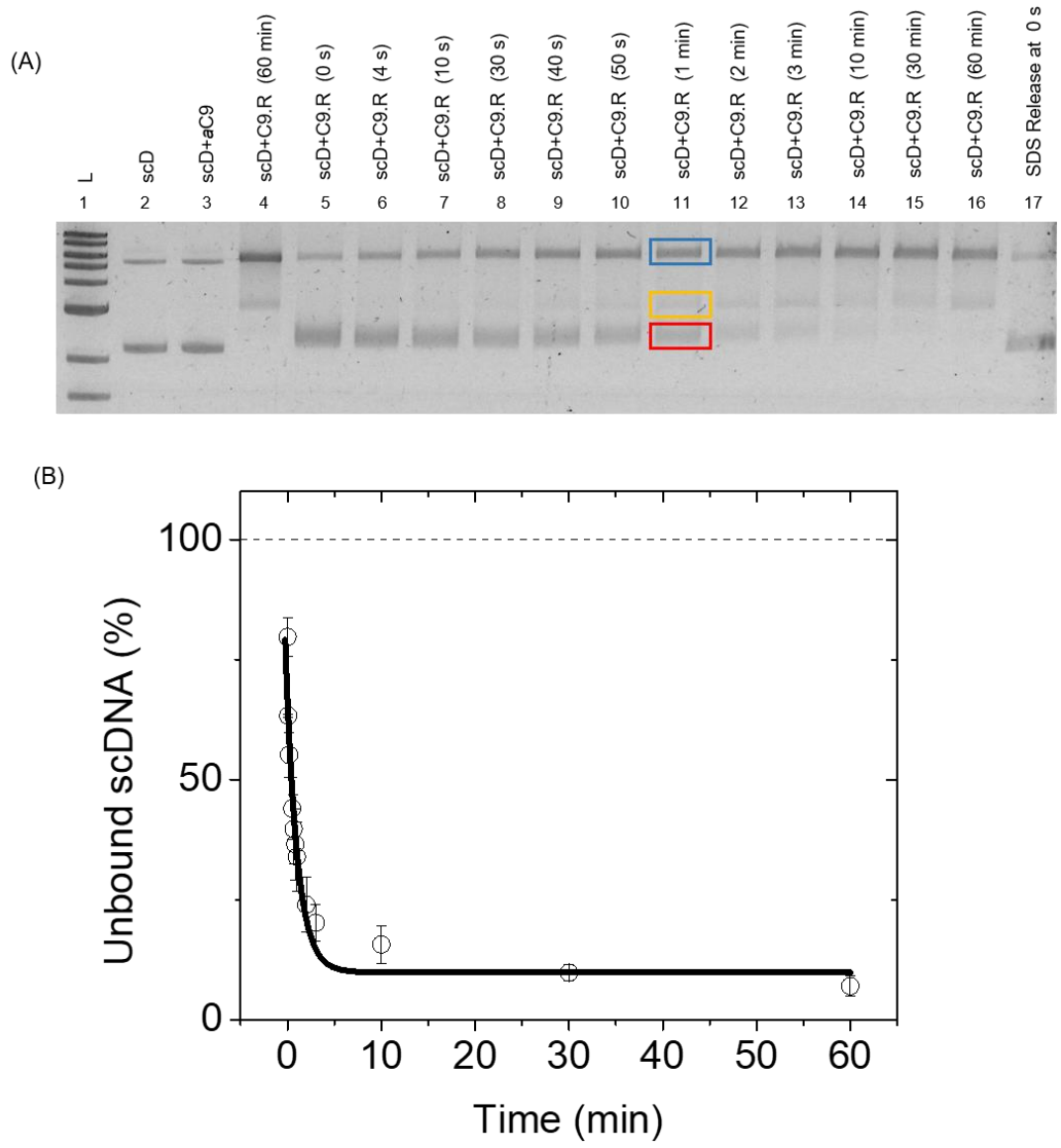


Figure 2.21*:* Time dependence of Cas9 binding to target DNA

(A)The gel shows the binding of Cas9 to the target DNA as a function of time and the subsequent arrest of the bound product with EDTA. Lane1: 1 kb ladder, lane2: control DNA, lane3: DNA+ apo-Cas9 at 37 °C, lane4: Cas9.RNA bound to the target DNA at 37 °C for 60 min which is the binding control. Lanes 5 to 16 show the binding of Cas9.RNA to the target DNA at 0 s, 4 s, 10 s, 30 s, 40 s, 50 s, 1 min, 2 min, 3 min, 10 min, 30 min, 60 min respectively. Lane 17 shows the SDS release of target DNA at 0 s to confirm no linear DNA is released. The blue, yellow and red boxes in lane 11 represent the typical positions for bound, released and the unbound gel bands used for DNA quantification. (B) shows quantification of the Cas9 binding rate constant from the gel band intensities of the unbound supercoiled DNA in lanes 5 to 16 respectively. (k_{bind}= 0.8 ± 0.2 min^{-1}) Solid line is fit to equation 2.29. (The mean and error bars are from N=3 independent experiments.

Here, y(t) is the percentage of unbound scDNA as a function of reaction time (t), y_0 is the initial percentage of unbound scDNA and c is the offset in percentage values. From the fit to equation (2), we obtain the rate constant of binding to be k_{bind} = 0.8 ± 0.2 min^{-1}.

The application of the above-mentioned techniques to study other systems involving DNA-protein interactions will be explored in the following chapters.

2.5 Conclusions

In this work, we demonstrate the application of quartz nanopore platform in high resolution measurement of conformational details of supercoiled DNA. We showed controlled fabrication of quartz nanopores of various sizes ranging from 13 nm to 93 nm with their conductance values in agreement with the standard theoretical model. The signal to noise ratio in nanopores was suitable for distinguishing folded configurations of the long λ-DNA due to generation of electrical signatures with conductance drop values that are multiples for that of unfolded DNA. We have also presented the principles of AFM and gel electrophoresis and also shown their application to characterize particularly the Cas9-DNA binding kinetics. This prepares us to later apply these techniques for investigation of biomolecules such as supercoiled DNA and chromatin.

CHAPTER 3

Fingerprinting branches on supercoiled plasmid DNA using quartz nanocapillaries

Abstract

DNA conformation, in particular it's supercoiling, plays an important structural and functional role in gene accessibility as well as in DNA condensation. Enzyme driven changes of DNA plasmids between its linear, circular and supercoiled conformations control the level of condensation and DNA distal-site interactions. Many efforts have been made to quantify the branched supercoiled state of the DNA to understand its ubiquitous contribution to many biological functions, such as packaging, transcription, replication etc. Nanopore technology has proven to be an excellent label-free single-molecule method to investigate conformations of the translocating DNA in terms of the current pulse readout. In this chapter, we present a comprehensive study to detect different branched-supercoils on individual plasmid DNA molecules. Using a detailed event charge deficit (ECD) analysis of the translocating molecules, we reveal, for the first time, the distributions in size and the position of the plectoneme branches on the supercoiled plasmid. Additionally, this analysis also gives an independent measure of the effective nanopore length. Finally, we use our nanopore platform for measurement of enzyme-dependent linearization of these branched-supercoiled plasmids. By simultaneous measurement of both single-molecule DNA supercoiled conformations as well as enzyme-dependent bulk conformational changes, we establish nanopore sensing as a promising platform for an in-depth understanding of structural landscapes of supercoiled DNA to decipher its functional role in different biological processes.

3.1 Introduction

Supercoiling which acts as a topological constraint on DNA is a key control mechanism in regulating transcription, replication, and compaction of DNA in both prokaryotes and eukaryotes [1,2,119]. It has been a topic of great interest to physicists and biologists alike, due to the strong interdependence between its structure and function [120–122]. Almost all organisms have negatively supercoiled DNA [4,5] except hyperthermophilic archaea that have positively supercoiled DNA [6]. The supercoiled DNA can assume a variety of interwound branched structures, called plectonemes, which forms due to the interplay between twisting and bending energy of the DNA [1,2]. Plectonemes with branches protruding out of the super helical axis were initially observed by electron microscope [1] and later studied in detail using AFM [7–9] under physiological conditions. Fluorescence and AFM based single molecule studies have revealed that the plectoneme formation depends on the underlying sequence as it governs the local DNA curvature [12,18,20]. The interwinding and branching can juxtapose multiple distant sites which are important for processes such as enhancer-promoter interaction [13,14]. The negative supercoiling favours the local untwisting of DNA thereby affecting the activity of transcription [123] and restriction digestion [124]. The studies on the kinetics of interconversion between the linear, circular and supercoiled plasmid conformations [125] need to be carried out as the DNA inside a cell is highly dynamic due to its interaction with various types of enzymes during the processes like nucleosome formation, recombination, viral genome integration, etc. Mechanisms involved in the maintenance of supercoiling, the kinetics of modulating the supercoiling of DNA by topoisomerases and gyrases have been studied using single molecule techniques like magnetic tweezers [126,127] and rotor bead assays [128]. However, there is a dearth of single molecule measurements required for quantification of the native branched structures and enzyme dependent conformational changes in DNA supercoiling.

Nanopore platform has recently emerged as an excellent platform for detecting various polymers. They have been used to detect molecules like DNA, RNA, proteins and protein-DNA complexes [70–75]. The application of nanopores in the detection of conformational changes in proteins is also demonstrated [78–80]. They have been used

for measuring the kinetics of biomolecular reactions [80,81] and to study conformations of DNA in the distribution of knot conformations, G-quadruplexes [86–88] and supercoiled DNA [89–91]. However, until now, high resolution detection of different branched populations and the length and position analysis of naturally occurring plectonemes on supercoiled plasmid DNA has not been possible.

In this work, we demonstrate the nanopore platform as a promising label-free technique to quantitatively measure a variety of branched DNA structures in aqueous solutions at nanometre resolution. We present a detailed event charge deficit (ECD) based analysis to quantify the DNA lengths in these supercoiled branches. This analysis also provided us a method to determine the effective length of the nanopore. Finally, we apply the nanopore measurements to study the kinetics of enzyme dependent conformational changes in supercoiled DNA. Our work extracts a wealth of information about the branched sub-structures in supercoiled DNA demonstrating facets of the nanopore platform useful in the study of the dynamic conformational changes of DNA at high resolution.

3.2 Materials and methods

3.2.1 Sample Preparation

For nanopore translocation experiments on λ-DNA (New England Biolabs), the sample was heated at 65^{o} C for 10 min to linearize the DNA and then added into the sample well at a final concentration of 4.2 ng/µl. The experiments with supercoiled DNA were performed using the plasmid pGEM3z601b [129,130] (a gift from Cees Dekker's lab). This plasmid was amplified by growing in *E. Coli* DH5α and extracted using a Midi Prep kit (Qiagen). The circular form of the plasmid was obtained by nicking with Nt. BspQI enzyme (New England Biolabs, NEB) and the linear form was obtained by single-cut digesting the supercoiled plasmid in 1X CutSmart Buffer (NEB) by either NdeI, ScaI or EcoRI restriction enzymes (NEB) at a concentration of 0.14 U/ul (1X). The samples are purified using a PCR purification kit (Qiagen) followed by ethanol precipitation. For enzyme kinetics experiments, the digestion reactions with NdeI enzyme were performed in 50 mM Tris, 100 mM NaCl, 5 mM $MgCl_2$ pH 7.5 [125] at 1/6X enzyme concentration

(see Fig 3.9). The digestion reaction was arrested at different time points (see Results section) by the addition of EDTA to a final concentration of 100 mM and plunging immediately on ice. All the translocation experiments on plasmids were performed at a final concentration of 0.8 ng/µl.

3.2.2 Nanopore Experiments

The nanopores were fabricated and prepared for experiments as explained in the previous chapter. Data acquisition was performed using NI DAQ PCI-6251 (National Instruments) at a sampling frequency of 200 kHz. The translocation data was acquired and analysed using custom-written codes in LabVIEW (National Instruments). For data analysis, the translocation data for plasmids was low-pass filtered at 25 kHz. A translocation event is defined as a trace where the current drops below the threshold value of 5 times the standard deviation of noise in the baseline. Branch analysis was performed using a custom-written code in MATLAB. When comparing samples measured on the same nanopore, the fluid cell and the nanopore were washed thoroughly with buffer and a stable baseline was confirmed before adding each sample to ensure that there is no sample left behind from the previous experiment.

3.3 Results

3.3.1 Characterization of plasmid conformations

Here, we compare different forms of a 3kb plasmid DNA which are in linear (lin-), circular (cir-) and supercoiled (sc-) conformations. AFM images clearly show the difference between three conformations of DNA (Fig 3.1 left panel). Each of the DNA molecule assumes various possible conformations on the substrate. Linear DNA and circular DNA assume open and closed conformations as expected. Supercoiled DNA is compact and shows multiple branches in each molecule. Representative current traces for different conformations are shown in Fig 3.1 right panel.

Fig 3.2A shows the representative electrical events of the three types of DNA conformations when compared on the same nanopore (pore diameter 19 nm). The

experiments were performed on the same nanopore, one sample at a time, with extensive washes between different sample measurements, as described in the Material and Methods section. Fig 3.2B shows the gel image of the 3 forms of the plasmid DNA.

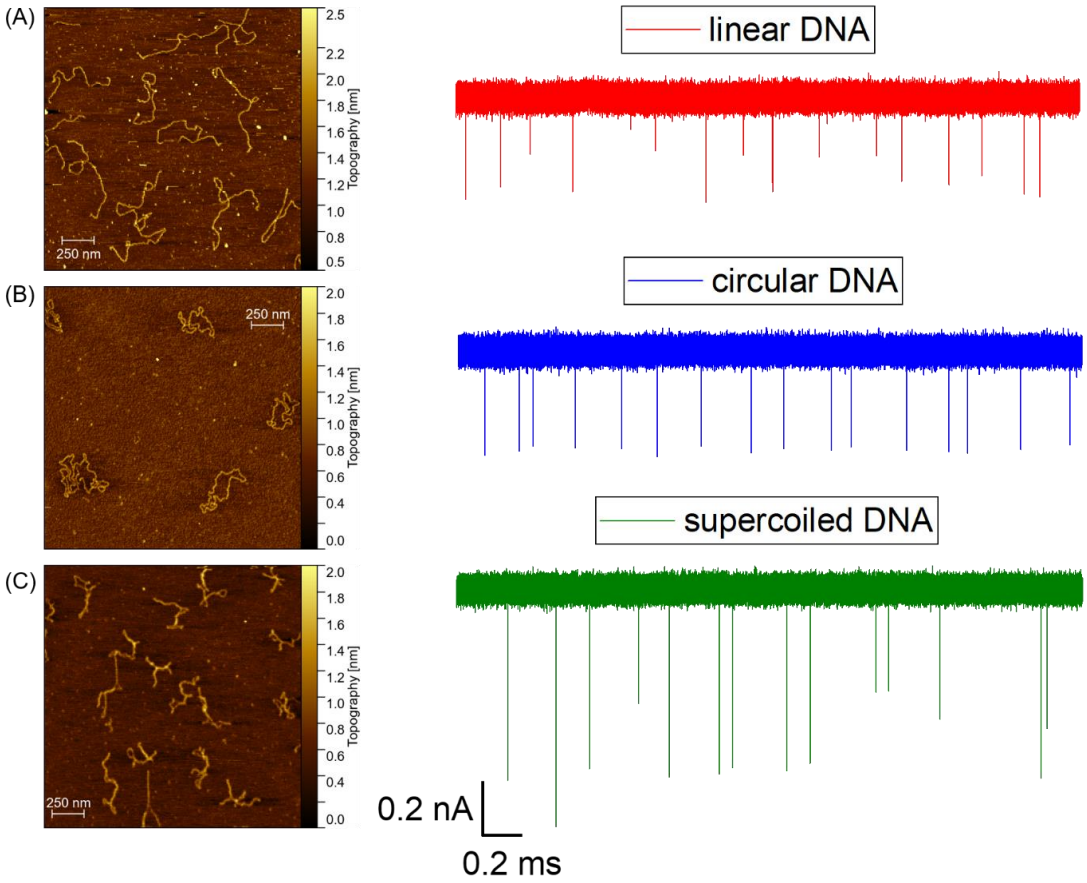


Figure 3.1: AFM images of plasmids and Nanopore current trace.

(A)-(C) AFM images of linear, nicked circular, and supercoiled plasmid forms respectively. (D) Current trace for different conformations of plasmid DNA from P31. Data presented here is for nanopore P31 of 19 nm, measuring translocation events of linear plasmid DNA at 300 mV.

Fig 3.2C-3.2E compares the ΔG, Δt, ΔG-Δt scatter plots and the ECD values of the detected events when each conformation of the plasmid DNA was measured individually. Gaussian fits were used to identify different peaks. Linear DNA sample shows two populations corresponding to ΔG values of 0.5 ± 0.1 nS and 1.0 ± 0.1 nS respectively and two populations in Δt histograms corresponding to dwell times of 0.26 ± 0.03 ms and 0.37 ± 0.05 ms (Fig 3.2C & 3.2D, lin-DNA, red). The ΔG value for the second peak is almost double that of the first peak indicating that the two populations correspond to unfolded (smaller ΔG, longer Δt values, marked by *) and the folded forms (higher ΔG, shorter Δt values, marked by **) of the linear plasmid DNA, respectively. The percentage of events in the unfolded DNA peak is 33.4% when measured using this pore. This value

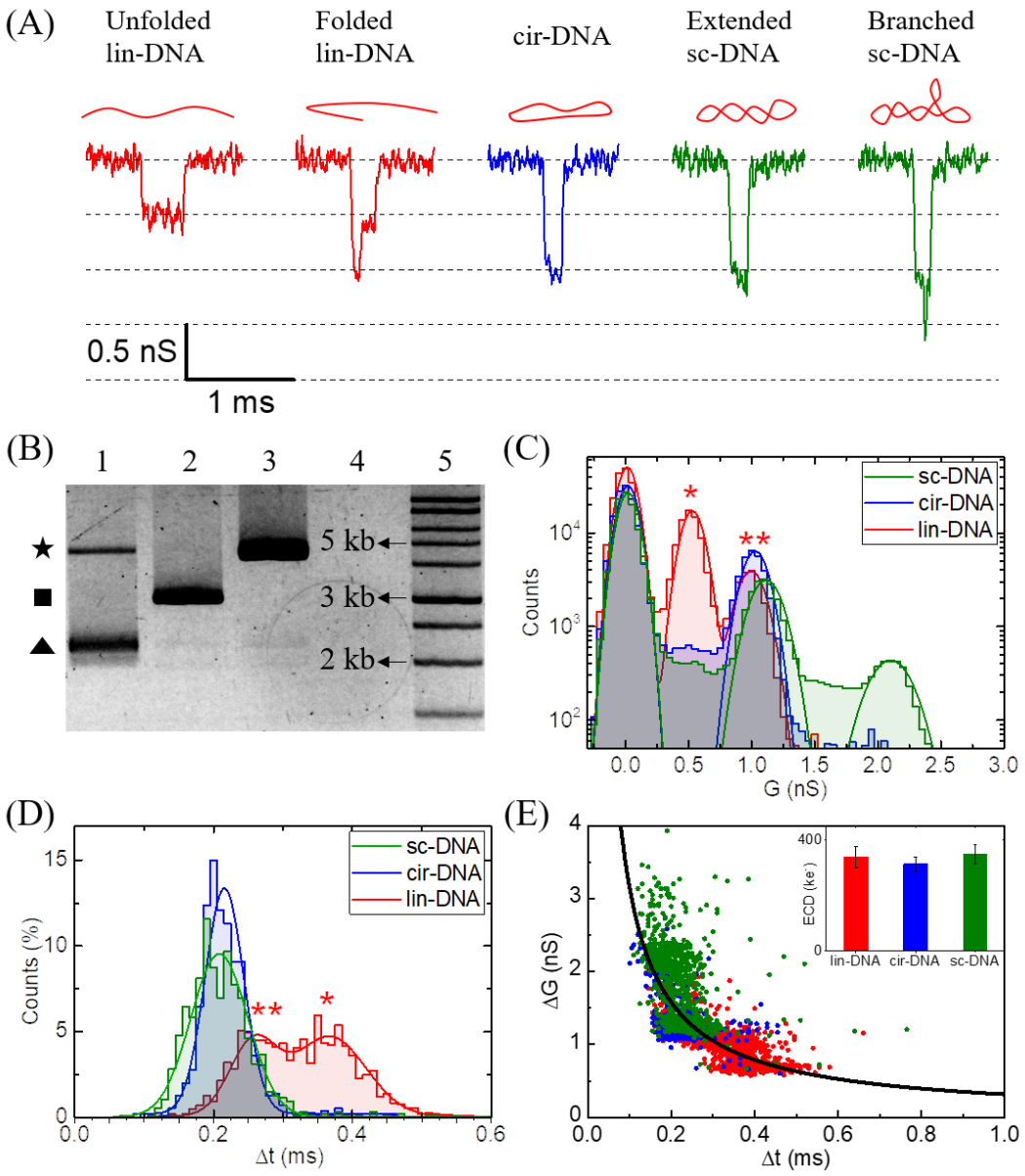


Figure 3.2: Characterization of different conformations of plasmid DNA.

Events corresponding to linear (lin-DNA, red), circular (cir-DNA, blue) and supercoiled (sc-DNA, green) 3kb plasmid translocating through a 19 nm pore (P31). (A) Representative events recorded during the translocation of individual conformations of plasmid DNA. Schematics of DNA conformation corresponding to each event is shown on the top. (B) Gel image showing three plasmid forms – Lane1 : sc-DNA (triangle), Lane2: lin-DNA (square), Lane3: cir-DNA (star), Lane4: empty and Lane5: 1 kb DNA ladder. (C) compares conductance (G (nS)) histograms for the three samples. We find that the lin-DNA and sc-DNA show multiple populations whereas cir-DNA showed only one population. (D) compares Δt histograms for the same data set showing that the folded events (marked by **) have higher ΔG values and shorter Δt when compared to unfolded (marked by *) DNA events. (E) shows the ΔG-Δt scatter plot. Here the solid line is the plot (not fit) for ECD value of 332 ± 18 ke$^-$. Inset: Bar plot showing the ECD values for the individual samples of different plasmid forms. The number of events, n = 1690 (lin-DNA), 1115 (cir-DNA), 960 (sc-DNA).

was measured using multiple 20 nm nanopores and we find it to be highly reproducible with a mean percentage of 36.5 ± 3.1 of the events in the unfolded DNA peak (Table 3.1). We later apply this approach to quantify the amount of linear DNA from a mixture

of lin-DNA and sc-DNA from individual nanopores (see Fig 3.7 and Fig 3.8). Circular DNA sample, measured in the same nanopore, shows a dominant single population (90.3% of total events) corresponding to ΔG value of 1.0 ± 0.1 nS and a single peak in Δt histogram at 0.22 ± 0.03 ms (Fig 3.2C & 3.2D respectively, cir-DNA, blue). In the cir-DNA sample, the absence of any folded events is possibly due to its total length of 514.25 nm being closer to the dsDNA persistence length when compared to the linear plasmid DNA (1028.5 nm total length). Longer DNA, such as λ-DNA in previous chapter, however shows many more levels of folding and supports this possibility.

Pore ID	**Sample**	**ΔG (nS)**	**Δt (ms)**	**ECD (ke^-)**	**% events in peak1**	**% events in peak2**
P28	lin-DNA	0.7 ± 0.1 1.3 ± 0.1	0.25 ± 0.04 0.36 ± 0.05	425 ± 46	34.1	65.9
P29	lin-DNA	0.6 ± 0.1 1.2 ± 0.1	0.23 ± 0.03 0.32 ± 0.06	373 ± 37	37.7	62.3
P31	lin-DNA	0.5 ± 0.1 1.0 ± 0.1	0.26 ± 0.03 0.37 ± 0.05	337 ± 38	33.4	66.6
P31	cir-DNA	1.0 ± 0.1	0.22 ± 0.03	312 ± 26	90.3	9.7
P31	sc-DNA	1.1 ± 0.1 2.1 ± 0.2	0.21 ± 0.04	347 ± 34	34.9	65.1
P32	lin-DNA	0.62 ± 0.08 1.18 ± 0.1	0.27 ± 0.04 0.35 ± 0.04	356 ± 37	39.0	61.0
P35	lin-DNA	0.4 ± 0.1 0.8 ± 0.1	0.24 ± 0.04 0.33 ± 0.04	232 ± 21	41.9	58.1
P36	lin-DNA	0.65 ± 0.1 1.3 ± 0.1	0.24 ± 0.03 0.36 ± 0.08	437 ± 61	35.2	64.8
P37	lin-DNA	0.9 ± 0.1 1.8 ± 0.1	0.19 ± 0.01 0.27 ± 0.05	458 ± 51	34.2	65.8

Table 3.1: Summary of population analysis for various pores used

The data in this table corresponds to Fig 3.2, Fig 3.7 and Fig 3.8.

Supercoiled DNA sample measured in the same nanopore showed a primary population with 34.9% of total events corresponding to a ΔG value of ΔG_1 = 1.1 ± 0.1 nS and Δt of 0.21 ± 0.04 ms (Fig 3.2C & 3.2D, green). Being similar to ΔG value for the circular plasmid, we attribute this population to an extended form of sc-DNA. Interestingly, we found a large number of events (65.1%) with a broad range of ΔG values which are higher than the first peak. Such a representative event is shown in Fig 3.2A. Since these deeper

events present themselves with additional current spikes which are not multiples of extended sc-DNA event level and also noting that the persistence length of sc-DNA is (~100 nm), about double that of lin- or cir-DNA [131], we disregarded these events as folded sc-DNA as they are energetically not favorable. However, since it is well known that due to the torsional energy stored in sc-DNA it can assume higher order branched structures [1,7], we attribute the deeper ΔG events detected in sc-DNA sample to these branched structures on the supercoiled plasmid. We note that these spiked events, due to the branched sc-DNA, are resolvable from the extended sc-DNA events only because of the choice of the nanopore. Later, we present the role of nanopore diameter on its resolution (see Fig 3.3). Different sized branched structures on sc-DNA produce current spikes of different amplitudes. It is evident that these structures are mechanically stable as the high electric field gradient near the pore was unable to extend them. These events have a broad range of ΔG values which finally culminates into a population seen as the second peak with a ΔG value of $\Delta G_2 = 2.1 \pm 0.2$ nS. This second population suggests either a large population of supercoiled DNA has a specific size of the branched structure or a cut-off in the detectable branch size that depends on the nanopore dimensions (see discussion for Fig 3.6). We also note that compared to the unfolded-linear conformation of DNA, the folded-linear, circular and supercoiled conformations of DNA take about half the time to translocate through the pore. This is supported by the fact that their lengths are half of the unfolded lin-DNA and suggests that the translocation dynamics of different DNA conformations remain the same. In our experiments of measuring both samples on the same pore, event rates for lin- and sc-DNA were found to be 0.82 ± 0.13 and 0.77 ± 0.54 sec^{-1} respectively from multiple nanopores ($D_P = 20 \pm 2$ nm). We note that in the current conditions, our experiments are not sensitive to conformation dependent changes in capture rates for lin- and sc-DNA of 3025 bp (~1.028 μm) plasmid.

In Fig 3.2E (inset) we show a comparison of the ECD values for the three forms of the plasmid DNA and find their values close to each other. This is expected as the ECD is the total amount of the charge displaced by a molecule when translocating through the pore and is independent of the conformation. This is further evident in Fig 3.2E where the mean of the three ECD values of 332 ± 18 ke^- (kilo-electrons) is plotted (solid line, not a fit) on the ΔG-Δt scatter plot and the curve passes through the scatter plots of all three forms of the plasmid DNA [89]. The calculation of ECD also allowed us to eliminate the occasionally detected collision peaks as shown in previous chapter. The

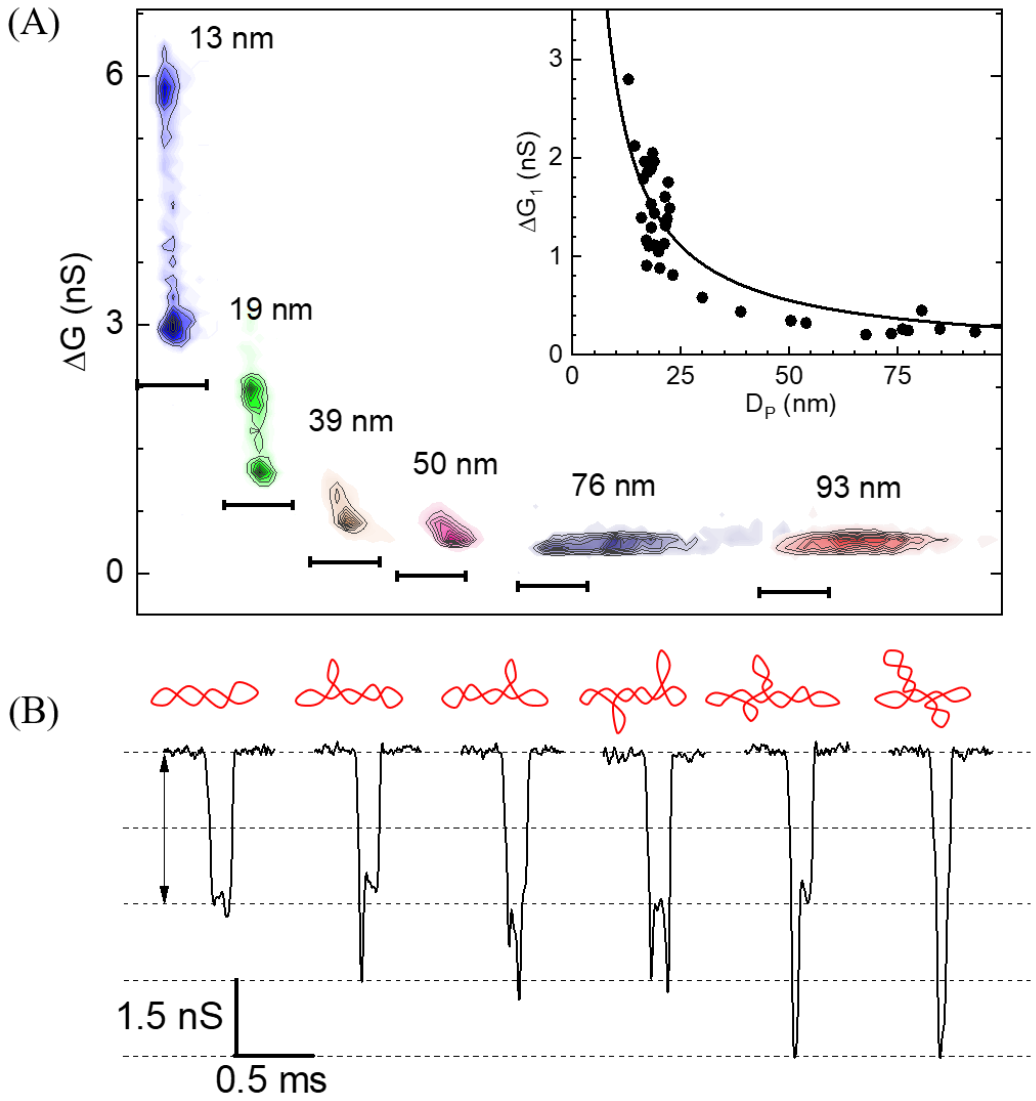


Figure 3.3: Role of nanopore size in resolving branched populations of supercoiled DNA

(A) Density contour plots of conductance drop versus dwell time are compared for various pore diameters – Pores used here are: P24 (13 nm, n = 752), P31 (19 nm, n = 960), P30 (39 nm, n = 1233), P08 (50 nm, n = 1405), P02 (76 nm, n = 637), and P03 (93 nm, n = 926). Here the darker regions represent higher density of data points. For visual clarity scatter plots have common Y-axis and independent time axis scale bars are shown at the bottom of each contour plot which starts from 0 ms and ends at 0.4 ms. Inset: Conductance drop (1st peak, ΔG_1) (300 mV) during translocation of a supercoiled plasmid is plotted as a function of pore diameter when measured for N = 37 different nanopores. The solid line is the fit to eqn (3.1). (B) Representative events showing different conformations of sc-DNA as resolved by a 13 nm pore (300 mV). Schematics of DNA conformation corresponding to each event is shown above it.

summary of ΔG, Δt, percentage of events and, ECD values for all the conformations are listed in Table 3.1. From the above observations, it is clear that the quartz capillary based nanopore platform can successfully distinguish between different plasmid conformations.

In Fig 3.3 we show the role of nanopore size in resolving multiple populations corresponding to different conformations on the sc-DNA plasmid. We note that in larger

pores, the entire sample set is detected as a single peak in the ΔG histogram and is unable to distinguish between branched and unbranched forms. For pore sizes of 39 nm and smaller we see the sc-DNA with branches being detected as a population separate from extended sc-DNA. This can be clearly seen from their ΔG histograms as well, as shown in Fig 3.4. Among all the pore diameters, we find the highest resolution in the 13 nm pore which presents us with a large plethora of event heterogeneity as seen in the representative events shown in Fig 3.3B. Possible branched conformation that may result in the event is shown as a schematic above every event. Given the high throughput, reproducibility and sufficient resolution, all further experiments to quantify the branched samples were performed with nanopores of diameters around 20 nm.

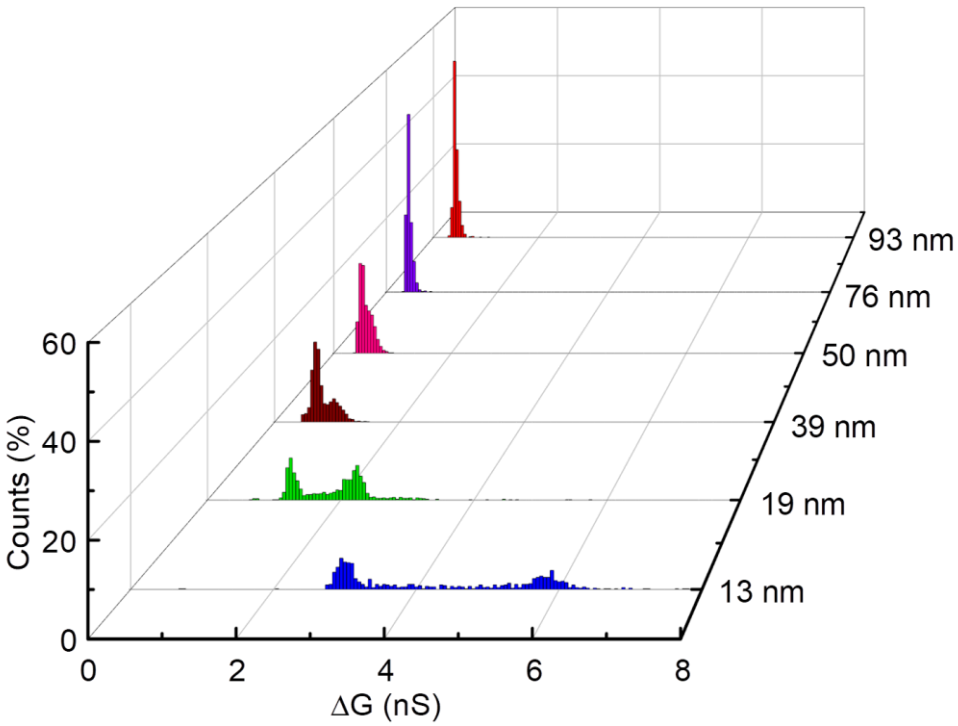


Figure 3.4: Role of nanopore size in resolving branched populations of supercoiled DNA.

Histograms of conductance drop are compared for various pore diameters – Pores used here are: P24 (13 nm), P31 (19 nm), P30 (39 nm), P08 (50 nm), P02 (76 nm), and P03 (93 nm).

Translocation experiments for sc-DNA were performed in a total of 37 nanopores with pore sizes ranging from 13 nm to 93 nm. We compare ΔG corresponding to the extended sc-DNA conformation (first ΔG peak in Fig 3.2C, ΔG_1) for different pore sizes. We show, in Fig 3.3A, that the sc-DNA can be detected in nanopore in this entire size range with the ΔG_1 values increasing non-linearly with decreasing pore diameter (see Fig 3.3A inset & Table A4.1). We consider the nanopore as a cylinder with effective length L_{eff} and

substituting $L_{eff} = K \times D_p$ in eqn (2.14) and using equation $D_{eff} = \sqrt{D_P^2 - 2D_{DNA}^2}$ as there are two DNA molecules in case of scDNA, we obtain [80,110,132,133],

$$\Delta G(D_P) = \sigma \left(\left(\frac{4K}{\pi D_P} + \frac{1}{D_P} \right)^{-1} - \left(\frac{4K}{\pi D_P \left(1 - \frac{2D_{DNA}^2}{D_P^2} \right)} + \frac{1}{\sqrt{(D_P^2 - 2D_{DNA}^2)}} \right)^{-1} \right) \quad (3.1)$$

The ΔG_1 versus D_P is fitted with this equation and obtained $L_{eff} = (3.73 \pm 0.21) \times D_P$. The variation in ΔG_1 values for similar pore sizes is due to the variation in internal profiles of the capillaries which are not accessible to us. A cluster of data points is visible around 20 nm because our experimental investigations are primarily focused on pores with diameters in this size range.

3.3.2 Quantification of branches in supercoiled DNA using ECD analysis

We quantify the branches in sc-DNA using ECD analysis of the translocation events. Fig 3.6A shows the typical translocation events of sc-DNA through a 17 nm pore. In each translocation event, the blockade level corresponding to an extended form of sc-DNA is shown in black and the red trace represents the detected spike on top of the event which corresponds to the branched region translocating through the pore. The size and location of the spike correspond to the size and location of the branch on the translocating sc-DNA molecule [87,92] The ECD of an event corresponds to the total amount of ions displaced by the molecule which depends on the size (length, in case of a linear polymer) of the molecule and is independent of its conformation during translocation. We use this analysis to quantify the linear lengths (L_b) and positions (Z_b) of the branches in sc-DNA using the following equations, for all events that display the branch signatures. A detailed step-by-step analysis procedure is shown in Fig 3.5.

We exploit the ECD for characterizing features on DNA for both linear and supercoiled DNA samples. We collect all events with spikes by considering events with $\Delta G > \Delta G_1 + 2\sigma_1$ and $\Delta G < \Delta G_2 + 2\sigma_2$ as branched (sc-DNA) events. The events with $\Delta G > \Delta G_2 + 2\sigma_2$ have faster dwell times and do not contain resolvable spikes as they possibly

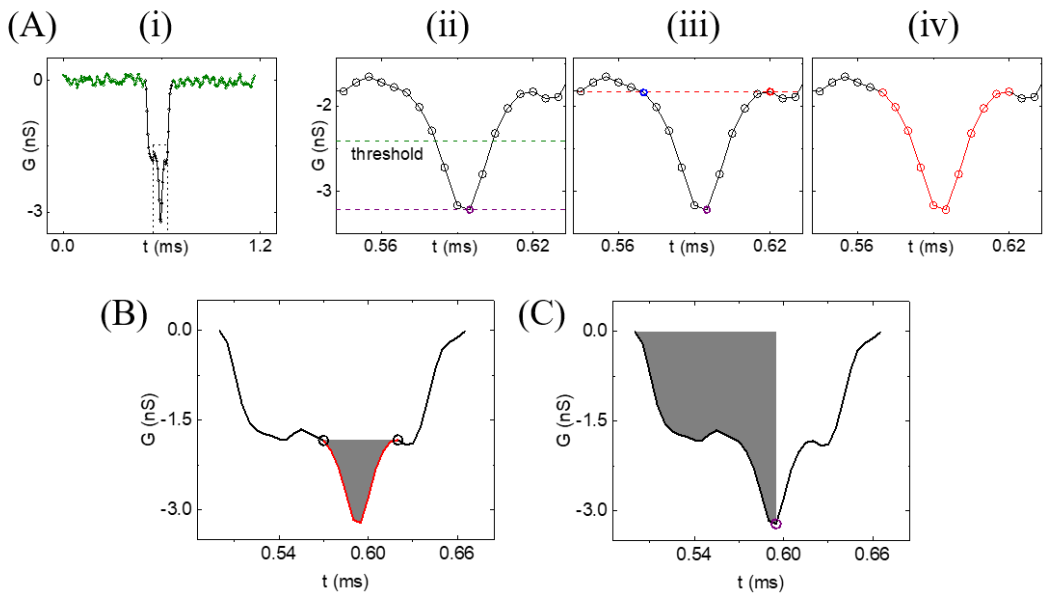


Figure 3.5: Branch identifications and characterization from spikes in translocation event.

(A) shows representative event (i) of sc-DNA translocating through a 17 nm nanopore (P25). The green trace is the baseline and the black trace is the translocation event with the current spike due to the branched region marked as the dotted region. Zoom of this current spike (dotted region) is shown on the right in steps (ii) – (iv), describing the three-step event processing. (B) & (C) shows the full event (black trace in (A)(i)) with the ECD of the branched region (red) and the ECD up to the location of the current spike (purple circle) shaded along the raw event.

represent highly compact structures. Here σ_1 and σ_2 are the standard deviations in the gaussian fit to first and second ΔG peaks respectively. We employ a simple algorithm to detect the spike within each event for quantitative analysis.

(ii): Find the data point index and the value of the global minimum (purple circle). Starting from a threshold (green dotted line, $T_h = 0.25 \times \Delta G$) above this global minimum, find on either side, the first data point where the slope of the curve changes sign.

(iii): Horizontal line from this data point defines the begin (blue circle) and the end (red circle) of the spike.

(iv): Trace between the start and end points defines the spike region for ECD and position calculation.

(B) Branch length (shaded region) is calculated as follows

$$L_b = \frac{ECD_b}{ECD_{event}} \times L_{sc} \tag{3.2}$$

where, ECD_{event} and ECD_b are respectively the ECDs of the entire event and the branched region only, L_{sc} is the linear length of sc-DNA which is 514.25 nm for 3025 bp pGEM3z plasmid.

(C) Relative position Z_b (shaded region) of the branch is calculated as follows:

$$Z_b = \frac{ECD_Z}{ECD_{event}} \tag{3.3}$$

where, ECD_Z is the ECD of the event up to the global minimum.

Most of the events collected in our experiments for Fig 3.6 have single branch. In the rare cases (< 10 %) where there are events with multiple spikes, the largest spike is considered for the calculation of L_b. Fig 3.6B shows the histogram of branch lengths on sc-DNA estimated for all events from the same pore. The mean branch length was found to be 68 ± 50 nm with a maximum branch length of ~ 200 nm. This result is in agreement with the AFM studies previously reported [12]. Here we note that the events with very small spikes would be lost in the noise floor of the data. Representative events with branch lengths of 25 nm, 50 nm, 100 nm and 150 nm are shown in Fig 3.6A.

We next correlate the branch length of individual molecule with the corresponding ΔG value. In Fig 3.6C, we plot the ΔG values against the estimated branch length (L_b) for all the branched molecules. We find that the ΔG increases linearly with L_b and then saturates to a constant value (ΔG_{sat}) for all higher L_b values. We see this behavior in the raw data (Fig 3.6A) and understand it by following the argument: increasing branch lengths block larger portion of the pore resulting in increasing ΔG values, however, for molecules with branch lengths (L_b) equal to or longer than the pore length (L_{eff}), it will result in ΔG values twice that of extended sc-DNA (ΔG_1). Fig 3.6A (Top) shows the cylindrical model of quartz nanopores with a length of L_{eff} and the shaded portion is its sensing region. The sc-DNA translocating through this cylindrical volume causes the conductance drop (event) and the branches on the sc-DNA cause the spikes on top of the event. Larger the branch length traversing the nanopore larger the ΔG value of the event. As we go from left to right in Fig 3.6A, we see that with increasing branch length the event depth (ΔG) increases. This argument is also supported by the fact that the second conductance drop peak (ΔG_2), as shown in Fig 3.6B (inset, bottom) matches the saturation conductance drop value (ΔG_{sat}) for the nanopore. We confirm this behavior to be consistent with

measurements across multiple pores (see Table 3.2). By fitting the data to a piecewise function which defines the two connected straight lines with different slopes, we can

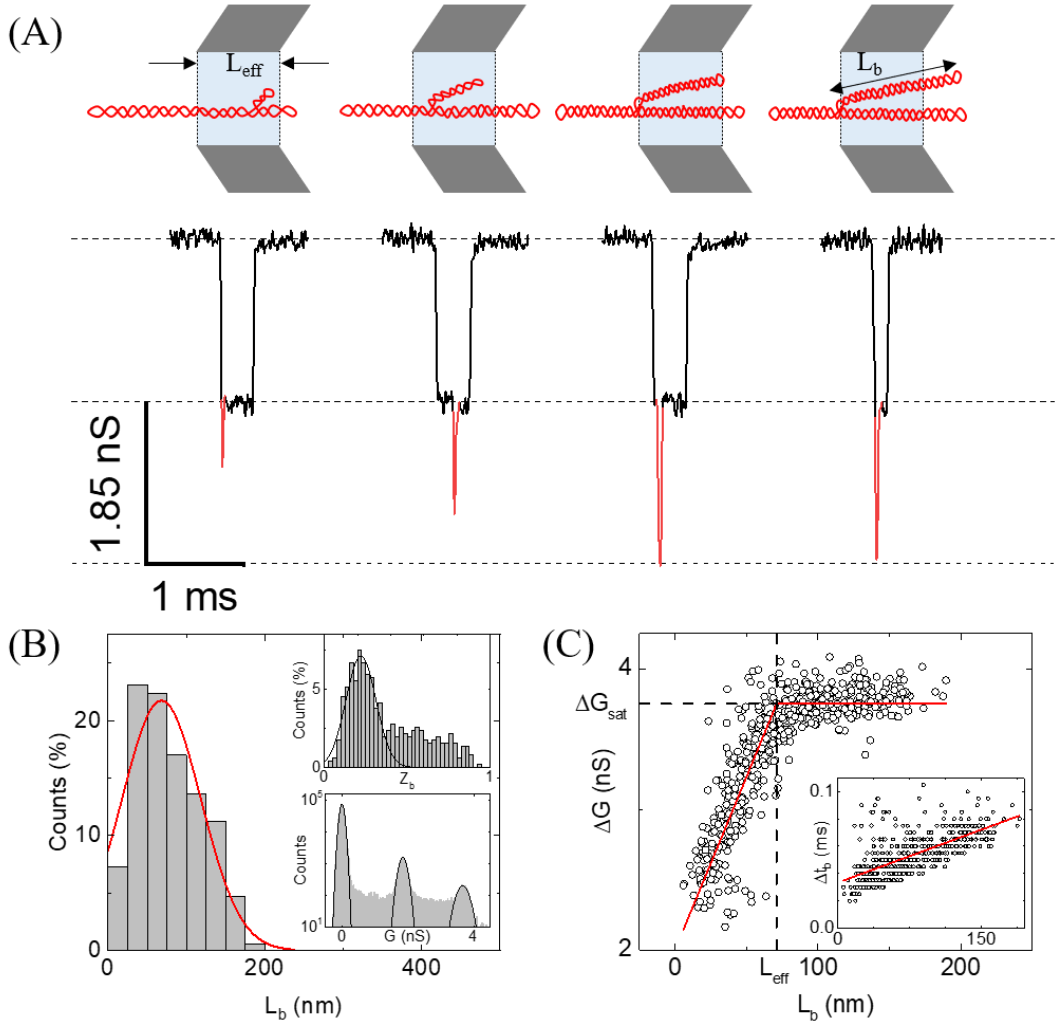


Figure 3.6: Analysis of branches in sc-DNA translocation.

(A) Typical events measured with sc-DNA showing different branch lengths translocating through the nanopore P25 (17 nm, n = 979) (left to right: 25, 50, 100, 150 nm length of the branched region). Here the sample is driven from outside of the pore to inside (left to right movement). Red traces show the current spikes due to branched regions of the sc-DNA sample. Schematics of possible DNA conformations corresponding to each event is shown on the top. (B) shows the histogram of branch length measured in the nanopore. Top inset: Histogram of relative branch positions along the length of the DNA molecule. Bottom inset: Conductance histogram for sc-DNA. (C) shows the conductance drop of the event (ΔG) plotted against the branch length (L_b). Inset: Dwell times of the branched region plotted against branch length showing a linear relationship. In Appendix 4 (Fig A4.1-A4.3) we compare the same trends across multiple other pores.

obtain the exact point at which the saturation happens. This value gives us the effective length of the pore which turned out to be 71 ± 1 nm for the pore P25 (17 nm) as shown in Fig 3.6C. This is also in close agreement with the pore length value of 63 ± 4 nm as estimated from the model of conductance drop vs pore diameter in Fig 3.3A (inset). This analysis method was applied to analyse folds in the linear DNA and obtained similar

results (See Appendix Fig A4.4). This analysis provided us a direct estimation of the effective length of the nanopore, one of the major bottlenecks in the accurate quantification of molecular sizing due to uncertainty in pore length in glass nanocapillaries [116,133].

				ΔG vs L_b for sc-DNA					
Pore ID	Pore diameter (nm)	ΔG_1 (nS)	ΔG_2 (nS)	L_b (nm)	Expected L_{eff} (nm) (3.73 x D_P)	L_{eff} (nm)	Slope1 (nS/nm)	Slope2 (nS/nm)	ΔG_{sat} (nS)
P12	21	1.6	3.2	66 ± 46	78	81	0.019	0	3.3
P19	19	2.05	4.0	66 ± 47	71	82	0.024	0	4.1
P24	13	2.8	5.6	70 ± 49	48	82	0.032	0	5.64
P25	17	1.85	3.7	68 ± 50	63	71	0.025	0	3.75
P26	18	1.9	3.8	66 ± 51	67	80	0.022	0	3.8
P29	18	1.3	2.6	65 ± 44	67	89	0.014	0	2.7
P31	19	1.1	2.1	68 ± 40	71	73	0.013	0	2.2

Table 3.2: Summary of ΔG vs L_b fitting from branch analysis for supercoiled DNA from a total of 7 nanopores

Fig 3.6C (inset) shows the variation in dwell time of the branch as a function of its length and shows a linear relationship. The slope of this curve provides the velocity of sc-DNA translocation. We found the average translocation velocity to be 12 ± 2 bp/μs when measured on multiple nanopores (N=7) with diameters of 18 ± 2 nm (see Table 3.3). This analysis helped us understand and model the relationship between event characteristics (ΔG and Δt) and polymer size in conical nanopores which had been difficult to do until now.

The supercoiled DNA can have the branch located at any position along the length of the molecule as observed from previous reports in the literature [1,7]. The branch location was estimated for each event using eqn (3.4) and plotted as branch position (Z_b) histogram in Fig 3.6B inset (Top). Here, Z_b is scaled such that Z_b =1 corresponds to the linear length of sc-DNA (514.25 nm). We observe that the location of branches on the sc-DNA is predominantly towards the beginning (peak value of 0.22 ± 0.1) of the event. The observed bias in the position of the branches is likely due to the greater propensity of the charge-dense branched region to enter the pore first. These results were tested and found consistent across multiple nanopores and the data is shown in Fig A4.1-A4.3 in

Appendix 4. From the above analysis, we understand that the sc-DNA in its natural form exists in both unbranched and branched structures with a wide range of sizes.

		Δt_b vs L_b for sc-DNA			
Pore ID	Pore diameter (nm)	Slope× 10^{-4} (ms/nm)	Intercept (ms)	Velocity (μm/s)	Velocity (bp/μs)
P12	21	2.27	0.031	4404	13
P19	19	2.22	0.034	4506	13
P24	13	3.37	0.034	2966	9
P25	17	2.58	0.033	3870	11
P26	18	2.10	0.032	4763	14
P29	18	2.35	0.030	4247	12
P31	19	3.31	0.028	3025	9

Table 3.3: Summary of Δt_b vs L_b fitting from branch analysis for supercoiled DNA from a total of 7 nanopores

3.3.3 NdeI-dependent changes in plasmid conformation

Here, we demonstrate the quantification of lin-plasmid DNA in a sample with a binary mixture of lin- and sc-DNA samples. We prepare samples containing a known mixture of linear and supercoiled forms with 0%, 25%, 50%, 75%, 100% of linear DNA. All the samples are analysed using gel electrophoresis (Fig 3.7A) and nanopore (Fig 3.7B). 100% linear plasmid shows two peaks in ΔG histograms with a first peak corresponding to the unfolded linear DNA and the second peak corresponding to folded linear DNA which also coincides with the first ΔG peak for supercoiled plasmid. However, since a known percentage of events (Table 3.1) are in the unfolded ΔG peak of a 100% linear DNA sample, we can use this correction factor to estimate the total amount of linear DNA in any sample. We use the following expression to accurately quantify the percentage of linear DNA in a mixture of plasmid conformations.

$$\%\ Linear\ DNA\ in\ a\ mixture = \%\ Unfolded\ events\ in\ a\ mixture \times \frac{100}{\%\ Unfolded\ events\ for\ 100\%\ linear\ DNA} \quad (3.4)$$

Finally, we perform quantification of NdeI dependent linearization of sc-DNA using our nanopore platform. The supercoiled plasmid was digested with NdeI enzyme for various durations of 0, 5, 10, 30 & 60 min and then the samples were analysed by both gel

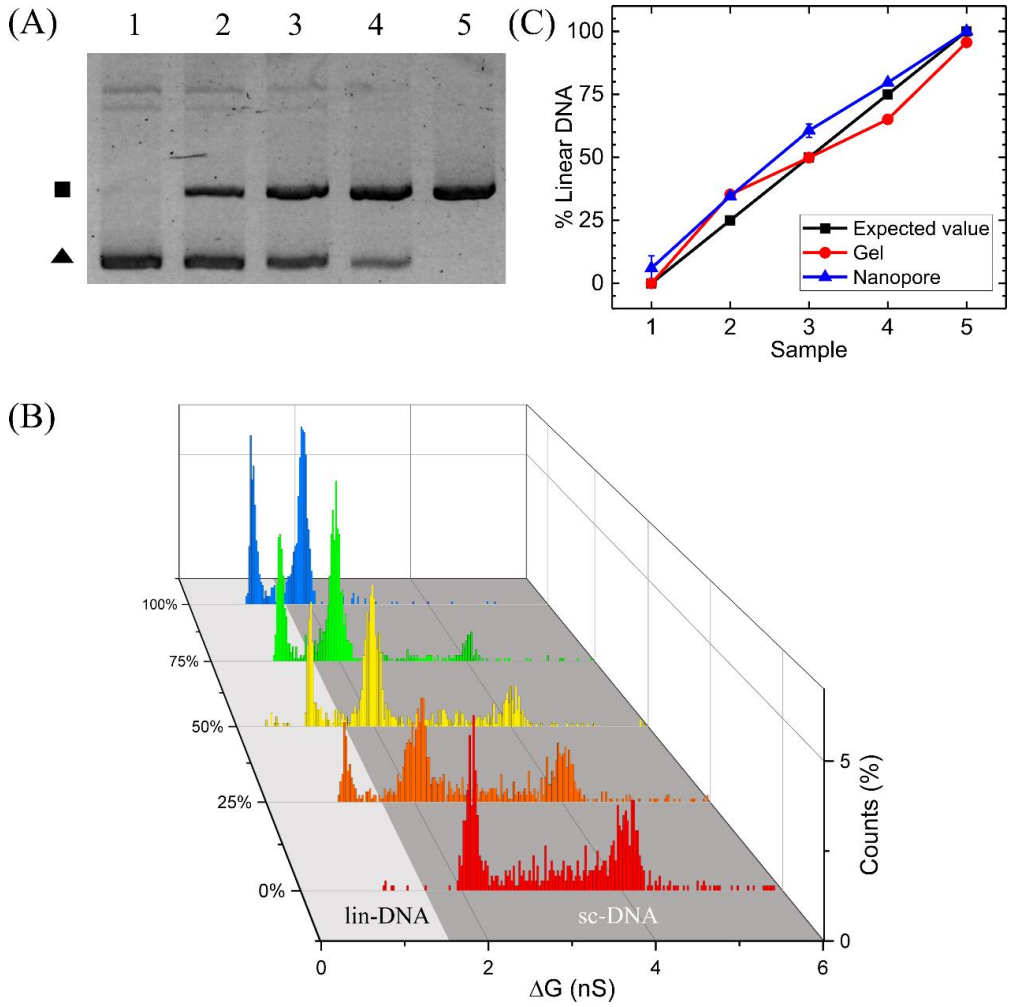


Figure 3.7: Demonstration of nanopore based quantification of linear plasmid using a known mixture of lin- and sc-DNA plasmid forms.

(A) Gel image with lanes 1-5 showing bands corresponding to supercoiled (triangle) and linear (square) plasmid in the mixed samples with 0% (completely supercoiled), 25%, 50%, 75%, & 100% (completely linearized) of linear plasmid respectively. (B) Nanopore data from P37 (19 nm) showing ΔG histograms for the same mixtures as (A). Here n = 925 (100%), 926 (75%), 872 (50%), 913 (25%), 686 (0%). (C) Comparison of the amount of lin-DNA calculated from gel band (red circles) analysis and nanopore (N=2, from P36 and P37) data (blue triangles) analysis. Black squares show the expected values.

electrophoresis (Fig 3.8A) and nanopore platform (Fig 3.8B & C). Since the NdeI enzyme is highly reactive, we chose to work at a low enzyme concentration (1/6 X) allowing us to access the intermediate reaction time points. Reaction rates at different enzyme concentrations are shown in Fig 3.9B. Fig 3.8B shows the translocation experiments performed on samples digested for 0, 5, 30 & 60 mins. All samples were measured back-to-back in the same nanopore of diameter 17 nm. The ΔG histogram for sample incubated

for 0 min (which is the undigested supercoiled plasmid) shows the primary peak at 1.32 nS and a plethora of higher ΔG events ending with a peak at 2.29 nS, similar to what was seen in Fig 3.2. After 5 min reaction, we observe the emergence of events with a mean

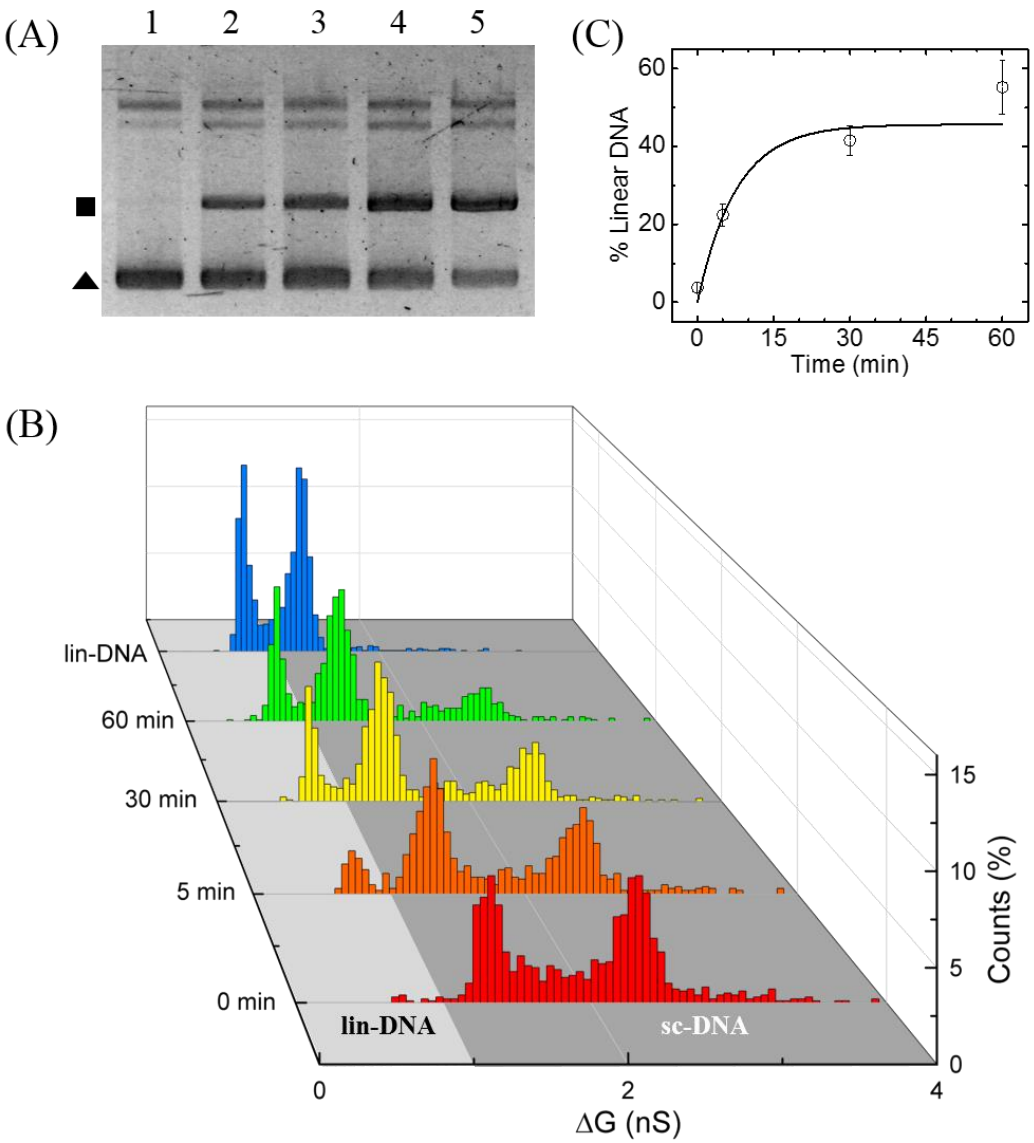


Figure 3.8: Enzymatic linearization of branched supercoiled DNA.

Time dependent linearization of supercoiled plasmid DNA by NdeI restriction enzyme was quantified using nanopores. (A) Gel image shows time-dependent linearization of supercoiled plasmid by NdeI restriction enzyme at 1/6X concentration. Lanes 1-5 show the result of restriction digestion assay when the reaction was stopped at 0, 5, 10, 30 and 60 min time points. The solid triangle marks the location of the supercoiled plasmid band and the square marks the linearized DNA band. (B) shows nanopore data of ΔG histograms for NdeI-digested samples with increasing reaction times (as shown on the left axis) obtained on a 17 nm pore (P32). The time dependent increase in the lin-DNA peak and the corresponding decrease in sc-DNA peak, is evident from the data. ΔG histogram for 100% linear DNA is also shown for comparison. Here, n = 951 (0 min), 905 (5 min), 1102 (30 min), 1018 (60 min), 1470 (lin-DNA). (C) shows quantification of lin-DNA (in percentage) produced by the enzyme, at different reaction time points as measured by our nanopore data (N=3, 2 datasets from P32 and 1 dataset from P35). The solid line is the fit to kinetics model (eqn (3.5)) giving the reaction rate constant k = 0.13 ± 0.07 min^{-1}.

ΔG of 0.77 nS confirming the formation of linearized-DNA upon NdeI digestion (as also

seen in Lane-2 of Fig 3.8A). For visual clarity, the regions in the ΔG histograms corresponding to linear and supercoiled conformations are shaded in light and dark grey respectively. With increasing reaction times, the lin-DNA peak increases, as expected. Given the percentage of linearized DNA showing up as folded (or partially folded), the correct amount of DNA in the lin-DNA peak is quantified as done in Fig 3.2B and confirmed with known mixtures of linearized and sc-DNA as shown in Fig 3.7. We quantified the amount of lin-DNA in samples of increasing reaction times to be 3.8 %, 22.4 %, 41.5 % and 55.2 % respectively as shown in Fig 3.8C and listed in Table 3.5.

Reaction time	**% events in peak1 after correction**
0 min	3.8 ± 1.4
5 min	22.4 ± 2.9
30 min	41.5 ± 3.8
60 min	55.2 ± 7.0

Table 3.4: Summary of nanopore quantification for enzyme digestion

This is the average of results obtained from 3 datasets in 2 nanopores for quantification of enzyme activity (as shown in Fig 3.8C). Two datasets were obtained on P32 and one dataset on P35.

The data is fitted with first-order rate constants, as shown in eqn (3.5) for the restriction digestion [134]:

$$[L(t)] = [SC]_{max}(1 - e^{-kt}) \tag{3.5}$$

Here k is the first-order rate constant of reaction, [L] and [SC] are the time-dependent concentrations of linear and supercoiled plasmid DNA and $[SC]_{max}$ is the maximum amount of sc-DNA that can be digested under the given conditions. The values of rate constant were found to be 0.13 ± 0.07 and 0.12 ± 0.02 min^{-1} from nanopore (Fig 3.8C) and gel data (Fig 3.9B, Black curve) respectively which are in excellent agreement with each other. Upon performing ECD analysis on these datasets, we find that the mean branch lengths (L_b) on sc-DNA remain constant as the enzyme reaction progresses confirming the well-established all-or-none mechanism [135] where the DNA cut by the enzyme removes all branches from sc-DNA. This work establishes the nanopore platform to study thermal as well as enzyme dependent conformational changes in biomolecules at the single molecule level.

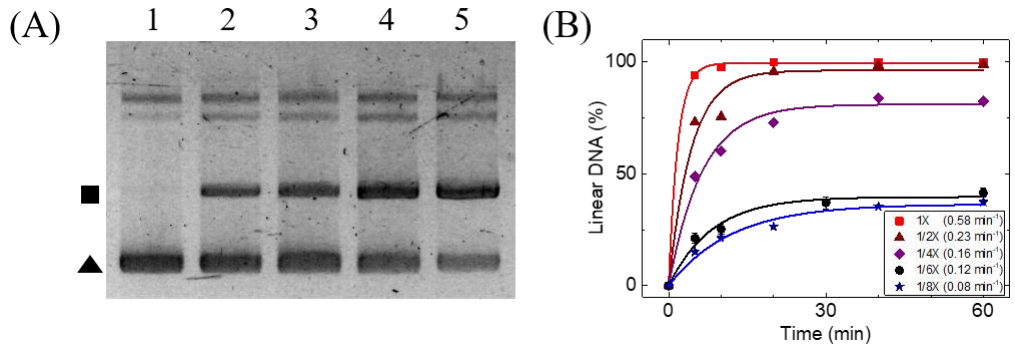


Figure 3.9: Gel quantification of rate constant for DNA digestion at different NdeI concentrations.

(A) Gel image shows time-dependent linearization of sc-DNA plasmid by NdeI restriction enzyme at 1/6X concentration. Lanes 1-5 shows the result of restriction digestion assay when the reaction was stopped at 0, 5, 10, 30- and 60-min time points. The marker triangle shows sc-DNA and the square shows linearized DNA. (Fig 3.8A is reproduced here for clarity). (B) Plot of the amount of lin-DNA as a function of reaction time for NdeI concentrations of 1X, 1/2X, 1/4X, 1/6X, 1/8X. Solid lines are the fits to kinetics model (eqn (3.5)) and the first-order rate constants calculated for each reaction condition are mentioned in their respective parenthesis in the legend.

3.4 Conclusions

In this work, we demonstrated the distinguishability of the three conformations of the plasmid DNA due to difference in their electrical signatures. We observed that supercoiled DNA has two sub-populations as seen in ΔG histogram, which were identified as unbranched (or extended) and branched supercoiled DNA. Using a customized ECD analysis, we could fingerprint the branch sizes as well as their relative positions on the molecule. An interesting outcome of this analysis was an independent measurement of the effective length of the pore which could lead to a significant technological advancement in accurate quantification of macromolecular sizes. Finally, we demonstrated nanopore based quantification of NdeI dependent conformational changes in supercoiled DNA. The values of first-order rate constants are calculated from nanopore and gel electrophoresis data and they agree reasonably well with each other. We conclude that the nanopore technique is highly promising for studying conformational changes in DNA and have potential applications in real time kinetics measurements of biomolecular reactions. The results presented here also pave the way for immediate further studies on the role of key parameters such as sequence, supercoiling density, environment, etc., on the global topology of the supercoiled DNA using nanopores.

CHAPTER 4

Detection of sub-nucleosomal structures

Abstract

The packaging of DNA, in nucleosomes, to form chromatin helps the biological cell to compact and regulate its genetic material. During the gene regulatory mechanisms, nucleosomes transiently exist in various structural forms. In this chapter, a novel, glass nanopore based investigation is shown that probes voltage dependent stability of nucleosome structures. It is shown that the canonical octasomal nucleosomes, comprising of octameric histone complexes on DNA, systematically breaks down into its partial structures, in a voltage dependent manner. This structural breakdown is attributed to the opposing electrical forces experienced by the positively charged histone protein complex and the negatively charged DNA, inside the nanopore. From the nanopore data analysis, we estimate the volumes of these various nucleosome structures. Finally, we present a model for the electric force-induced breakdown of nucleosome complexes into their sub-structures.

4.1 Introduction

Nucleosome is the basic unit of chromatin in the cell and it is formed when the DNA is wrapped around the histone octamers. It is stabilized due to the electrostatic interactions between the negatively charged DNA and the positively charged histone complex. Changes in nucleosome core particle structure [24] have attracted wide scientific interest, however, their breakdown into partial-structures and its relevance in genetic processes is less studied. It was observed that the occurrence of partial nucleosomes such as hexasomes and tetrasomes (due to loss of histone dimers) is directly related to the gene transcription levels which suggests that partial breakdown of nucleosomes could be assisting the transcription process [53,55]. The existence of sub-nucleosomal particles with a single copy of all the four core histones referred as half-nucleosomes was also suggested from genome-wide studies [57]. Single molecule studies were done to observe and quantify structural forces in various partial nucleosome species using AFM [58,59] and FRET [60] studies under varied salt conditions. They were also suggested to be the intermediates in the histone exchange pathways but the stability or kinetics of these structures is not well studied, likely due to the transient nature of these structures. In order to understand the electrostatic mechanisms contributing to the stability of these structures, biophysical experiments that directly apply forces to the nucleosome structures, need to be developed. Though the techniques such as optical and magnetic tweezers, that apply forces to underlying DNA, reveal the interactions contributing to the stability of the molecule, precisely localizing the sites where those interactions are at play has not been possible. A single-molecule analysis technique with high resolution and throughput is necessary to study the interactions responsible for stabilizing large macromolecular structures.

Nanopores have shown great potential in detecting the molecular conformations as well as in applying forces on the molecules, thereby providing us a possibility to investigate both the equilibrium structures as well as dynamic changes in it [64,67,69,71–73,78–80,117]. A quantitative analysis of mononucleosomes and its sub-components was earlier reported using nanopores [101,136]. Previous studies also demonstrated the measurement of interactions involved in stabilizing streptavidin-biotin complex [97], EcoRI-DNA complex [98]. In order to understand the interactions stabilizing chromatin, there has been a growing interest on applying this technique to study the stability of

chromatin systems but were limited to simpler systems such as distinguishing unmodified from methylated [99,100] and ubiquitinated nucleosomes [137]. These force-spectroscopy experiments were confined to sub-5 nm pore sizes where the DNA enters the pore but the bound protein is restricted from entering the pore due to mechanical confinement allowing to probe the stability of the complexes under increasing voltage.

In this chapter, we present a way to profile the structural states of individual nucleosomes (partial or full) while they freely translocate through the nanopore. Our strategy would be very helpful in studying the biomolecular complexes that are stabilized primarily due to the electrostatic interactions.

4.2 Materials and methods

4.2.1 Sample preparation

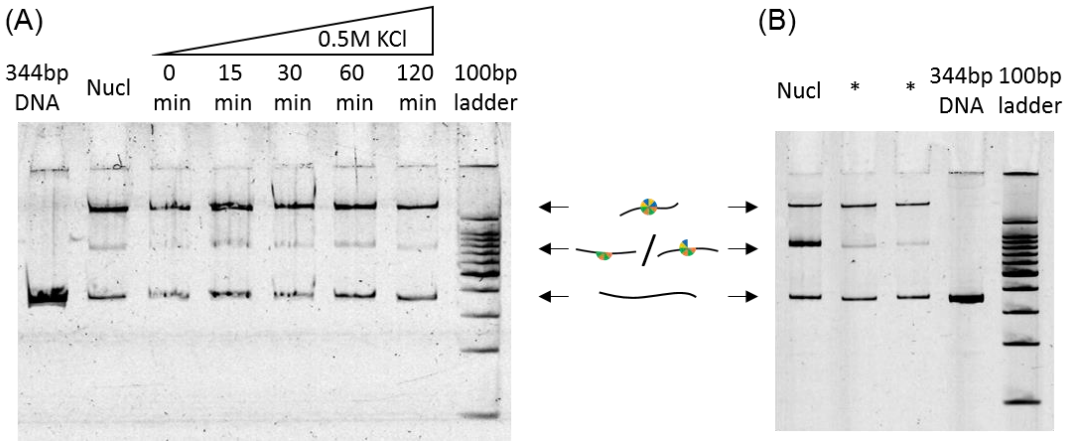


Figure 4.1: Mononucleosome reconstitution and its salt-stability check over time.

(A) Gel image showing the nucleosome sample with faint intermediate band (Sample 2). Lane1: 344 bp DNA, Lane2: Nucleosomes assembled on 344 bp DNA, Lanes3-7: demonstrate nucleosome stability in 0.5 M KCl at 0, 15, 30, 60, 120 min respectively and Lane8: 100 bp DNA ladder. (B) Gel image showing the nucleosome sample with bright intermediate band (Sample 1). Lane1: Nucleosomes assembled on 344 bp DNA, Lanes2-3 are not relevant here, Lane4: 344 bp DNA and Lane5: 100 bp DNA ladder. Sample shown in lane 1 was used for collecting the datasets shown in this figure. The schematics of biomolecules corresponding to each band are shown with arrows.

Histone octamers (16-0001, Epicypher) were made glycerol-free by dialyzing them against the nucleosome assembly buffer (NAB, 2 M KCl or NaCl, 20 mM Tris-HCl, 1 mM EDTA, 1 mM DTT, 0.5 mM Benzamidine, pH 7.5) using 3.5 kDa MWCO dialysis cups (Cat# 69552, Thermo Scientific) for 2 hrs. For mononucleosomes, 344 bp DNA containing a 1x601 sequence was PCR amplified from pGEM-3z/601 plasmid. The

mononucleosomes and the 12mer nucleosome arrays were prepared as described previously [101]. Briefly the DNA template along with histone octamers were mixed in nucleosome assembly buffer (NAB) and salt gradient dialysis was performed against low salt buffer (250 mM KCl or NaCl, 20 mM Tris-HCl, 1 mM EDTA, 10 mM DTT, 0.5 mM Benzamidine, pH 7.5) using peristaltic pump (ISMATEC) at a rate of 0.2 ml/min for 70 hours to let the salt concentration change from 2 M to about 250 mM. Finally, the samples are dialyzed against the final storage buffer (TCS buffer, 20 mM Tris-HCl, 1 mM EDTA, 1 mM DTT, pH7.5) for 18 hrs. Nucleosome assembly was tested on native PAGE gel (Fig 4.1) and then stored at 4 ^{o}C until required.

Commercial nucleosomes (with 147 bp DNA) were purchased from Epicypher (16-0009) and were directly used for our experiments (Fig 4.4A). For translocation experiments, the final concentrations of the samples in the nanopore well were 0.4 nM for 5 kb lin-DNA, 15 – 35 nM for 147 bp nucleosomes and 0.4 – 1.8 nM for 344 bp nucleosomes respectively. For checking stability of nucleosome complexes in the nanopore buffer 5 µl of sample was incubated in the experimental (NPB-0.5) buffer for different time durations (see text). Sample was then checked by loading with 20 % sucrose in a 6 % native PAGE gel (in 1x TBE for 35 min at 130 V) for monucleosomes (38 – 40 nM) and visualized using Sybr Gold staining.

4.2.2 AFM imaging

For AFM imaging of the samples, mica sheets (71853-15, Muscovite Mica V4 grade, Electron Microscopy Sciences) were initially cleaned using methanol and dried by nitrogen. Freshly cleaved mica substrates were dipped into 0.03 % (v/v %) APTES solution (50 ml methanol, 2.5 ml acetic acid, 15 µl APTES) for 10 min, then washed with methanol and dried using nitrogen flow. Nucleosome samples were diluted in 1X TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8) to a final concentration of 1 ng/µl and 50µl of the prepared sample was dropped on APTES coated mica (AP-mica) sheet and incubated for 10 min at room temperature. The surface was then washed twice with Milli Q water and dried under nitrogen. Prepared sample substrates were desiccated for 10 min to dry them completely and then imaged in air under AFM. Molecular Imaging AFM (Pico Plus - Pico Scan 3000) was operated in tapping mode (air) using TAP190-G probes

(190 kHz resonant frequency, 48 N/m spring constant, < 10 nm tip radius, Budget Sensors). The system was controlled using PicoView 1.14 software. The imaging was performed over a region of 2 × 2 μm^2 with a scan speed of 4 µm/s (line scanning rate of 1 Hz) and a resolution of 512 × 512 px^2.

4.3 Results and discussion

4.3.1 Voltage-dependent stability of 344 bp mononucleosomes

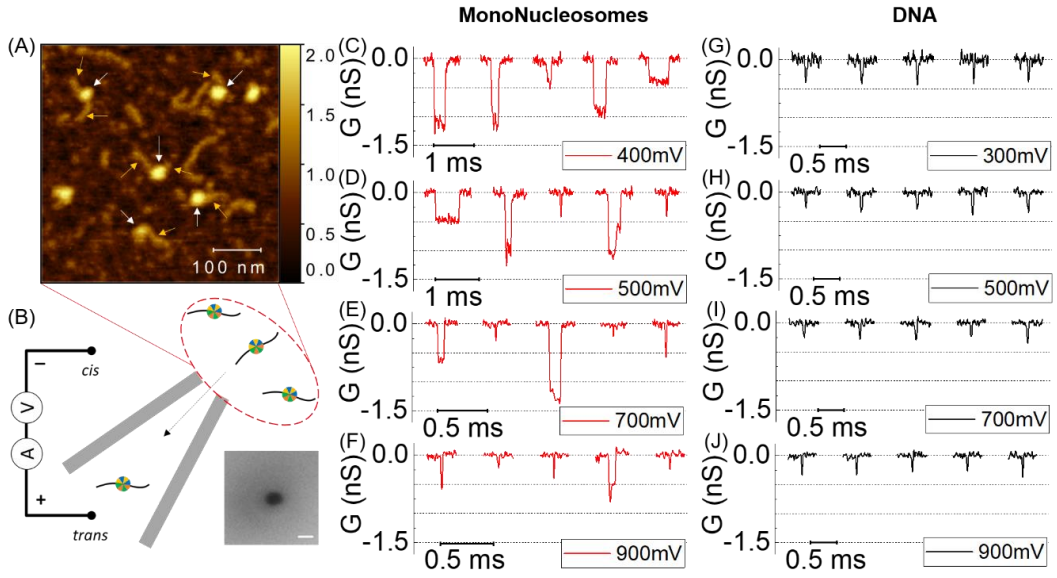


Figure 4.2: Translocation analysis of nucleosomes.

(A) AFM image of nucleosomes in dry mode showing the histone protein complexes and the surrounding DNA. The white and yellow arrows point towards the nucleosome core particle and the DNA overhangs respectively. (B) The schematic of the experimental setup showing the nanopore at tip of quartz capillary, the measurement electronics and the voltage driven translocation of nucleosome particles from cis to trans chamber through the nanopore. SEM image of the nanopore (Pore-1, 18 nm, See Table A7.1 for more details about the pores used in this study) is shown at the bottom right with a scale bar of 25 nm. (C)-(F) Representative events for nucleosome translocation at 400, 500, 700 and 900 mV respectively. (G)-(J) Representative events for 5 kb linear DNA translocation at 300, 500, 700 and 900 mV respectively.

We first show detection of mononucleosomes assembled on 344 bp DNA. These nucleosomes were assembled, tested for stability and imaged under AFM (Fig 4.2A) as well as on native PAGE (see Fig 4.1). In the AFM images we could see nucleosome core particle (white arrows) as well as the DNA overhangs (yellow arrows) on either side (344 bp DNA has one 147 bp nucleosome positioning '601' sequence with 82 bp and 115 bp

on either side) since our DNA is longer than what is required for a canonical nucleosome particle (147 bp). These mononucleosome complexes were translocated through an 18 nm pore, under an applied potential, as schematically shown in Fig 4.2B. SEM image of a typical pore used in these measurements is shown in bottom-right inset of Fig 4.2B. In a nanopore, with a stable baseline and a linear I-V curve, the nucleosome sample is added into the *cis*-chamber and the voltage is applied across the nanopore [69,138]. Nucleosomes, with net negative charge, translocate through the nanopore towards the *trans*-chamber where the positive electrode is placed. With the biasing voltage, we detect the electrical events consistent with resistive pulse sensing signatures, as the mononucleosomes translocate through the pore. The translocating particle hinder the ion-flow of the open nanopore causing transient (Δt, dwell time inside the pore) reduction in pore conductance. This drop in conductance (ΔG) is proportional to size of the translocating complex [70,80,112,139,140]. In a typical experiment we collect from few hundreds to thousands of such events to build statistics on the translocation events of sample population. In Figure 4.2C we show representative events of nucleosome translocations through 18 nm pore under 400 mV bias voltage. We see these translocation events are detected with excellent S/N ratio owing to the low-noise properties of glass nanopores [69]. The ΔG values of the translocation events correspond to the molecular volumes of the mononucleosomes (see text below). Interestingly, among these detected events for mononucleosomes, we note a population of events with ΔG of discrete lower values. Upon filtering out collisional events and electrical noise [69,99,137,138], we reason that these events with lower ΔG values correspond to analytes of smaller sizes. For comparison, we show in Fig 4.2G, representative events of 5 kb DNA translocation through the same nanopore, under the same experimental buffer and voltage, as control. These events, primarily show single ΔG level for the DNA molecules [68,69]. Note that 344 bp DNA (without the histone complex) was too small to be detected under these experimental conditions, hence the 5 kb linear DNA was used as control throughout this work. Figures 4.3A and 4.3E show ΔG histograms of all events for mononucleosome and bare DNA events, respectively, measured on the same nanopore.

We clearly see that the DNA sample has a single ΔG peak corresponding to a population of DNA molecules with similar size, whereas ΔG histograms of mononucleosome sample not only shows the ΔG peak at 1.02 ± 0.07 nS (mean of 6 different pores, see Table 4.1) but also a secondary peak at lower ΔG value, which corresponds to structures of lower

molecular volumes. We note that the events corresponding to low-molecular size structures are present in a relatively high proportion, however, no such equilibrium structures were seen in the gel images of the sample (see Fig 4.1A), when tested before and after the nanopore experiments. We reason that these multiple peaks in ΔG histograms may correspond to populations of histone-DNA complexes that were formed by breakdown due to the electrical forces on the complex, during translocating through our nanopore. The high electric field near the nanopore mouth (~ 6 x 10^6 V/m at 400 mV considering effective pore length of 67 ± 5 nm (refer Table 4.1)) applies opposing forces on the positively charged multi-component histone core (pI ~ 11) and the negatively charged DNA wrapped around the histones. The forces experienced by the wrapped & flanking DNA of our 344 bp nucleosomes inside the nanopore is estimated using the conversion factor of 50 fN/mV [141], which in our experiments range from 20 pN (400 mV) to 45 pN (900 mV). This force range is consistent with forces applied in force-spectroscopy experiments to either unwrap the DNA or break nucleosomes into its sub-structures [43,47,142,143]. We hypothesize that due to the applied electrical force on the complex, canonical nucleosomes (nucleosomes/octasomes: eight histones wrapped around by DNA) may break into its sub-structures of hexasomes (6 histones wrapped around by DNA) resulting in lower ΔG peak. These are known stable sub-structures formed during nucleosome assembly or disassembly [59,60,99,100,130,144,145]. A testable prediction from this hypothesis would be a reduction in the proportion of events in the higher ΔG peak with increasing voltage, resulting from the breaking down of intact nucleosome into its sub-structures as they experience higher pulling forces. To test this hypothesis, we performed nanopore translocation experiments at different bias voltages. Fig 4.2C-4.2F and Fig 4.2G-4.2J show the representative events, at various voltages, for translocation of nucleosomes and 5 kb DNA respectively. Figure 4.3A-4.3D and Figure 4.3E-4.3H show histograms of ΔG values from all events (ΔG histograms) at different voltages for nucleosomes and 5 kb control DNA, respectively. Note that for a given data set, all voltage experiments (for sample and control) were done on the same nanopore (under identical experimental conditions) and the order of applied voltage was randomized to eliminate any possible effects of systematic breakage of nucleosomes with time. For completion, dwell time histograms and scatter plots of ΔG and Δt for nucleosome translocations are shown in Figures A5.1E-A5.1H and Figures A5.1I-A4.1L respectively of the Appendix 5. The ΔG histograms for nucleosomes show interesting features at different voltages. At 400 mV, we find the entire sample population with in

two peaks with mean values of 0.58 ± 0.06 and 1.02 ± 0.07 nS. At a slightly higher voltage of 500 mV, we see a significant reduction in the higher peak and it continues to reduce till 900 mV where finally the peak corresponding to intact nucleosomes completely vanishes as shown in Fig 4.3A-4.3D. This observation unequivocally proves our hypothesis that the nucleosomes are indeed breaking down into its sub-structures in the presence of electric field. To our surprise, at 500 mV, we see the emergence of a third peak (population) at 0.31 ± 0.03 nS which is located at a significantly lower ΔG than the values obtained at 400 mV. Interestingly, we note that this peak, at smaller ΔG value, becomes more prominent with further increase in voltage. At 900 mV, the highest voltage applied, we see that the entire population is present only with in the lower two peaks. This suggests that the nucleosomes break down into further smaller sub-structures of histone-DNA complexes at higher voltages. These are most likely to be the tetrasomes (4 histones wrapped around by DNA) formed by further removal of histones compared to that of hexasomes. These three peaks, for the nucleosome sample, are termed Pk1 (mean ΔG_T = 0.31 ± 0.03 nS), Pk2 (mean ΔG_H = 0.58 ± 0.06 nS) and Pk3 (mean ΔG_N = 1.02 ± 0.07 nS) in rest of this work. The datasets for the complete voltage range are shown in Fig A5.2. We ascertain (in Fig A5.2G) that the 500 mV data set consistently showed the same 2nd and 3rd peaks even when this measurement was repeated at the end of the entire voltage set or randomly in between other voltages. This confirms that the nucleosome sample in the reservoir is unchanged and the emerging structures are purely due to voltage driven changes in nucleosome complex, during translocation. For comparison, we performed translocation of 5 kb DNA at various voltages to confirm that there are no voltage dependent effects. We find, the ΔG histograms for 5 kb DNA remains single peaked (Fig 4.3E-4.3H) and depends marginally on voltage with mean value of about 0.41 ± 0.04 nS at 300 mV and 0.35 ± 0.05 nS at 900 mV (Figure 4.3J) and the mean Δt value reduces from 0.08 ± 0.02 ms at 300mV to 0.06 ± 0.01 ms at 900mV (Fig A5.3). The representative events as well as population histograms show that at all voltages, DNA translocation demonstrates a single peak in ΔG histograms, on the other hand, nucleosomes re-distribute between, primarily three, different populations in a voltage dependent manner. We re-confirmed this voltage dependent re-distribution of structurally different populations in multiple nanopores (Fig A5.4-A5.5) and with independent batches of lab-assembled nucleosome samples (Fig 4.1B, Fig A5.6). Fig 4.3I shows the ΔG values corresponding to the three peaks plotted as a function of

voltage. We observe that the peaks are distinct and their positions remain unchanged with voltage. Fig 4.3J shows the ΔG histogram for 5kb DNA remains single peaked with its

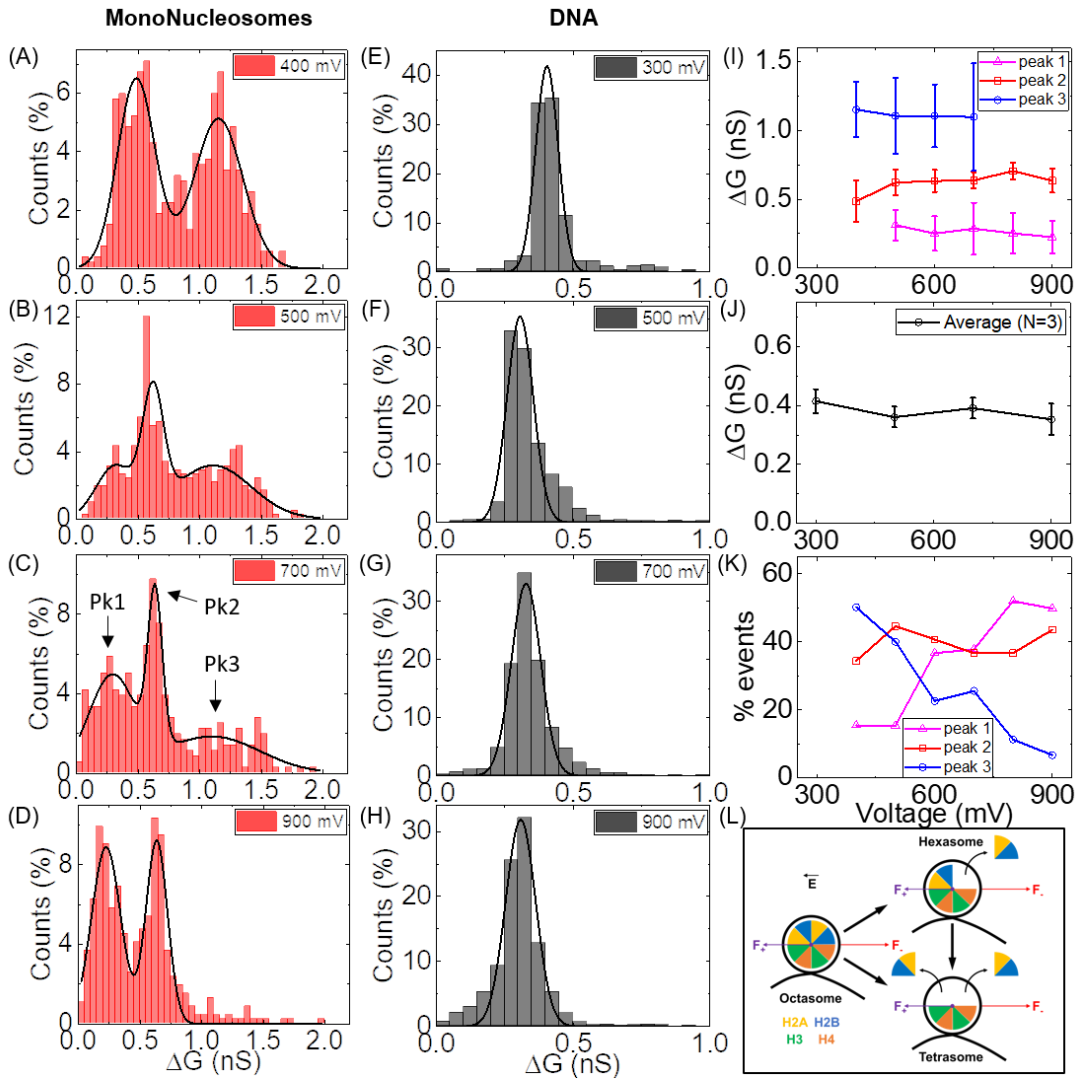


Figure 4.3: Statistical analysis of nucleosome translocation.

(A)-(D) ΔG histograms for nucleosome translocation at 400 (N = 538), 500 (N = 417), 700 (N = 360) and 900 (N = 464) mV respectively. See Figure A5.2 for data sets of all voltages measured. (E)-(H) ΔG histograms for 5 kb linear DNA translocation at 300 (N = 390), 500 (N = 1510), 700 (N = 1006) and 900 mV (N = 1259) respectively recorded using the same nanopore (Pore-1, dia = 18 nm). (I)-(J) show the trend in ΔG peak positions plotted as a function of voltage for nucleosomes and DNA respectively. Error bars in (I) are the standard deviation values of Gaussian fitting to the respective ΔG histograms. Error bars in (J) are the standard deviation values in the mean values obtained from multiple datasets (N = 3, Pores-1, 2, 16). (K) Quantification of % events in each peak in ΔG histograms for nucleosomes at various voltages. (L) Model showing the possible pathways for disassembly of nucleosome into its sub-structures in the presence of electric field. The negatively charged DNA and positively charged histones experience electric-field generated forces in opposite directions, shown by red (F_-) and purple (F_+) arrows respectively, resulting in their breakage.

position unchanged for the voltage range reported here. For nucleosomes, the voltage dependent re-distribution of sample populations estimated from the percentage of events

in the three ΔG peaks is quantified as a function of voltage in Fig 4.3K. We find that the percentage of events in peak 3 decreases and concurrently, that in peak 1 increases with voltage, whereas population in peak 2 remains constant. This indicates that with increasing voltage, the peak positions (molecular sizes of the analytes) remain unchanged, however, the population is redistributed between them. Our understanding is that the population 3 (peak 3) is either directly converted to population 1 (peak 1) or via the population 2 (peak 2) as an intermediate. Since the population ratios in the three peaks at lower voltages remained the same irrespective of the order in which the voltage was applied, indicates that the nucleosomes were broken just before entering the nanopore. Note that it is impossible to ascertain when and where along their travel through the nanopore did the nucleosome break down into its sub-structures. As the nanopore signal is directly proportional to volume of the analyte, we assign the peaks 1, 2 and 3 to the translocation of tetrasomes, hexasomes and octasomes respectively. Here, the possibility of disomes (histone dimer attached to the 344 bp DNA) is neglected, as they were reported to exist only under higher salt conditions (above 1 M) [59]. We also consider that the free positively charged histone proteins will not go through the nanopore as they will be moving in the opposite direction of the negatively charged nucleosomes.

We next use these resistive pulse sensing measurements to estimate molecular volumes of the three types of structures in our mononucleosome translocations. This volumetric quantification is done by using the 5 kb DNA data for the size calibration. The effective length of nanopore is estimated from the equation (2.18) [80,101] written in terms of conductance drop:

$$\Delta G = \gamma \sigma V / L_{eff}^2 \tag{4.1}$$

In the case of DNA, $V = \pi d_{DNA}^2 L_{eff}/4$, the diameter of DNA ($d_{DNA}$) is 2.2 nm and the value of shape factor (γ) is 1. With the conductivity (σ) of NPB-0.5 buffer measured to be 5.54 S/m and ΔG_{DNA} = 0.31 ± 0.02 nS at 500 mV, the L_{eff} is calculated to be 67 ± 5 nm which is in agreement with the model-independent estimation reported previously [138]. The volume of nucleosomal structures from their respective mean ΔG values is estimated using the same equation with the shape factor of the disc-shaped nucleosomes (a cylinder of diameter 10 nm and length 6 nm) of 1.36 (Refer Appendix 6). Table 4.1 summarizes the data for 344 bp nucleosome translocation and also shows the estimated volumes for the structures corresponding to three peaks. The average volumes obtained from multiple datasets for the three nucleosomal sub-structures (using

the ΔG values at 500 mV) are 190 ± 40 nm^3, 350 ± 60 nm^3 and 600 ± 100 nm^3 respectively. The relative volumes of octasomes, hexasomes and tetrasomes estimated here agrees very well with ratios reported in literature [59]. The molecular volumes of the octasomal structures (peak 3) measured here are in excellent agreement with earlier estimates of 532 nm^3 from X-ray studies [24,146], 580 nm^3 from AFM studies [147,148] and 509 nm^3 from planar nanopore studies [101]. The volumetric quantification confirms that the peak 3 indeed corresponds to the intact nucleosomes (octasomes). The DNA being negatively charged and the histone positively charged, they experience oppositely directed forces, shown by red and purple arrows respectively. This leads to breaking of the H2A-H2B dimers as depicted in a schematic in Fig 4.3L. Depending on the number of dimers removed, the octasomes break down into hexasomes and/or tetrasomes. The re-assembly of nucleosomes is not considered in this study as there are no free histone octamers in the experimental condition.

5 kb DNA	**Nucleosomes on 344 bp DNA**				**Estimated volumes for Nucleosomal structures**		
ΔG_{DNA} (nS)	ΔG_T (nS)	ΔG_H (nS)	ΔG_N (nS)	L_{eff} (nm)	V_T (nm^3)	V_H (nm^3)	V_N (nm^3)
0.31 ± 0.02	0.31 ± 0.03	0.58 ± 0.06	1.02 ± 0.07	67 ± 5	190 ± 40	350 ± 60	600 ± 100

Table 4.1: Volume estimation of nucleosomal structures.

The volume estimation from mean ΔG values nucleosome (344 bp) and 5 kb DNA translocation data at 500 mV. Mean and error bars represent the data obtained from 6 independent nanopore measurements (Pores1-6).

4.3.2 Voltage-dependent stability of 147 bp mononucleosomes

We re-confirmed this voltage dependent redistribution of structurally different populations in commercially available mononucleosomes assembled on 147 bp DNA (Fig 4.4 & Fig A5.7). Note that a total of only two peaks in ΔG are observed in the case of 147 bp-nucleosomes, in contrast to three peaks for 344 bp mononucleosomes (see Fig 4.3). This is possibly due to the reason that the excess linker DNA in 344 bp-nucleosomes stabilized more structures than these 147 bp-nucleosomes. The single peak in ΔG at 150

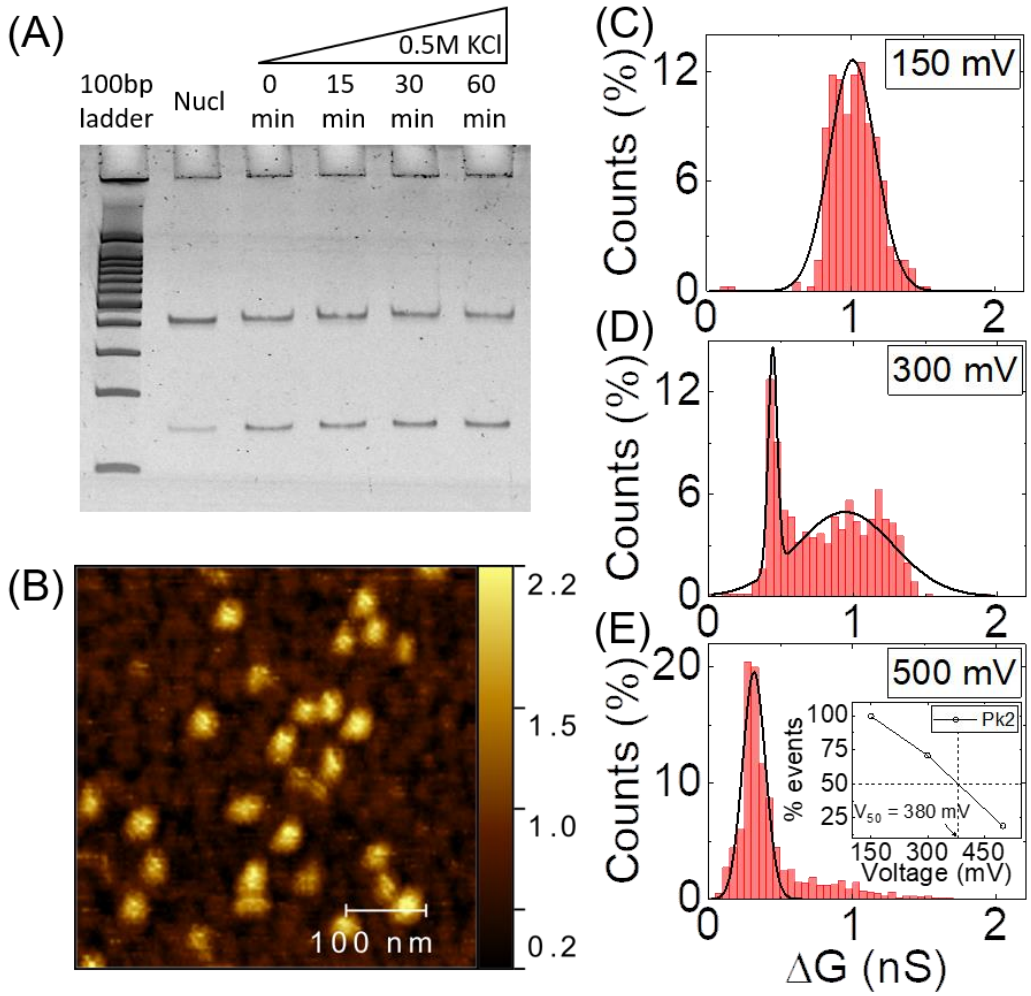


Figure 4.4: Statistical analysis of 147 bp nucleosome translocation.

(A) Gel image showing the stability of nucleosomes in NPB-0.5 buffer over time. Lane1: 100 bp DNA ladder, Lane2: 147bp-Nucleosomes, Lanes3-6: nucleosome samples in 0.5 M KCl incubated for at 0, 15, 30, 60 min respectively. (B) AFM image of 147bp-nucleosomes in dry mode. (C) – (E) ΔG histograms for nucleosomes at 150 (N = 416), 300 (N = 818) and 500 mV (N = 1614) respectively, in Pore-8 (18 nm) and in the order 300, 500 and 150 mV respectively. Inset shows the quantification of number of events in Pk2 corresponding to intact nucleosomes. From the linear interpolation, we found that the V50, the voltage at which the % events in Pk2 drops to 50 % is about 380 mV.

mV, corresponds to octasomes which breaks down at higher voltages. We found that at about 380 mV (19 pN), 50 % of the nucleosomes breakdown due to electric forces (Fig 4.4E inset). Our threshold force estimate matches with the literature value as measured from optical tweezer and AFM experiments [42,47].

4.4 Conclusions

In this work, we presented a novel nanopore-based scheme that allowed us to investigate the stability of the nucleosomal structures at the level of single nucleosomes as well as on 12mer nucleosome arrays. Firstly, the voltage dependent experiments have revealed the stability of the nucleosomes by displaying the transition of the octasomes to hexasomes and tetrasomes during their translocation. Calibrating the nanopore with DNA translocation experiment allowed us to estimate the volume of nucleosomes and its sub-structures which supports our reasoning for the assignment of lower value ΔG peaks to partial nucleosomes. The nucleosomes break due to the electrical forces in opposite directions on DNA and core histones. We understand that the removal of the single histone dimer (H2A-H2B) resulted in hexasomes and that of their pair resulted in H3-H4 tetrasomes on the DNA. In principle, smaller sub-structures (for example, disomes) are possible but they could not be confirmed under our current experimental conditions. Though the hexasomes were detected at a low voltage of 400 mV, the proportion of tetrasomes systematically increases starting from 500 mV and a significant proportion is attained only at about 900 mV. Intact nucleosomes are known to be more stable compared to the partial nucleosomes due to the higher number of contact points between the DNA and the histones. This might explain why the translocation analysis of another batch of nucleosomes that contained significant amount of partial nucleosomes also resulted in the similar proportions of events in different peaks. Due to the lower stability of partial nucleosomes, the already existing populations of these structures break down into free 344 bp or 147 bp DNA which goes undetected during their translocation through nanopores. Theoretical modelling of the electric-field induced nucleosome disassembly would help us get a better insight into the histone removal mechanism which would have implications in understanding various genetic processes including the histone exchange process. Our approach of voltage-dependent application of force using nanopores would be helpful in studying the energetics of sensitive biophysical interactions relevant in various scenarios.

CHAPTER 5

Voltage dependent stability of nucleosome arrays

Abstract

The packaging of DNA, in nucleosomes, to form chromatin helps the biological cell to compact and regulate its genetic material. Nucleosomes exist in various structural and positional combinations which allows for efficient DNA packaging and gene regulation. In this chapter, the stability of 12-mer nucleosome arrays is investigated which revealed discrete multi-level events corresponding to discrete structures on the array molecules. A detailed analysis of the voltage-dependence of these discrete levels showed that they correspond to single, multiple as well as partial nucleosomes on the array molecule. This work shows, for the first time, force dependent structural transitions in nucleosome-DNA complexes using the glass nanopore platform and opens up future biophysical studies in the context of force-based epigenetic control of chromatin structure by molecular motors.

5.1 Introduction

The chromatin folding into higher order structures helps in the genome packaging and regulation by adding physical barrier along with the individual nucleosomes in the gene regulatory mechanism [25,30,31]. The global chromatin structure is dynamic and was observed to vary depending on the stage of the cell cycle, assuming euchromatin form in the interphase and heterochromatin form in the mitotic phase. Numerous *in vitro* experiments have been performed to understand the maintenance and modulation of chromatin structure by techniques such as SAXS [37], ultracentrifugation [38], electron microscopy [39], fluorescence microscopy [40], AFM [47,48,149] etc. The inter-histone-dependent nucleosomal interactions, further modulated by epigenetic modifications, play a key role in regulating chromatin structure, as studied using bulk analysis techniques such as ultracentrifugation [36], and single-molecule techniques such as optical [42] and magnetic tweezers [46], DNA origami-based force spectrometer [49] etc. All the force-spectroscopy based techniques require tethering of molecules to surfaces which limit the number of measurements that can be made. In chromatin biology, there are a multitude of factors such as epigenetic modifications, histone variants, architecture chromatin proteins and many other that affect the structure of chromatin inside the cell. So, studying their individual contribution is not possible with these conventional techniques. We need to develop approaches that allow us to perform force measurements with high-resolution and throughput.

In this chapter, we extend the force application scheme demonstrated in the previous chapter to investigate the voltage-dependent transitions in 12-mer nucleosome array molecules, as they are driven through the nanopores. From the electrical signatures of their translocation, we also profile the structural states of nucleosomes that are positioned on a 12-mer nucleosome fiber using this technique. The experimental paradigm established in this work is promising in studying the force-dependent changes in chromatin structures and could have important applications in epigenetic studies.

5.2 Materials and methods

5.2.1 Nucleosome preparation

Histone octamers (16-0001, Epicypher) were made glycerol-free by dialyzing them against the nucleosome assembly buffer (NAB, 2 M KCl or NaCl, 20 mM Tris-HCl, 1 mM EDTA, 1 mM DTT, 0.5 mM Benzamidine, pH 7.5) using 3.5 kDa MWCO dialysis cups (Cat# 69552, Thermo Scientific) for 2 hrs. for nucleosome arrays, linearized form of 12x601-pUC18 plasmid was used (gifts from Cees Dekker's lab, TU Delft) To prepare linear DNA for assembly of nucleosome arrays, the plasmid (5102 bp) was digested with Kpn1 (NEB) and EcoRI (NEB) to create non-complimentary staggered ends that prevents re-circularization of the linearized plasmid. Digestion site was chosen such that the 12-repeats of the nucleosome positioning '601' sequence were on one half (2436 bp) of the linearized DNA with the other half being the plasmid backbone (2658 bp). The 12mer nucleosome arrays were prepared as described previously [101]. Briefly the DNA template along with histone octamers were mixed in nucleosome assembly buffer (NAB) and salt gradient dialysis was performed against low salt buffer (250 mM KCl or NaCl, 20 mM Tris-HCl, 1 mM EDTA, 10 mM DTT, 0.5 mM Benzamidine, pH 7.5) using peristaltic pump (ISMATEC) at a rate of 0.2 ml/min for 70 hours to let the salt concentration change from 2 M to about 250 mM. Finally, the samples are dialyzed against the final storage buffer (TCS buffer, 20 mM Tris-HCl, 1 mM EDTA, 1 mM DTT, pH7.5) for 18 hrs. Nucleosome array assembly was tested on agarose gel and then stored at 4 ^{o}C until required. For translocation experiments, the final concentrations of the samples in the nanopore well was 0.4 nM for 5 kb lin-DNA and 0.04 – 0.1 nM for arrays respectively. For checking stability of nucleosome complexes in the nanopore buffer 5 µl of sample was incubated in the experimental (NPB-0.5) buffer for different time durations (see text). Sample was then checked by loading with 20 % sucrose in a 1.5% agarose (in 1x TBE for 90 min at 90V) for nucleosome arrays (2.4 nM) and visualized using Sybr Gold staining (Fig 5.1).

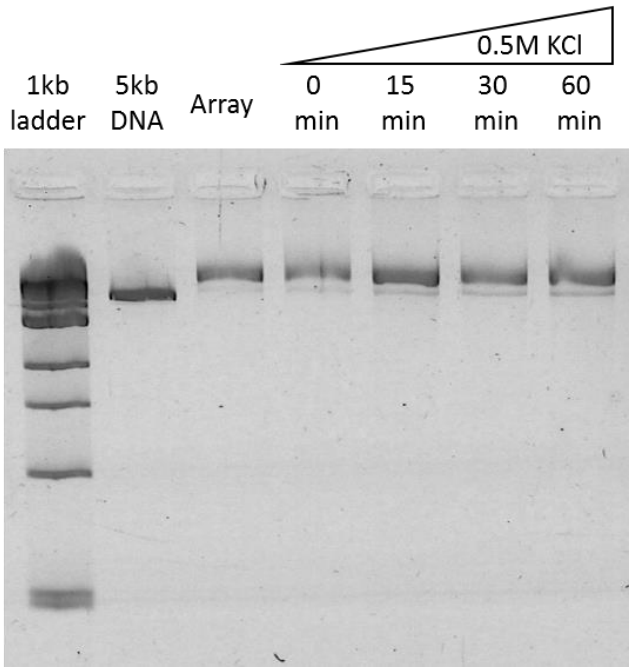


Figure 5.1: Gel image showing the stability of arrays in salt over time.

Lane1: 1 kb DNA ladder. Lane2: 5 kb linear DNA, Lane3: Arrays assembled on 5 kb linear DNA, Lanes4-7: demonstrate array stability in 0.5 M KCl at 0, 15, 30 and 60 min respectively.

5.2.2 Sample preparation for AFM imaging

For AFM imaging of the samples, mica sheets (71853-15, Muscovite Mica V4 grade, Electron Microscopy Sciences) were initially cleaned using methanol and dried by nitrogen. Freshly cleaved mica substrates were dipped into 0.03 % (v/v %) APTES solution (50 ml methanol, 2.5 ml acetic acid, 15 µl APTES) for 10 min, then washed with methanol and dried using nitrogen flow. Nucleosome samples were diluted in 1X TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8) to a final concentration of 1 ng/µl and 50µl of the prepared sample was dropped on APTES coated mica (AP-mica) sheet and incubated for 10 min at room temperature. The surface was then washed twice with Milli Q water and dried under nitrogen. Prepared sample substrates were desiccated for 10 min to dry them completely and then imaged in air under AFM (Fig 5.2). Molecular Imaging AFM (Pico Plus - Pico Scan 3000) was operated in tapping mode (air) using TAP190-G probes (190 kHz resonant frequency, 48 N/m spring constant, < 10 nm tip radius, Budget Sensors). The system was controlled using PicoView 1.14 software. The imaging was

performed over a region of 2 × 2 μm^2 with a scan speed of 4 µm/s (line scanning rate of 1 Hz) and a resolution of 512 × 512 px^2.

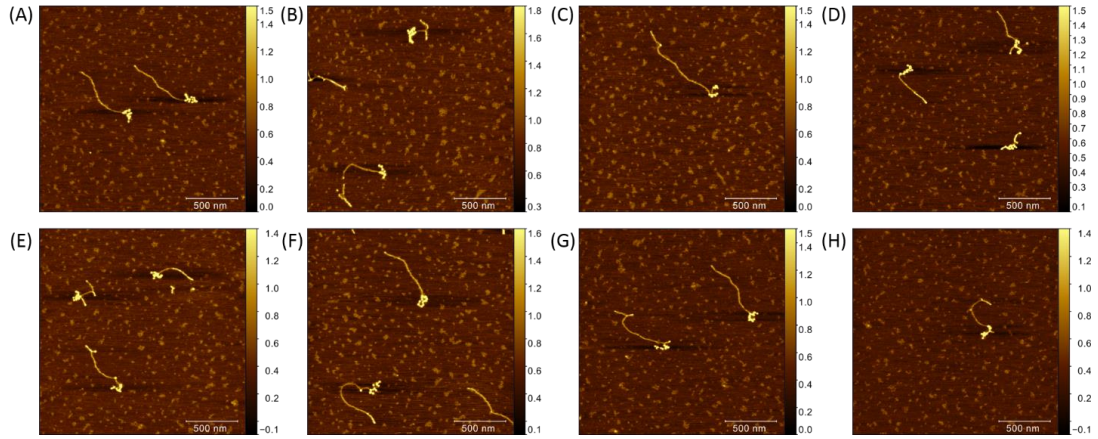


Figure 5.2: AFM images of nucleosome arrays imaged in dry mode.

5.3 Results and discussions

5.3.1 Voltage-dependent stability of nucleosomal arrays

Translocation of 12-mer arrays:

The nucleosome arrays were assembled on the linearized 5 kb plasmid DNA such that the 12 repeats of nucleosome positioning '601' sites are in one-half of the linear DNA (see schematic in Fig 5.3A). They were visualized using air mode AFM imaging (Fig 5.2 & Fig 5.3B) as well as gel imaging (Fig 5.1). In the AFM images, we see the nucleosome arrays on linear DNA with protein complexes appearing as bright dots towards one end of the DNA. We perform nanopore experiments with our *in vitro* assembled 12-mer array molecules (Fig 5.3A).

In a nanopore, with a stable baseline and a linear I-V curve, the nucleosome sample was added into the *cis*-chamber and the voltage was applied across the nanopore [69,138]. With the biasing voltage, we detected the electrical events consistent with resistive pulse sensing signatures, as negatively charged nucleosome array molecules translocated through the pore. The translocating particles hinder the ion-flow of the open nanopore causing transient (Δt, dwell time inside the pore) reduction in pore conductance. This drop in conductance (ΔG) is proportional to the size of the translocating

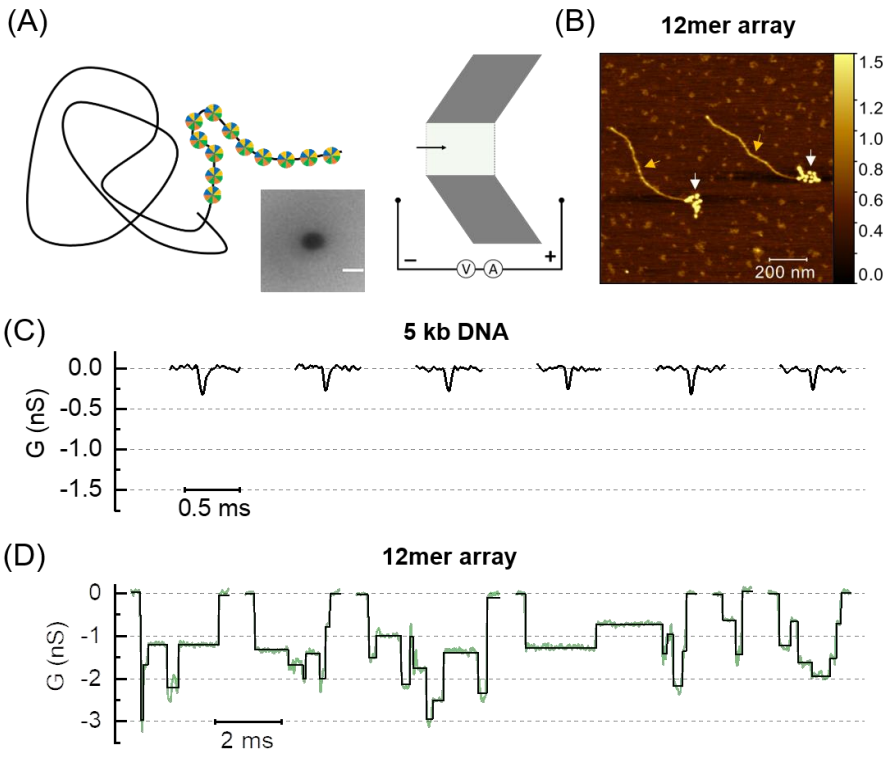


Figure 5.3: Translocation analysis of arrays.

(A) Schematic showing the experimental geometry of nucleosome arrays translocating through the nanopore. Bottom inset shows the SEM image of a typical nanopore (Pore-10, pore dia = 23 nm; scale bar is 25 nm) used in the array translocation experiments. (B) AFM image of arrays in dry mode. The white and yellow arrows point towards the cluster of nucleosome core particles and the DNA backbone respectively. (C) & (D) Representative events for translocation of 5 kb linear DNA using Pore-1 (pore dia of 18 nm) and 12-mer arrays using Pore-10 (23 nm dia) recorded at 500 mV. Data filtered at 10 kHz.

complex [70,80,112,139,140]. In a typical experiment, we collected between a few hundreds to thousands of such events to build statistics on the translocation events of the sample population. The translocation of free 5 kb DNA through an 18 nm pore at 500 mV resulted in events with a conductance drop of 0.25 nS ± 0.04 nS as shown in Fig 5.3C. The translocation of 12-mer arrays was then performed, and representative events recorded through a 23 nm pore at the same voltage are shown in Fig 5.3D. We see that the events are significantly longer as compared to bare 5 kb DNA. One possible reason for this is the presence of positively charged histone octamers in the case of 12-mer arrays. The slowdown in translocation helped us capture the intra-event structures in more detail. We also note that the events are complex that show multiple clearly-resolved and discrete levels in pore blockades. It is unlike the uniform events recorded in the case of bare DNA. In any given translocation event, we expected the array of nucleosomes to be detected as individual spikes of similar depths. However, the multi-level nature of these events was quite unexpected and interesting. This could be due to following

reasons: (I) folding of the arrays due to inter-nucleosome interactions, (II) detection of multiple nucleosomes present within the sensing region of nanopore. Case I is not likely since the size of the nanopore (~20 nm) used in this study is not big enough to allow the translocation of nucleosome aggregates. Also, the weak [42,45,49] inter-nucleosome interactions are expected to break due to the electrical forces at the nanopore (see discussion section for an estimate on force magnitudes). Case II, would imply distinct levels in the events with specific ΔG values correspond to multiple nucleosomes in the nanopore. We quantify the event levels to obtain a better insight into the structures corresponding to individual levels as well as the role of the applied electric field on the integrity of these structures.

5.3.2 CUSUM algorithm for fitting events with levels

Once an event is detected and recorded with the padding on either side using analysistranslocation.vi, the CUSUM algorithm [150] which is adopted from OpenNanopore.m [151] is used to detect the levels and fit them. Input parameters for level fitting analysis are expected change in level value (δ), positive detection threshold (h) and standard deviation of baseline (σ). The optimum value of h is considered to be δ/σ.

The algorithm identifies the time point (t_i) after which the mean value of the signal, y(t) changes from μ to μ ± δ where δ is the expected shift in level. In other words, $y(t_{i-1} < t \leq t_i)$ belongs to a gaussian distribution with mean value μ and $y(t_i < t \leq t_{i+1})$ belongs to the distribution with mean value μ ± δ. Here, i = 1 to n, where n is the total number of change points and $t_0 = 0$ and t_{n+1} = size(y). In order to check if the values are going away from a mean value μ to μ ± δ, it is reasonable to compare the values with the midpoint between μ to μ ± δ which is μ ± δ/2. Mathematically, the shift in level is detected by continuously monitoring the points where $y(t) > \mu + \delta/2$ or $y(t) < \mu - \delta/2$ and calculating the differences between the current value and the midpoints in both directions, sp(t) or sn(t) respectively which are the likelihood ratios. They are defined in such a way that the expected changes would lead to positive quantities.

$$sp(t) = y(t) - \left(\mu + \frac{\delta}{2}\right) \tag{5.1}$$

$$sn(t) = \left(\mu - \frac{\delta}{2}\right) - y(t) \tag{5.2}$$

The decision is taken at $t=t_{gi}$ when the decision function(gp(t) or gn(t)) raises beyond a threshold value (h) that a change in level has taken place. The decision function resets to zero value when it becomes negative. This retains its sensitivity to detect a change over the length of the event.

$$gp(t) = \sum_{t=t_{g_{i-1}}}^{t_{g_i}} \max(sp(t), 0) \tag{5.3}$$

$$gn(t) = \sum_{t=t_{i-1}}^{t_i} \max(sn(t), 0) \tag{5.4}$$

The actual change point (t_i) would be the time point where the cumulative sum value Sp(t) or Sn(t) is minimum between t_{gi-1} and t_g depending on whether $gp(t_{gi})$ or $gn(t_{gi})$ is greater than h respectively. The decision function decides that a change is detected and re-initializes the algorithm but the cumulative sum decides the change time point (t_i).

$$Sp(t) = \sum_{t=t_{i-1}}^{t_i} sp(t) \tag{5.5}$$

$$Sn(t) = \sum_{t=t_{i-1}}^{t_i} sn(t) \tag{5.6}$$

$$t_i = \begin{cases} argmin\{Sn(k) : t_{g_{i-1}} < k < t_{g_i}\}, & gn(t_{g_i}) > h \\ argmin\{Sp(k) : t_{g_{i-1}} < k < t_{g_i}\}, & gp(t_{g_i}) > h \end{cases} \tag{5.7}$$

'*argmin*' is the mathematical form to represent the index of the minimum element in a set. Once t_i is found, the mean value of the levIl i is calculated to be mean of $y(t_{i-1} < t < t_i)$. The mean value of the last level which is the padding is calculated to be the mean value of $y(t_n < t \leq t_{end})$.

$$\mu_i = \frac{\sum_{t=t_{i-1}}^{t_i} y(t)}{t_i - t_{i-1}} \tag{5.8}$$

Once a change point is detected, the algorithm is re-initialized and repeated which makes sure that the algorithm remains sensitive to the changes in distribution of y(t) over the complete duration of the event with the given set of parameters.

We applied this algorithm to perform level fitting to quantify certain parameters from the events as discussed in the following section.

Quantification of multi-level events:

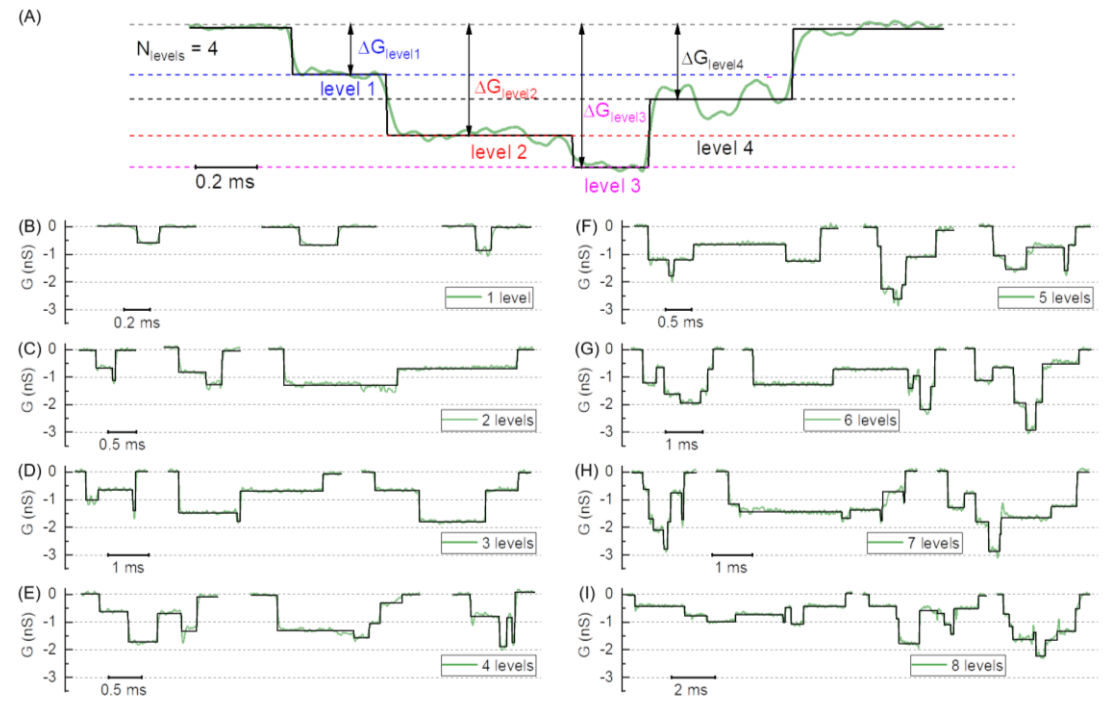


Figure 5.4: Multi-level event analysis.

(A) A typical array translocation event with 4 levels (N_{levels} = 4) is shown labelled with its individual levels, level 1, level 2, level 3 and level 4 corresponding to their ΔG values as ΔG_{level1}, ΔG_{level2}, ΔG_{level3} and ΔG_{level4} respectively. (B)-(I) Representative events along with their level fits shown with increasing number of levels from 1 to 8 respectively (filtered at 10 kHz). Statistical analysis of number of levels per event (N_{levels}) and level-depths (ΔG_{level}) in array translocation events.

We quantified the multiple levels in the events for 12-mer sample. Histograms of pore conductance (G-histograms) are typically used in literature to identify multiple blockade states in nanopore data. However, the G-histograms are susceptible to giving different weightage to different dwell times of the levels in the events. To avoid this, we performed level-fitting to individual events to extract the number of levels per event (N_{levels}) and the ΔG value for each such detected level (ΔG_{level}). The events were fitted using CUSUM-based algorithm [151]. Fig 5.4A shows a typical multi-level event demonstrating the

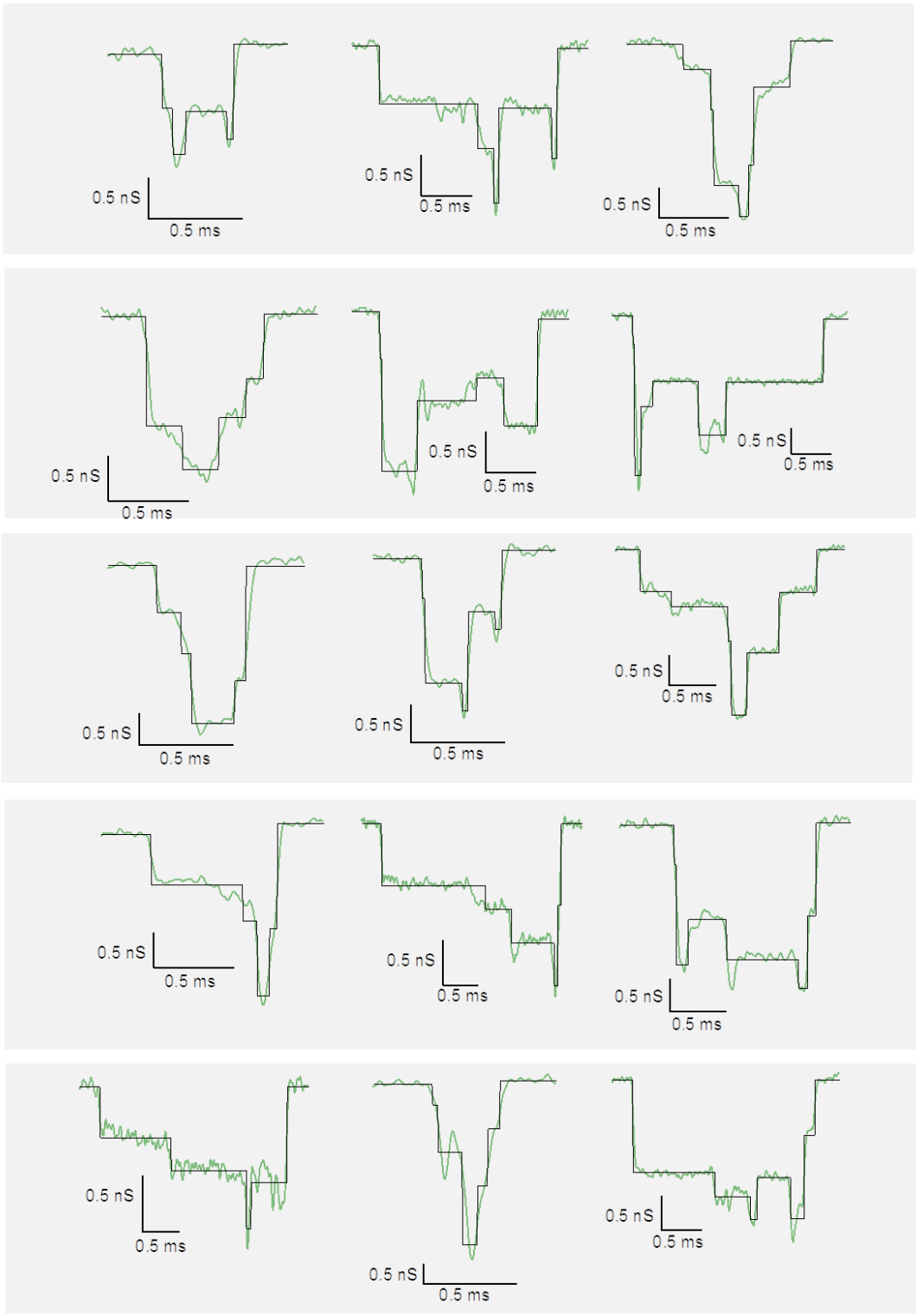


Figure 5.5: Library of multi-level events from 12mer translocation.

Representative events along with their level fits as recorded at 500 mV. Data collected using 20210823-p62-p2000f-02 (23 nm).

definitions of parameters, N_{levels} and ΔG_{level} which are obtained from the level fitting procedure. Application of the level fitting algorithm to a typical multi-level event data

(green curve) resulted in the fitted event (black curve). The individual levels are labelled with their respective ΔG_{level} markers. Figures 5.4B-5.4I show examples of events with varying complexity (N_{levels} = 1 to 8) in their structures which correspond to the heterogeneity in the local structures on arrays. At 500 mV, almost 50 % of array molecules, translocating through the pore, show multi-level events and a library of them is shown in Fig 5.5 & Fig A7.1. In contrast, the control DNA of 5 kb length shows minimal folded events (Fig A7.3E-A7.3H). These multi-level events, as seen in the 12-mer sample, establish our method of quantifying the complex signatures that contains information about the spatial and structural details of nucleosomes in the array.

Voltage dependence of nucleosome array translocations:

We next test the effect of applied voltage on the multi-level events of the 12-mer arrays. Fig 5.6A-5.6D shows representative events of 12-mer translocation through a 23 nm glass nanopore (see Fig A7.2 for representative events at complete voltage range). Interestingly, we find that the array molecules continue to show deep multi-level events at voltages up to 500 mV. However, above 500 mV, we no longer see the complex multi-level events, and most of the events are of a single level. To see the statistics of these multi-level events, we extract the events from the current time series, plot the histograms of the number of levels (N_{levels}) per event and compare them (Fig 5.6E-5.6H) for each voltage. The N_{level} histograms show the number of different local structures present on each array molecule which was not accessible from the G-histograms. At lower voltages, we find that events display a large number of levels, sometimes up to 10. The voltage-dependent changes in N_{levels} histograms show that the percentage of multi-level events reduces from about 50 % at 300 mV to 4 % at 900 mV. Also, the number of discrete levels per event dropped significantly from 300 mV to 900 mV. We find that for bare DNA translocations, most events are of a single level at voltages 500 mV and above, with only about 20 % of events showing multiple levels at 300 mV (see Fig A7.4A-A7.4D). This corresponds to the folded form of DNA, as reported earlier in the literature under similar experimental conditions [152]. Although the analysis of the patterns in N_{levels} histograms qualitatively indicates changes in local structures on arrays during their translocation, further quantitative analysis is necessary to understand the nature of these structures. For this purpose, we analyzed the statistics of the level-depths (ΔG_{level}) obtained directly from the CUSUM-based level fitting. The ΔG_{level} histograms, plotted in Fig 5.6I-5.6L, quantify size of local structures observed on each array, whereas G-

histograms provide only the ensemble average information. G-histograms help quantify events with recurring shapes and depths that result in multiple clearly resolved peaks. Heterogeneity in sample, like the 12-mer array sample used in this study, can arise from the fact that the nucleosomes can occupy up to 12 positions in different combinations (full & partial occupancy of 12 nucleosome positioning sites on the template DNA) as well as nucleosomes, themselves,

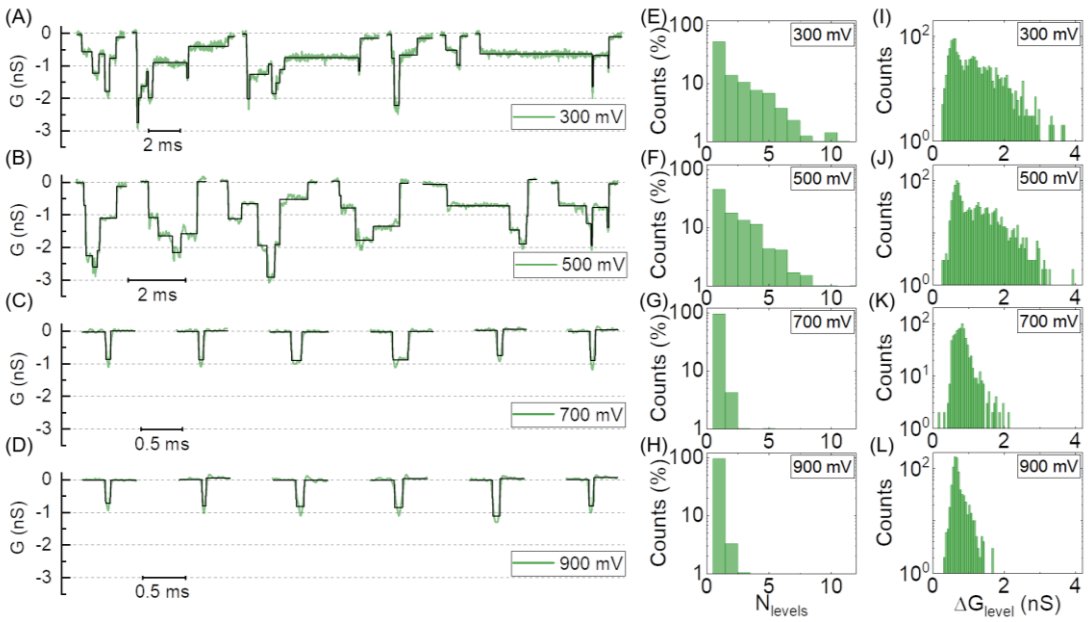


Figure 5.6: Statistical analysis of number of levels per event (N_{levels}) and level-depths (ΔG_{level}) in array translocation events.

(A)-(D) Representative events along with their level fits (black) for array translocation recorded at 300, 500, 700 and 900 mV respectively in Pore-10 (23 nm dia). (E)-(H) compares histograms of number of levels per event (N_{levels}) for array translocation at 300 (N = 564), 500 (N = 534), 700 (N = 929) and 900 mV (N = 1063) respectively. (I)-(L) shows ΔG_{level} histograms for array translocation at 300, 500, 700 and 900 mV respectively.

can be structurally complete or partial [59,60].

The ΔG_{level} values quantify these possible local structures present on each translocating molecule. The ΔG_{level} histograms in Figure 5.6I-5.6L show that the sample displays both deep levels corresponding to larger structures as well as a large variability in ΔG_{level} values. This corresponds to heterogeneity in the molecular sizes of the structures on the array molecule. We find that the depth of ΔG_{level} reduces upon an increase in the bias voltage from 300 mV to 900 mV. This observation is consistent with what we see from G histograms (Fig A7.3A-A7.3D). In the measured voltage range, the ΔG_{level} histograms for 5 kb DNA (Fig A7.4E-A7.4H) showed no voltage dependence, as expected. Since ΔG_{level} directly corresponds to the molecular size of the translocating molecule, a reduction in ΔG_{level} values (at higher voltage) would indicate the breaking down of these

complexes into smaller sub-structures. Note that the ΔG_{level} values of multi-level events up to 500 mV as well as the single-level events at higher voltages remain higher than the ΔG values for the control bare DNA translocation. This indicates the presence of nucleosomal sub-structures on the DNA even at voltages as high as 900 mV. These results were reproduced using multiple pores (Fig A7.5-A7.7) and with independent sample preparation of arrays (Fig A7.8).

To quantify the molecular size of individual nucleosomal structures in the 12-mer array, it is important to read out the structural-state of each nucleosome in the array. It can be achieved by converting the individual ΔG_{level} levels to molecular volumes. To calibrate ΔG_{level} to its corresponding molecular size, we next performed translocation of individual mononucleosomes and observed their structural transitions as a function of voltage.

Interpretation of multi-levels for nucleosome array translocation:

The typical changes in the global conformation of a nucleosome array during various stages of the translocation process are depicted in Fig 5.7A. To understand the nature of the local structures, present on the translocating array molecules, we correlate the ΔG_{level} of these multi-level events to the ΔG values obtained in the mononucleosome experiments (see Fig 4.3). In Fig 5.7B, the vertical red dotted line shows the ΔG_N value corresponding to the mononucleosome (octasome) (Peak 3 of 1.02 nS in Fig 4.3A).

The blue and black dotted lines show the estimated integral ΔG values for 2 (ΔG_{2N} = 2.04 nS) and 3 nucleosomes (ΔG_{3N} = 3.06 nS), respectively. We estimated that a maximum of three nucleosomes could fit in the nanopore with a sensing length (L_{eff}) of 67 ± 5 nm (see Table 4.1). At 500 mV, we find that most of the data points lie within the black dotted line (ΔG_{3N}), which supports our understanding that there are no more than 3 consecutively positioned nucleosomes that are detected at any instant during the translocation. It also confirms that the 12-mer array fiber is in its extended conformation with the nucleosomes separated apart as the force experienced at this voltage exceeds the inter-nucleosome interaction force. The changes in local conformations of arrays are summarized in the schematic shown in Fig 5.7B. The events showing discrete levels lower than ΔG_N are reasoned to be due to partial structures of nucleosomes (hexasomes and tetrasomes shown as magenta and orange dotted lines, respectively). The events with discrete levels in between the red-blue and blue-black regions, have ΔG values which are

non-integral multiple of ΔG_N. These, are possible, only due to various sequential

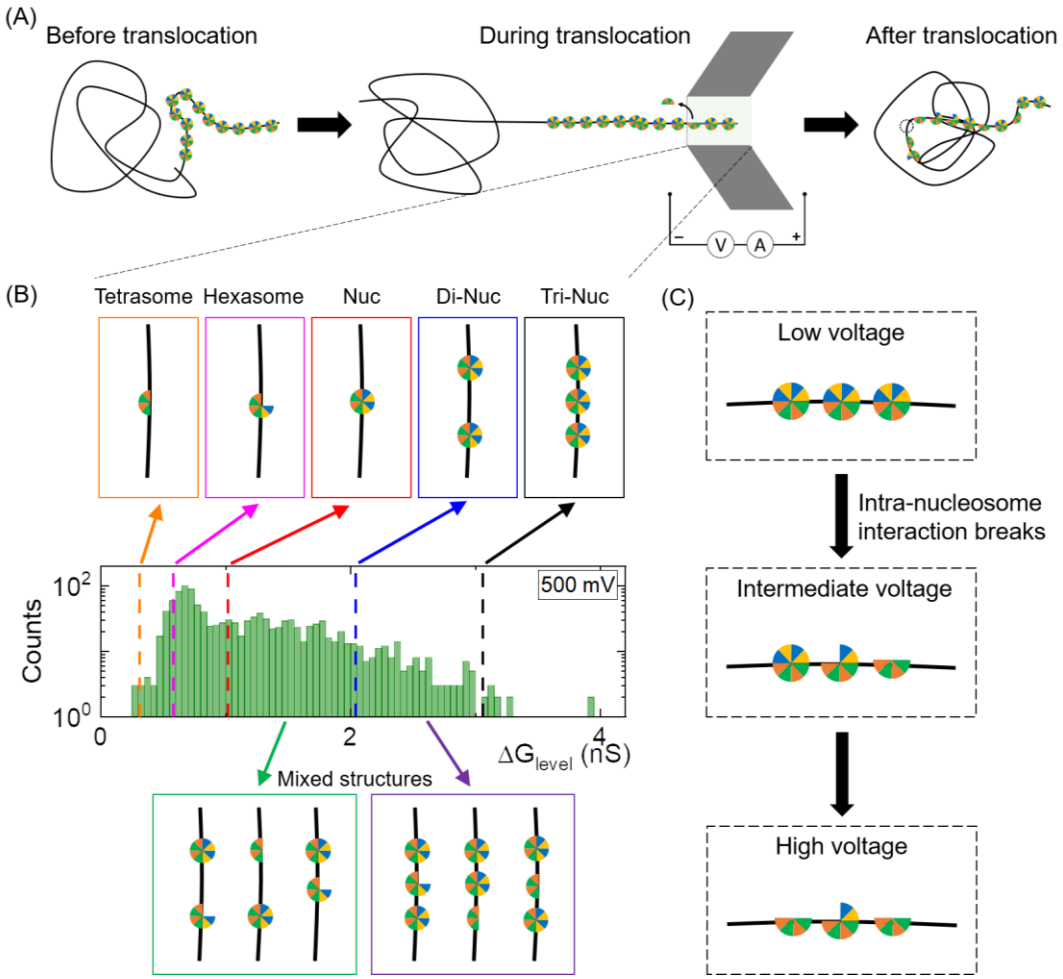


Figure 5.7: Mechanistic model of the array conformational dynamics during its translocation.

(A) The nucleosome array conformations during the process of its translocation, also depicting the stage where the nucleosome structures are modified due the electric-field induced breakdown. (B) shows ΔG_{level} histogram for 12-mer sample measured at 500 mV. The red vertical dotted line indicates the average conductance for full mononucleosome (ΔG_N). The blue and black vertical dotted lines indicate the expected conductance drop for two (ΔG_{2N}) and three nucleosomes (ΔG_{3N}) respectively. The orange and magenta vertical dotted lines indicate the average conductance drops for tetrasomes (ΔG_T) and hexasomes (ΔG_H) respectively (top). The region corresponding to mixture of full, partial and no nucleosome structures is shown by green and purple arrows and are schematically shown (bottom). (C) Schematic showing changes in the array conformations as a function of voltage.

combinations of full, partial and loss of nucleosomes as shown in Fig 5.7B. The voltage dependence of ΔG_{level} for array translocation (Fig 5.6I-5.6L) indicates that with an increase of voltage from 500 mV to 900 mV, the applied force was sufficient to break some of the nucleosomes into their partial structures (Fig 5.7C). We note that a typical array molecule displays a complex arrangement of levels corresponding to partial and full nucleosomes. This points to the complexity ($1 \leq N_{levels} \leq 8$) of histone structures on individual array molecules (Fig 5.4B-5.4I).

5.4 Conclusions

In this work, we presented a novel nanopore-based scheme that allowed us to investigate the stability of the nucleosomal structures on 12mer nucleosome arrays. In our 12-mer array translocation experiments we find a rich structural complexity in terms of multiple well-defined discrete levels. These events did not show any distinct levels for free backbone DNA (the 2.7 kb of DNA template without nucleosome positioning sequence). This is because the backbone DNA (would take about half the dwell time for 5 kb DNA (0.08 ± 0.02 ms at 300 mV) which is close to the detection limit of our amplifier and hence did not result in significant number of data points recorded for the 12-mer array events. The understanding obtained from the mononucleosome experiments made it possible to observe and interpret the heterogeneous multi-level events to be a result of a wide variety of local and global structures on our model chromatin system (12mer nucleosome arrays). The deeper events were reasoned to have been caused by the detection of a sequence of nucleosomes. The application of CUSUM-based level fitting analysis to the multi-level events helped us confirm that they are indeed caused by the translocation of arrays in various structural states resulting from many combinations of partial and full nucleosomes. Voltage dependence experiments allowed us to coarsely manipulate these global structures finally leading to breaking of nucleosomes on the arrays resulting in reduction of event complexity, quantified as the number of levels per event (N_{levels}) which reveals the local heterogeneity in the arrays. This could be useful in determining the regions of the genome with uniformly and non-uniformly positioned nucleosomes as they would result in simple (lower N_{levels}) and complex (higher N_{levels}) event signatures respectively. At voltages beyond 500 mV, there is a significant reduction in levels arising from the intact nucleosome translocation telling us that a significant number of nucleosomes on the arrays are broken down at those forces similar to the case of mononucleosomes. Our experiments did not show signatures of inter-nucleosome interaction possibly because of the high forces generated by the applied voltages, compared to their interaction forces. The direct detection of the higher order structures arising from such interactions could be done using pores slightly bigger than those used in this study as the translocation of nucleosome aggregates should be entropically favorable. Confirming whether the nucleosomes are conferred with additional stability

due to the presence of neighboring nucleosomes in the context of arrays compared to isolated nucleosomes demands fine-tuned voltage-dependent studies, which are currently on-going. In summary, we have shown that the nucleosomal sub-structures formed on arrays due to high electrical force experienced by them during their translocation generates a distinct quantitative signature in terms of its conductance drop being a non-integral multiple of the conductance drop due to full nucleosomes.

Further optimization of the technique could help us in studying nucleosomes with epigenetic modifications. The voltage-dependent scheme for probing stability of biomolecular complexes is unique as we get access to a different pathway where the forces are applied directly to the histone complexes rather than what is explored in the optical tweezer based pulling experiments where the forces are applied to the underlying DNA. The information about the possibility and the propensity to form these sub-structures could help us better understand the local and global conformations of the chromatin. For a more realistic picture of the chromatin, extension of this study to the context of nucleosomes with epigenetic modifications is desired.

5.5 Preliminary studies on PRC2-DNA binding

It is found that the chromatin folding at higher levels require proteins called the chromatin architectural proteins [50]. Polycomb repression complexes is an important family of proteins belonging to this class and especially, the Polycomb Repression Complex 2 (PRC2) spreads the methylation marks on the chromatin that leads to gene silencing [153–155]. It was recently discovered that the PRC2 can bind to DNA even in absence of nucleosomes and lead to its bending or looping which are important mechanisms of DNA folding [156–158]. The interaction between PRC2 and DNA or chromatin is not well studied and there is scope for interesting observations. Hence, we intend to carry out experiments using nanopore to understand PRC2 binding to DNA and later extend these studies to explore more about modes of PRC2 binding to chromatin to study their relevance in compaction mechanisms. In this section, we present our preliminary results on nanopore analysis of PRC2-DNA complexes.

5.5.1 Sample preparation

Polycomb repressive complex 2 (PRC2 from Active motif Cat # 31387) and linear pUC18-12X(601) 5kb plasmid (a gift from Cees Dekker's lab) was used in the binding experiment. The reaction buffer (50mM Tris Hcl (pH 7.5), 100mM potassium chloride, 2.5mM Magnesium chloride, 0.1mM Zinc chloride, 2mM β-Mercaptoethanol, 0.1mg/ml Bovine Serum Albumin, 5% (v/v) glycerol) was prepared, to which PRC2 and DNA was added. The plasmid was linearized by digestion with Kpn1 (NEB) and EcoRI (NEB) to create non-complimentary staggered ends that prevents re-circularization of the linearized plasmid. The 5 kb linear DNA and PRC2 was added in different ratios of 2.5 nM: 62.5 nM, 2.5 nM: 125 nM and 2.5 nM: 250 nM respectively to check their binding. 10ul of the reaction mixture was incubated at 30 °C for 30 min. The binding was analyzed on a 0.7% agarose prepared in 1X Tris Borate EDTA (TBE) against same buffer for 90min at 130V as shown in Fig 5.8A.

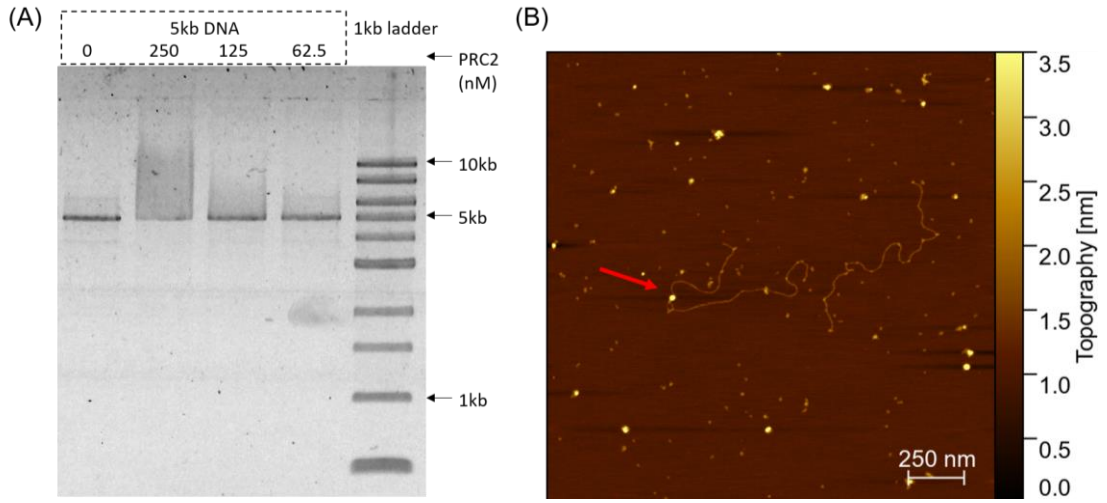


Figure 5.8: Characterization of PRC2 binding to 5 kb linear DNA.

(A) Lanes 1-4: Titration of 12mer array with various concentrations of PRC2 respectively as shown on top of the gel image. Lanes 5-8: Titration of 5 kb DNA with various concentrations of PRC2 respectively as shown on top of the gel image. (B) AFM image shows the PRC2 (pointed by red arrows) binding to DNA.

5.5.2 AFM imaging

For AFM imaging of the samples, mica sheets (71853-15, Muscovite Mica V4 grade, Electron Microscopy Sciences) were initially cleaned using methanol and dried by nitrogen. Freshly cleaved mica substrates were dipped into 0.03 % (v/v %) APTES

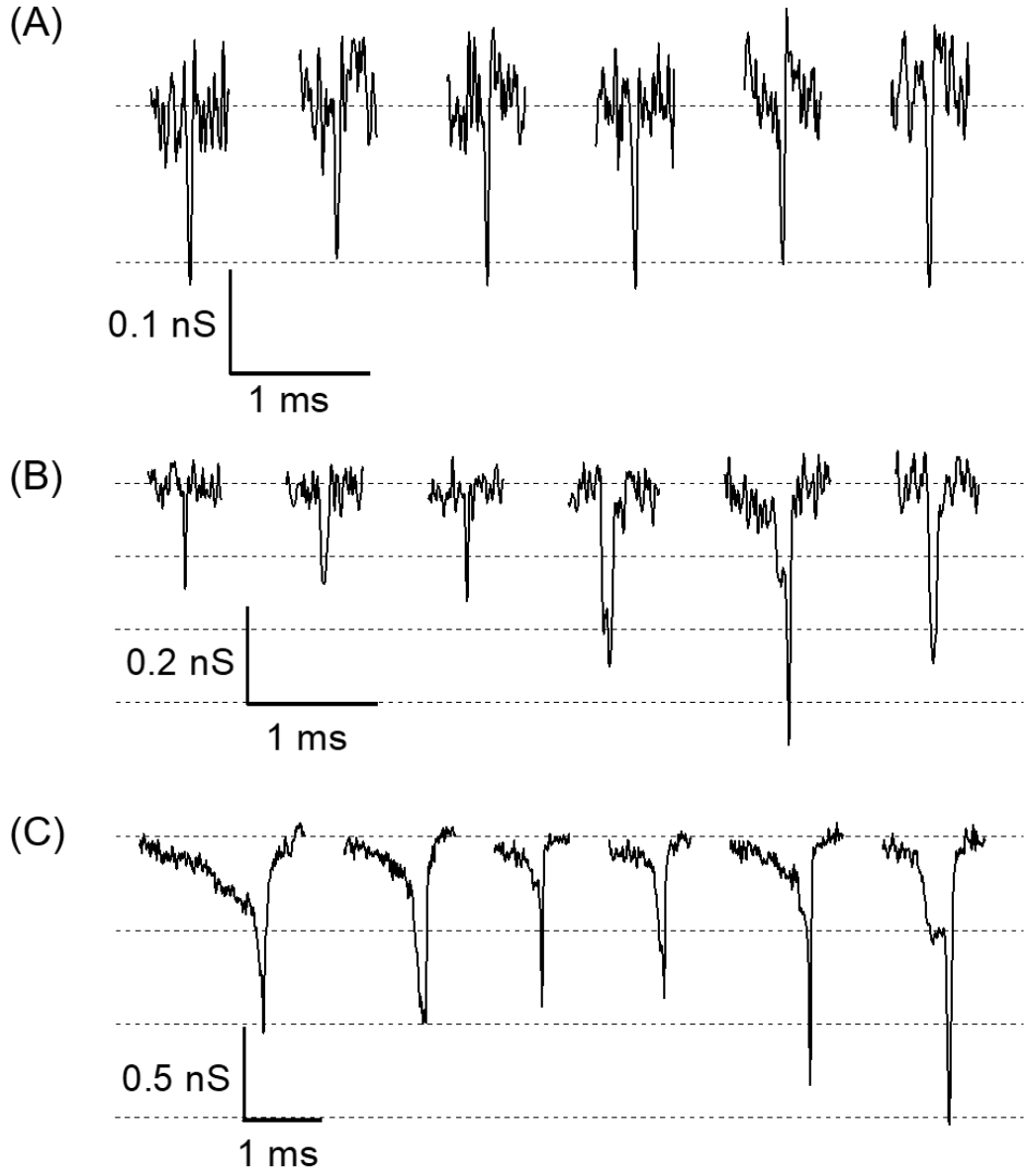


Figure 5.9: Nanopore translocation of PRC2-DNA complexes

(A) shows the representative events corresponding to 5 kb DNA translocation and (B) and (C) show the representative shallow and deep events respectively recorded in case of PRC2-DNA complex translocation. The experiments were performed in 0.25 M KCl at 900 mV and the data was low-pass filtered at 25 kHz.

solution (50 ml methanol, 2.5 ml acetic acid, 15 µl APTES) for 10 min, then washed with methanol and dried using nitrogen flow. 1 µl of DNA-PRC2 reaction mixture (ratio of 2.5 nM:250 nM) sample was diluted in 100 µl PRC2 buffer and mixed well by tapping. 30µl of the prepared sample was dropped on APTES coated mica (AP-mica) sheet and incubated for 10 min at room temperature. The surface was then washed twice with Milli Q water and dried under nitrogen. Prepared sample substrates were desiccated for 10 min to dry them completely and then imaged in air under AFM (Fig 5.8B). Molecular Imaging

AFM (Pico Plus – Pico Scan 3000) was operated in tapping mode (air) using TAP190-G probes (190 kHz resonant frequency, 48 N/m spring constant, < 10 nm tip radius, Budget Sensors). The system was controlled using PicoView 1.14 software. The imaging was performed over a region of 2 × 2 μm^2 with a scan speed of 4 µm/s (line scanning rate of 1 Hz) and a resolution of 512 × 512 px^2.

5.5.3 Nanopore analysis

We performed titrations of DNA with various concentrations of PRC2 in the ratios of 2.5 nM: 62.5 nM, 2.5 nM: 125 nM and 2.5 nM: 250 nM and checked the complex formation using gel electrophoresis as shown in Fig 5.8A. We observed that at a low PRC2 concentration of 62.5 nM, there is no binding but at 250 nM, most of the DNA is bound and shifted. The binding of PRC2 to 5 kb DNA was also confirmed by AFM imaging as shown in Fig 5.8B where the red arrow point towards the PRC2.

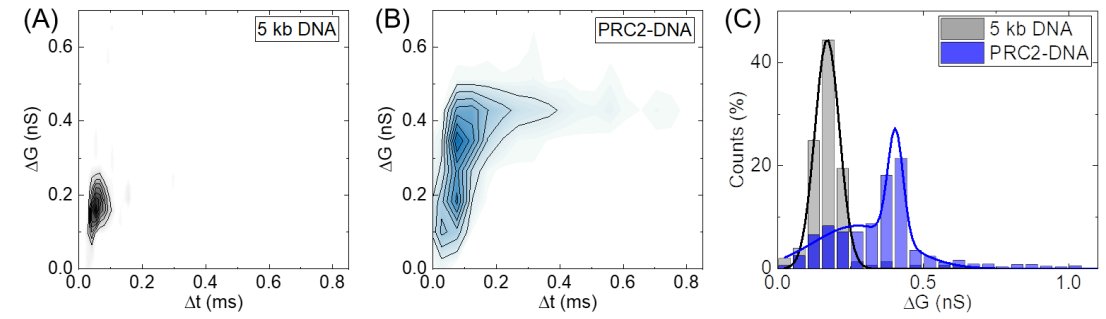


Figure 5.10: Statistical analysis of PRC2-DNA translocation.

(A) & (B) show the ΔG-Δt scatter plots for 5 kb DNA and PRC2-DNA complex. (C) shows the comparison of ΔG-Δt histograms for 5 kb DNA and PRC2-DNA complex.

Next, we performed the nanopore experiments to detect PRC2-DNA complexes. All the translocation data shown here was recorded at 900 mV. The translocation events corresponding to 5 kb DNA is shown in Fig 5.9A which are quite uniform and shallow with short duration spikes. In the case of PRC2-DNA complex, there are two types of translocation events – shallow (Fig 5.9B) and deep (Fig 5.9C). The scatter plot of ΔG-Δt shows a huge spread with greater ΔG and Δt values in case of PRC2-DNA complex (shown in Fig 5.10B) when compared to 5 kb DNA (shown in Fig 5.10A). The ΔG histogram for DNA shows a single peak located at 0.17 ± 0.04 nS (Fig 5.10A). On the other hand, the ΔG histogram for PRC2-DNA complex shows two peaks located at 0.3 ±

0.2 nS and 0.40 ± 0.03 nS respectively. The data clearly shows a good contrast between PRC2 bound DNA and the 5 kb DNA. The huge spread in the scatter plot of PRC2-DNA complex most likely represents the heterogeneity in the number of PRC2 bound to the DNA sample. The events with ΔG values in different ranges are shown in different columns. The shape of the events corresponding to bare DNA (Fig 5.9A) is symmetric triangular whereas the population in second ΔG peak of PRC2 bound DNA (Fig 5.9C) are asymmetric. Further experiments involving the volumetric quantification of these complexes would help us identify the modes of PRC2 binding to DNA and also help us find the cause of asymmetry in the event shapes.

CHAPTER 6

Summary and future directions

In living systems, it is observed that the structure of an entity is directly related to its function. This relation is clearly seen starting ranging from the organ shapes required to perform their distinct functions, to the different cell types varying in shapes that allow them to perform distinct functions, to the structures of biomolecules such as proteins and DNA that allow them to involve in specific interactions and forms the basis of the lock and key model. This relation is valid in many more scenarios but we will limit our discussion to the case of chromatin folding. Before the cell division, the chromatin is highly compacted to form chromosomes so that the mitotic spindles can pull them apart and segregate them more efficiently. In a cell, the open extended form of chromatin called euchromatin allows expression of genes but the closed compact form results in gene repression. The chromatin allows for lot more functions to take place by letting the nucleosomes exist in multiple states such as hexasomes and tetrasomes. These structures have particular functions in favouring the transcription process. Similarly, the other interesting phenomena which are the supercoiled structures have huge relevance to their functions. The supercoiling in the closed DNA molecule seem to favour the communication between the distant DNA sites required for gene regulation and genetic recombination. The branched structures are thought to be more efficient in these processes as the distant sites can contact each other by sliding of the branches on the DNA. These are some of the facts showing the correlation between structure and function of an entity in biology. This gives us an interesting perspective into how we can better understand and investigate the nature around us, that studying a structure could tell us about its function or vice versa.

In this thesis, we have studied important aspects of chromatin folding such as supercoiled, and nucleosome structures. In chapter 2, we have introduced the nanopore technique and demonstrated its characterization required for high resolution studies of biomolecular structures.

In the Chapter 3, we have shown that nanopores can distinguish the different conformations of plasmids – linear, circular and supercoiled. Though they have the same ECD values, they are clearly distinguishable from their conductance drop values. The supercoiled being the most compact and have branches, they generate higher dG values compared to circular DNA which in turn generates the value that is double that of linear DNA. Then we investigated the heterogeneity in the supercoiled DNA structures, and shown the single and multi-branched structures. Though the quantification done in the thesis was limited to the longest branches that generate the deepest signals, there is scope for extracting more information about these structures by extending the analysis techniques to quantify all the branches on each DNA molecule. We have also shown the quantification of the restriction digestion process in converting the supercoiled to linear forms and obtained accurate kinetics of the reaction comparable to the standard gel electrophoresis technique. There is still more to be understood about the physics of supercoiling, especially, it is not yet known as to when the plasmid forms the branched structures. This is due to the technological limitations to observe the branches on DNA. Though the gel electrophoresis is extremely good at resolving the changes in linking number of 1, it cannot give information about the branched and unbranched structures. Electron microscopy and AFM can be used to visualize the branched structures, but the surface effects could interfere with the conformations assumed by the plasmid and makes the readout unreliable. Apart from the surface artefacts, it is also difficult to quantify the length of the branches on DNA. Many single-molecule based studies using fluorescence microscopy and magnetic tweezers have studied the supercoils on the DNA, but they are performed on linear DNA tethered on the surface and the supercoils are generated either by mechanically twisting the DNA or by using an intercalant that unwinds the DNA leading to the twisting of DNA. This generates coils on the DNA that are studied as a function of force, DNA sequence, etc. The branched structures have not been studied in these experiments. Nanopore is highly appropriate to study such systems and the phenomena by letting us study the molecular structures in their free flight translocation.

In Ch4, we have demonstrated a novel approach to perform the bond rupture experiments using nanopores where the nucleosomes were broken into their sub-structures. Nucleosome being the basic unit of chromatin is extremely important as it helps in packaging the long DNA inside the nucleus. This is possible due to its stability acquired by the electrostatic interactions but there is enough evidence that the nucleosome stability

is significantly modulated upon post-translational modifications which are relevant in several molecular mechanisms including the processes that decide the cell-fate. Quantifying the stability changes due to these modifications is important as it gives us an insight into how it affects the binding of other proteins that needs access to the nucleosomes thereby affecting the respective cellular processes. It would also help us in developing sensors to detect the modifications that will help us read the histone code on single molecule that is difficult to characterize using the existing technologies.

In Ch5, we have analyzed the nucleosome arrays by translocating them through the nanopore. Contrary to our expectation of obtaining a spike for each nucleosome on the array, the generated event signatures have multiple-levels and have high event-to-event heterogeneity. By implementing a event level fitting algorithm, we extracted information about the commonly occurring levels in various events. This information along with what we learnt from the previous chapters regarding the sensing length as well as the nucleosome instability at higher voltages, helped us interpret the data on arrays. We concluded that the multi-levels signify the nucleosomes being in a combination of full or partial structures. This technique could be applied at lower voltages to study the transition of arrays from compact to extended forms. The important factor to consider while performing this experiment is the sticking of densely charged histones of nucleosomes to the capillary surface. This can be overcome by coating the nanopore with inert molecules such as PEG or lipids which was already demonstrated by other groups. These experiments are important as the study of proteins leading to chromatin compaction and the kinetics of compaction and decompaction that are relevant in defining and limiting the timescales of the mechanotransduction processes where the gene regulation is affected by the external stimuli is crucial. We can understand the fundamental limits on the timescales for externally induced gene regulation. This could have implications in understanding the speed of cell cycle and cell differentiation.

Later part of this chapter discusses some of the plausible extensions of the work presented in this thesis.

Future possibility-1: Study supercoiling density vs branch length

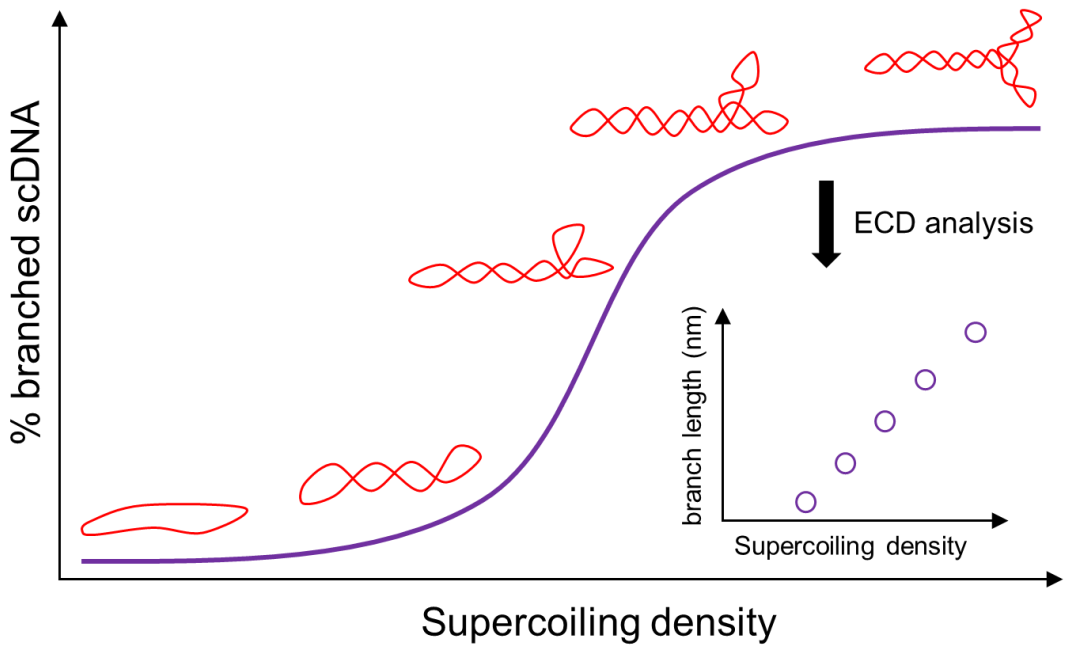


Figure 6.1: Proposal to study the scDNA at various supercoiling densities.

Main figure shows the percentage of branched ScDNA estimated from the relative number of translocation events in the second peak of conductance drop histogram. Inset shows the branch length estimated by ECD analysis for each species of varying supercoiling density.

Currently, we have no information about the turns as the ΔG for scD is exactly same as cirD. Getting information about multiple parameters from a single quantity (ECD) may not be feasible. We also do not know how the branch size is related to the linking number. I think in order to estimate the linking number, we should perform experiments with many samples of scD with known linking number and measure the branch sizes to obtain the correlation between them. There are some computational papers on branched DNA but the relation between branch length and linking number is not yet explored or mentioned so far. Topoisomerase is used to modulate the supercoiling density of DNA. The circular closed plasmid is treated with varying amount of intercalant so that it attains varying amount of supercoiling. Then it is relaxed with topoisomerase and then the intercalant is removed. Now the DNA gets back its supercoils. 2D gel electrophoresis is done with a gradient in intercalant concentrations which can resolve such small changes in linking number. The branched scDNA structures are not yet detected in vivo. If we find some proteins that selectively bind to the tri-junctions, it would be a simple way to detect these branch nodes and their dynamics could be observed, telling us about the gene

regulation mechanism. When the plasmid is relaxed (σ=0), it is in open form. As the magnitude of σ increases, the supercoils starts forming and the number of turns increases making it more compact. At what point of σ does the plasmids start forming the branches is not yet known. This can be studied by preparing scDNA of different σ and analysing their structures using nanopore, we would know when the transition of events happens by looking at their ΔG values, moving from ΔG_1 peak to ΔG_2 peak of the ΔG histogram. Branches have been studied computationally and found to occur at high salt concentrations as the DNA charges need to be screened for them to overcome the repulsion to come closer to each other for forming branched structures. At low salt conditions, the twisted DNA tends to remain in open form due to electrostatic repulsion. Another interesting question would be to know what factors decide the number of branches on a given plasmid.

Future possibility-2: Energetics of nucleosome sub-structures

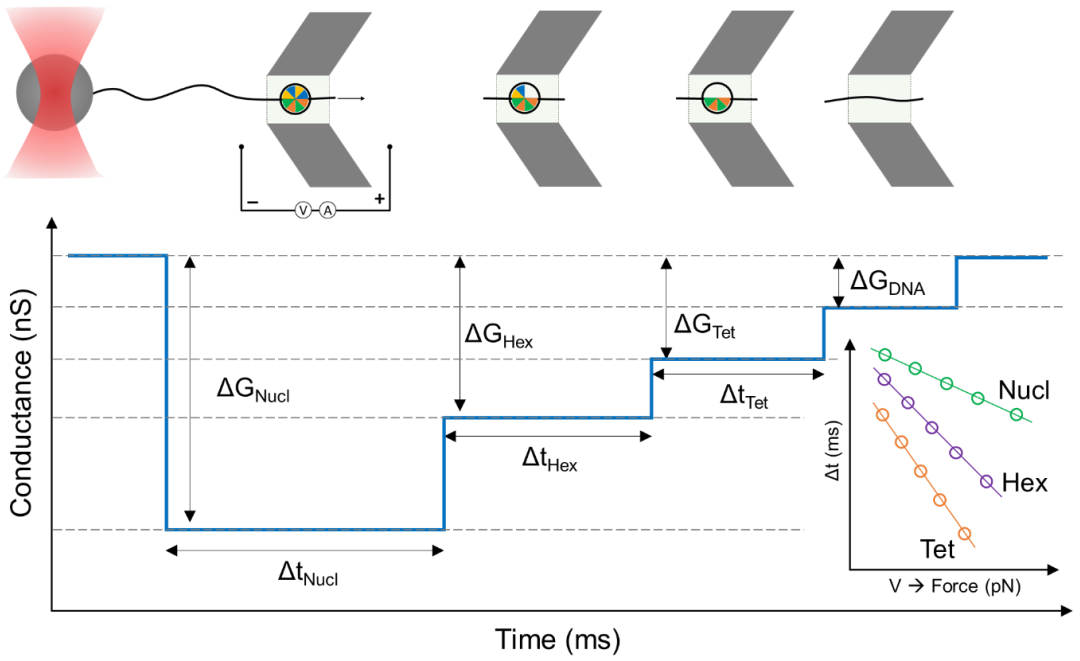


Figure 6.2: Proposal to measure the dynamics of nucleosome structures using LOT-Nanopore hybrid setup

A nucleosome-tethered bead is trapped in LOT and the nucleosome is pulled into the nanopore by application of a voltage V. At a given voltage, the conductance change is observed over time letting us measure the lifetimes of each species.

Many techniques such as AFM, optical tweezers, FRET, etc., are available to study the energetics of biomolecular complexes. In all these techniques, molecules have to be

labelled with specific molecules and anchored to the substrate. But there is a significant amount of work done in performing force spectroscopy without anchoring them using nanopores. Such experiments were performed using both biological nanopores and planar silicon nitride nanopores. The measurement of energy landscape for STV-Biotin interaction was done by Vincent Tabard Cossa using planar nanopore. We would like to use the similar technique of nanopore force spectroscopy to measure the interaction between nucleosomes and architectural proteins. Force spectroscopy allows us to probe the interactions between molecules. This could be addressed using an LOT-Nanopore hybrid setup. Trap a bead with a mononucleosome. Place the mononucleosome inside the nanopore and observe the conductance drop due to mononucleosome. Now increase the voltage gradually, and see at what voltage and in what duration, does the nucleosome break into its sub-structures like the hexasomes and then the tetrasomes and finally the bare DNA. This experiment will show the reduction in conductance drop as a function of voltage. This will also give information about the energetics and kinetics of the nucleosome breaking in multiple steps. From our experiments and with known understanding about the formation of nucleosome structure, we concluded that the nucleosomes break by the sequential removal of H2A-H2B dimers one by one. But the strength between each pair of histones and with the DNA in the nucleosome is not yet known. It is important to experimentally demonstrate which histones are broken at what point of the nucleosome dissociation.

Future possibility-3: Measuring inter-nucleosome interactions

At the lowest level in hierarchy, there are interactions between the neighboring nucleosomes. Then it is the interactions between the nucleosomes and the CAPs that lead to compaction or decompaction. Protein complexes like HP1, PRC2, etc. cause the compaction of chromatin by bridging the two nucleosomes and brings them closer. There are other proteins such as RCC1 which binds to the acidic patch of the nucleosome and blocks this important region from letting other nucleosomes access it, leading to the decompaction of chromatin. In events such as gene silencing, the chromatin compaction plays a key role while in events such as gene expression or DNA replication, the chromatin decompaction is important. Here, we discuss the experimental proposal of using the same experimental platform for measurement of forces between nucleosomes.

Here we exploit the fact that nanopore can distinguish conformational changes in the biomolecular complexes. Consider a di-nucleosome where two nucleosomes are separated by a linker DNA and they reach their equilibrium state. In the complexes where the two nucleosomes do not interact, they remain in extended form. When they interact, the complex is in compact form. When these molecules translocate through the nanopore, the extended form gives shallower but longer signal, whereas the compact form generates deeper and short duration signal. By analyzing the statistics of these events, we can find out the relative proportion of the extended and compact forms at various voltages and interpolate it to the case of zero voltage which is the natural state. This would reveal the binding energy of a pair of nucleosomes. This approach could be implemented to differently modified nucleosomes to characterize various possible modulations in the inter-nucleosome interactions. Such experimental information would help us accurately model the large-scale chromatin systems in various conditions of interest.

Future possibility-4: Epigenetic sequencing using nanopores for molecular diagnostics

Very recently, a hybrid technique that utilizes fluorescence microscopy and Sanger sequencing techniques was developed by Efrat Shema, et.al. [159] to obtain the multiple modifications on each nucleosome along with its underlying DNA sequence. It would be more efficient and faster if the epigenetic mapping could be done from the intact chromatin fragments and by reading the location of the nucleosome from the relative position on the whole fragment. Nanopore is a promising technology for this application as it is fast and quantitative and works independent of immobilization and labelling techniques. This would be more suitable for epigenetic diagnostics applications. For simplicity, consider a nucleosome array of certain length. This generates spikes equal to the number of nucleosomes on the array with uniform depth. Now add an antibody to detect the epigenetic mark under investigation, then the antibody only binds to its respective modification on a particular nucleosome. This array when translocated through the nanopore would generate a deeper spike at the position where the antibody is bound. This could be repeated with various antibodies of interest on the same sample and this would tell us about the multiple modification marks on the chromatin fragments. It would be ideal to combine this information to the underlying DNA sequence.

Future possibility-5: Mapping gene activity levels using nanopore

Here, we propose a simple experimental scheme that could be useful for technological applications, with the knowledge that the activity of a gene is positively correlated to the formation of partial nucleosomes [53,55,57]. The translocation of chromatin fragments obtained from known regions of active (repressive) genes would result in electrical signals with levels of depths comparable to that of partial (full) nucleosomes presented in the case of 12mer array translocation events. The genes with differing transcriptional activity would generate different relative proportions of ΔG_{level} values corresponding to partial and full nucleosomes which could then be mapped on to the genome. The nanopore technique being label-free and requiring minimal sample processing steps are being explored for their applicability in the genome wide studies but they are still in their nascent stage [160,161]. Our work lays foundation for the development of portable sensors for the rapid survey of differential gene activity across the genome. For this to be possible, we suggest performing experiments at lower voltages that preserves the already existing nucleosomal structures on the chromatin fragments as the pulling forces would be reduced. An important question that still remains to be addressed is how the chromatin folding is affected due to the heterogeneity in the nucleosomal states (partial or full). Our study presents a good model for the chromatin in the cell that continuously undergoes conformational changes at the global level due to dynamic interactions between full and partial nucleosomal structures, during cellular processes. The information about the possibility and the propensity to form these sub-structures could help us better understand the local and global conformations of the chromatin. For a more realistic picture of the chromatin, extension of this study to the context of nucleosomes with epigenetic modifications is desired.

REFERENCES

[1] J. Vinograd, J. Lebowitz, R. Radloff, R. Watson, and P. Laipis, *The Twisted Circular Form of Polyoma Viral DNA.*, Proc. Natl. Acad. Sci. U. S. A. **53**, 1104 (1965).

[2] Andrew D. Bates and Anthony Maxwell, *DNA Topology* (Oxford University Press, 2005).

[3] F. B. Fuller, *The Writhing Number of a Space Curve.*, Proc. Natl. Acad. Sci. U. S. A. **68**, 815 (1971).

[4] A. Worcel and E. Burgi, *On the Structure of the Folded Chromosome of Escherichia Coli*, J. Mol. Biol. **71**, 127 (1972).

[5] P. R. Kramer and R. R. Sinden, *Measurement of Unrestrained Negative Supercoiling and Topological Domain Size in Living Human Cells*, Biochemistry **36**, 3151 (1997).

[6] P. Forterre, A. Bergerat, and P. Lopez-Garcia, *The Unique DNA Topology and DNA Topoisomerases of Hyperthermophilic Archaea*, FEMS Microbiol. Rev. **18**, 237 (1996).

[7] Y. L. Lyubchenko and L. S. Shlyakhtenko, *Visualization of Supercoiled DNA with Atomic Force Microscopy in Situ*, Proc. Natl. Acad. Sci. U. S. A. **94**, 496 (1997).

[8] P. Bettotti, V. Visone, L. Lunelli, G. Perugino, M. Ciaramella, and A. Valenti, *Structure and Properties of DNA Molecules over the Full Range of Biologically Relevant Supercoiling States*, Sci. Rep. **8**, 6163 (2018).

[9] T. Brouns, H. De Keersmaecker, S. F. Konrad, N. Kodera, T. Ando, J. Lipfert, S. De Feyter, and W. Vanderlinden, *Free Energy Landscape and Dynamics of Supercoiled DNA by High-Speed Atomic Force Microscopy*, ACS Nano **12**, 11907 (2018).

[10] D. A. Koster, A. Crut, S. Shuman, M. A. Bjornsti, and N. H. Dekker, *Cellular Strategies for Regulating DNA Supercoiling: A Single-Molecule Perspective*,

Cell.

[11] R. R. Sinden, *Supercoiled DNA: Biological Significance*, J. Chem. Educ. **64**, 294 (1987).

[12] E. A. Oussatcheva, J. Pavlicek, O. F. Sankey, R. R. Sinden, Y. L. Lyubchenko, and V. N. Potaman, *Influence of Global DNA Topology on Cruciform Formation in Supercoiled DNA*, J. Mol. Biol. **338**, 735 (2004).

[13] Y. Liu, V. Bondarenko, A. Ninfa, and V. M. Studitsky, *DNA Supercoiling Allows Enhancer Action over a Large Distance*, Proc. Natl. Acad. Sci. U. S. A. **98**, 14883 (2001).

[14] R. Kanaar, P. Van De Putte, and N. R. Cozzarelli, *Gin-Mediated DNA Inversion: Product Structure and the Mechanism of Strand Exchange*, Proc. Natl. Acad. Sci. U. S. A. **85**, 752 (1988).

[15] J. F. Marko, *The Internal "slithering" Dynamics of Supercoiled DNA*, **244**, 263 (1997).

[16] T. R. Strick, J.-F. Allemand, D. Bensimon, A. Bensimon, and V. Croquette, *The Elasticity of a Single Supercoiled DNA Molecule*, Science (80-.). **271**, 1835 (1996).

[17] A. Crut, D. A. Koster, R. Seidel, C. H. Wiggins, and N. H. Dekker, *Fast Dynamics of Supercoiled DNA Revealed by Single-Molecule Experiments*, PNAS **104**, 11957 (2007).

[18] M. T. J. Van Loenhout, M. V. De Grunt, and C. Dekker, *Dynamics of DNA Supercoils*, Science (80-.). **338**, 94 (2012).

[19] R. Vlijm, A. Mashaghi, S. Bernard, M. Modesti, and C. Dekker, *Experimental Phase Diagram of Negatively Supercoiled DNA Measured by Magnetic Tweezers and Fluorescence*, Nanoscale 3205 (2015).

[20] S. Hyun Kim, M. Ganji, E. Kim, J. van der Torre, E. Abbondanzieri, and C. Dekker, *DNA Sequence Encodes the Position of DNA Supercoils*, Elife **7**, e36557 (2018).

[21] J. G. Gall, *Chromosome Fibers Studied by a Spreading Technique*, Chromosoma

20, 221 (1966).

[22] J. T. Finch, M. Noll, and R. D. Kornberg, *Electron Microscopy of Defined Lengths of Chromatin*, Proc. Natl. Acad. Sci. U. S. A. **72**, 3320 (1975).

[23] C. L. F. Woodcock, J. P. Safer, and J. E. Stanchfield, *Structural Repeating Units in Chromatin. I. Evidence for Their General Occurrence*, Exp. Cell Res. **97**, 101 (1976).

[24] K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, and T. J. Richmond, *Crystal Structure of the Nucleosome Core Particle at 2.8 Å Resolution*, Nature.

[25] R. Everaers and H. Schiessel, *The Physics of Chromatin*, J. Phys. Condens. Matter **27**, (2015).

[26] K. J. Polach and J. Widom, *Mechanism of Protein Access to Specific DNA Sequences in Chromatin: A Dynamic Equilibrium Model for Gene Regulation*, J. Mol. Biol. **254**, 130 (1995).

[27] K. J. Polach and J. Widom, *A Model for the Cooperative Binding of Eukaryotic Regulatory Proteins to Nucleosomal Target Sites*, J. Mol. Biol. **258**, 800 (1996).

[28] B. Alberts, A. Johnson, J. Lewis, D. Morgan, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell*, 6th ed (Garland Science, n.d.).

[29] A. Pyne, R. Thompson, C. Leung, D. Roy, and B. W. Hoogenboom, *Single-Molecule Reconstruction of Oligonucleotide Secondary Structure by Atomic Force Microscopy*, Small **10**, 3257 (2014).

[30] M. A. Ricci, C. Manzo, M. F. García-Parajo, M. Lakadamyali, and M. P. Cosma, *Chromatin Fibers Are Formed by Heterogeneous Groups of Nucleosomes in Vivo*, Cell **160**, 1145 (2015).

[31] A. N. Boettiger, B. Bintu, J. R. Moffitt, S. Wang, B. J. Beliveau, G. Fudenberg, M. Imakaev, L. A. Mirny, C. T. Wu, and X. Zhuang, *Super-Resolution Imaging Reveals Distinct Chromatin Folding for Different Epigenetic States*, Nature **529**, 418 (2016).

[32] A. Allahverdi, Q. Chen, N. Korolev, and L. Nordenskiöld, *Chromatin Compaction under Mixed Salt Conditions: Opposite Effects of Sodium and*

Potassium Ions on Nucleosome Array Folding, Sci. Rep. **5**, (2015).

[33] W. Iwasaki, Y. Miya, N. Horikoshi, A. Osakabe, H. Taguchi, H. Tachiwana, T. Shibata, W. Kagawa, and H. Kurumizaka, *Contribution of Histone N-Terminal Tails to the Structure and Stability of Nucleosomes*, FEBS Open Bio **3**, 363 (2013).

[34] R. Jaenisch and A. Bird, *Epigenetic Regulation of Gene Expression: How the Genome Integrates Intrinsic and Environmental Signals*, Nat. Genet. **33**, 245 (2003).

[35] A. Bertin, A. Leforestier, D. Durand, and F. Livolant, *Role of Histone Tails in the Conformation and Interactions of Nucleosome Core Particles*, Biochemistry **43**, 4773 (2004).

[36] Y. Liu, C. Lu, Y. Yang, Y. Fan, R. Yang, C. F. Liu, N. Korolev, and L. Nordenskiöld, *Influence of Histone Tails and H4 Tail Acetylations on Nucleosome- Nucleosome Interactions*, J. Mol. Biol. **414**, 749 (2011).

[37] S. Mangenot, A. Leforestier, P. Vachette, D. Durand, and F. Livolant, *Salt-Induced Conformation and Interaction Changes of Nucleosome Core Particles*, Biophys. J. **82**, 345 (2002).

[38] N. Korolev, A. Allahverdi, Y. Yang, Y. Fan, A. P. Lyubartsev, and L. Nordenskiöld, *Electrostatic Origin of Salt-Induced Nucleosome Array Compaction*, Biophys. J. **99**, 1896 (2010).

[39] S. A. Grigoryev, G. Arya, S. Correll, C. L. Woodcock, and T. Schlick, *Evidence for Heteromorphic Chromatin Fibers from Analysis of Nucleosome Interactions*, Proc. Natl. Acad. Sci. **106**, 13317 (2009).

[40] K. Maeshima et al., *Nucleosomal Arrays Self-assemble into Supramolecular Globular Structures Lacking 30-nm Fibers*, EMBO J. **35**, 1115 (2016).

[41] M. L. Bennink, B. G. De Grooth, J. Greve, S. H. Leuba, G. H. Leno, and J. Zlatanova, *Unfolding Individual Nucleosomes by Stretching Single Chromatin Fibers with Optical Tweezers*, Nat. Struct. Biol. **8**, 606 (2001).

[42] Y. Cui and C. Bustamante, *Pulling a Single Chromatin Fiber Reveals the Forces That Maintain Its Higher-Order Structure*, Proc. Natl. Acad. Sci. **97**, 127 (2002).

[43] L. H. Pope, M. L. Bennink, K. A. Van Leijenhorst-Groener, D. Nikova, J. Greve, and J. F. Marko, *Single Chromatin Fiber Stretching Reveals Physically Distinct Populations of Disassembly Events*, Biophys. J. **88**, 3572 (2005).

[44] M. Kruithof, F. T. Chien, A. Routh, C. Logie, D. Rhodes, and J. Van Noort, *Single-Molecule Force Spectroscopy Reveals a Highly Compliant Helical Folding for the 30-Nm Chromatin Fiber*, Nat. Struct. Mol. Biol. **16**, 534 (2009).

[45] H. Meng, K. Andresen, and J. Van Noort, *Quantitative Analysis of Single-Molecule Force Spectroscopy on Folded Chromatin Fibers*, Nucleic Acids Res. **43**, 3578 (2015).

[46] A. Kaczmarczyk, A. Allahverdi, T. B. Brouwer, L. Nordenskiöld, N. H. Dekker, and J. Van Noort, *Single-Molecule Force Spectroscopy on Histone H4 Tail-Cross-Linked Chromatin Reveals Fiber Folding*, J. Biol. Chem. **292**, 17506 (2017).

[47] G. V. Soni, L. Brar, F. M. Hameed, A. K. Raychaudhuri, and G. V. Shivashankar, *Distinct Levels in the Nanoscale Organization of DNA-Histone Complex Revealed by Its Mechanical Unfolding*, Appl. Phys. Lett. **90**, 1 (2007).

[48] F. M. Hameed, G. V. Soni, H. Krishnamurthy, and G. V. Shivashankar, *Probing Structural Stability of Chromatin Assembly Sorted from Living Cells*, Biochem. Biophys. Res. Commun. **385**, 518 (2009).

[49] J. J. Funke, P. Ketterer, C. Lieleg, S. Schunter, P. Korber, and H. Dietz, *Uncovering the Forces between Nucleosomes Using DNA Origami*, Sci. Adv. **2**, (2016).

[50] S. J. McBryant, V. H. Adams, and J. C. Hansen, *Chromatin Architectural Proteins*, Chromosom. Res. **14**, 39 (2006).

[51] B. Hendrich and W. Bickmore, *Human Diseases with Underlying Defects in Chromatin Structure and Modification*, Hum. Mol. Genet. **10**, 2233 (2001).

[52] T. Misteli, *Higher-Order Genome Organization in Human Disease.*, Cold Spring Harb. Perspect. Biol. **2**, (2010).

[53] M. L. Kireeva, W. Walter, V. Tchernajenko, V. Bondarenko, M. Kashlev, and V. M. Studitsky, *Nucleosome Remodeling Induced by RNA Polymerase II: Loss of*

the H2A/H2B Dimer during Transcription, Mol. Cell **9**, 541 (2002).

[54] J. Zlatanova, T. C. Bishop, J. M. Victor, V. Jackson, and K. van Holde, *The Nucleosome Family: Dynamic and Growing*, Structure **17**, 160 (2009).

[55] H. A. Cole, J. Ocampo, J. R. Iben, R. V. Chereji, and D. J. Clark, *Heavy Transcription of Yeast Genes Correlates with Differential Loss of Histone H2B Relative to H4 and Queued RNA Polymerases*, Nucleic Acids Res. **42**, 12512 (2014).

[56] S. Venkatesh and J. L. Workman, *Histone Exchange, Chromatin Structure and the Regulation of Transcription*, Nat. Rev. Mol. Cell Biol. **16**, 178 (2015).

[57] H. S. Rhee, A. R. Bataille, L. Zhang, and B. F. Pugh, *Subnucleosomal Structures and Nucleosome Asymmetry across a Genome*, Cell **159**, 1377 (2014).

[58] I. Nazarov, I. Chekliarova, G. Rychkov, A. V. Ilatovskiy, C. Crane-Robinson, and A. Tomilin, *AFM Studies in Diverse Ionic Environments of Nucleosomes Reconstituted on the 601 Positioning Sequence*, Biochimie **121**, 5 (2016).

[59] G. N. Rychkov, A. V. Ilatovskiy, I. B. Nazarov, A. V. Shvetsov, D. V. Lebedev, A. Y. Konev, V. V. Isaev-Ivanov, and A. V. Onufriev, *Partially Assembled Nucleosome Structures at Atomic Detail*, Biophys. J. **112**, 460 (2017).

[60] A. Gansen, A. Valeri, F. Hauger, S. Felekyan, S. Kalinin, K. Tóth, J. Langowski, and C. A. M. Seidel, *Nucleosome Disassembly Intermediates Characterized by Single-Molecule FRET*, Proc. Natl. Acad. Sci. U. S. A. **106**, 15308 (2009).

[61] J. Mazurkiewicz, J. F. Kepert, and K. Rippe, *On the Mechanism of Nucleosome Assembly by Histone Chaperone NAP1*, J. Biol. Chem. **281**, 16462 (2006).

[62] W. H. Coulter, *Means for Counting Particles Suspended in a Fluid. US Patent, 2656508*, (1953).

[63] A. J. Storm, C. Storm, J. Chen, H. Zandbergen, J. F. Joanny, and C. Dekker, *Fast DNA Translocation through a Solid-State Nanopore*, Nano Lett. **5**, 1193 (2005).

[64] J. Li, M. Gershow, D. Stein, E. Brandin, and J. A. Golovchenko, *DNA Molecules and Configurations in a Solid-State Nanopore Microscope*, Nat. Mater. **2**, 611 (2003).

[65] L. Liu, C. Yang, K. Zhao, J. Li, and H. C. Wu, *Ultrashort Single-Walled Carbon Nanotubes in a Lipid Bilayer as a New Nanopore Sensor*, Nat. Commun. **4**, 2989 (2013).

[66] G. F. Schneider, S. W. Kowalczyk, V. E. Calado, G. Pandraud, H. W. Zandbergen, L. M. K. Vandersypen, and C. Dekker, *DNA Translocation through Graphene Nanopores*, Nano Lett. **10**, 3163 (2010).

[67] L. J. Steinbock, R. D. Bulushev, S. Krishnan, C. Raillon, and A. Radenovic, *DNA Translocation through Low-Noise Glass Nanopores*, ACS Nano **7**, 11255 (2013).

[68] L. J. Steinbock, O. Otto, C. Chimerel, J. Gornall, and U. F. Keyser, *Detecting DNA Folding with Nanocapillaries*, Nano Lett. **10**, 2493 (2010).

[69] J. A. Bafna and G. V. Soni, *Fabrication of Low Noise Borosilicate Glass Nanopores for Single Molecule Sensing*, PLoS One **11**, e0157399 (2016).

[70] C. Dekker, *Solid-State Nanopores*, Nat. Nanotechnol. **2**, 209 (2007).

[71] D. Branton et al., *The Potential and Challenges of Nanopore Sequencing*, Nature Biotechnology.

[72] M. Wanunu, *Nanopores: A Journey towards DNA Sequencing*, Physics of Life Reviews.

[73] S. M. Iqbal and R. Bashir, *Nanopores: Sensing and Fundamental Biological Interactions* (Springer, 2011).

[74] R. D. Bulushev, L. J. Steinbock, S. Khlybov, J. F. Steinbock, U. F. Keyser, and A. Radenovic, *Measurement of the Position-Dependent Electrophoretic Force on DNA in a Glass Nanocapillary*, Nano Lett. **14**, 6606 (2014).

[75] N. E. Weckman, N. Ermann, R. Gutierrez, K. Chen, J. Graham, R. Tivony, A. Heron, and U. F. Keyser, *Multiplexed DNA Identification Using Site Specific DCas9 Barcodes and Nanopore Sensing*, ACS Sensors **4**, 2065 (2019).

[76] S. M. Lu, Y. Y. Peng, Y. L. Ying, and Y. T. Long, *Electrochemical Sensing at a Confined Space*, Anal. Chem. **92**, 5621 (2020).

[77] R. J. Yu, Y. L. Ying, R. Gao, and Y. T. Long, *Confined Nanopipette Sensing:*

From Single Molecules, Single Nanoparticles, to Single Cells, Angew. Chemie - Int. Ed. **58**, 3706 (2019).

[78] P. Waduge, R. Hu, P. Bandarkar, H. Yamazaki, B. Cressiot, Q. Zhao, P. C. Whitford, and M. Wanunu, *Nanopore-Based Measurements of Protein Size, Fluctuations, and Conformational Changes*, ACS Nano **11**, 5706 (2017).

[79] H. Chae, D. K. Kwak, M. K. Lee, S. W. Chi, and K. B. Kim, *Solid-State Nanopore Analysis on Conformation Change of P53TAD-MDM2 Fusion Protein Induced by Protein-Protein Interaction*, Nanoscale **10**, 17227 (2018).

[80] X. Wang, M. D. Wilkinson, X. Lin, R. Ren, K. R. Willison, A. P. Ivanov, J. Baum, and J. B. Edel, *Single-Molecule Nanopore Sensing of Actin Dynamics and Drug Binding*, Chem. Sci. **11**, 970 (2020).

[81] R. Nouri, Y. Jiang, X. L. Lian, and W. Guan, *Sequence-Specific Recognition of HIV-1 DNA with Solid-State CRISPR-Cas12a-Assisted Nanopores (SCAN)*, ACS Sensors **5**, 1273 (2020).

[82] A. Fennouri, C. Przybylski, M. Pastoriza-Gallego, L. Bacri, L. Auvray, and R. Daniel, *Single Molecule Detection of Glycosaminoglycan Hyaluronic Acid Oligosaccharides and Depolymerization Enzyme Activity Using a Protein Nanopore*, ACS Nano **6**, 9672 (2012).

[83] T. Ma, E. Balanzat, J. M. Janot, and S. Balme, *Single Conical Track-Etched Nanopore for a Free-Label Detection of OSCS Contaminants in Heparin*, Biosens. Bioelectron. **137**, 207 (2019).

[84] N. Giamblanco, J. M. Janot, A. Gubbiotti, M. Chinappi, and S. Balme, *Characterization of Food Amyloid Protein Digestion by Conical Nanopore*, Small Methods **4**, 1900703 (2020).

[85] T. Ma, J. M. Janot, and S. Balme, *Dynamics of Long Hyaluronic Acid Chains through Conical Nanochannels for Characterizing Enzyme Reactions in Confined Spaces*, Nanoscale **12**, 7231 (2020).

[86] C. Plesa, D. Verschueren, S. Pud, J. Van Der Torre, J. W. Ruitenberg, M. J. Witteveen, M. P. Jonsson, A. Y. Grosberg, Y. Rabin, and C. Dekker, *Direct Observation of DNA Knots Using a Solid-State Nanopore*, Nat. Nanotechnol. **11**,

1093 (2016).

[87] R. Kumar Sharma, I. Agrawal, L. Dai, P. S. Doyle, and S. Garaj, *Complex DNA Knots Detected with a Nanopore Sensor*, Nat. Commun. **10**, 4473 (2019).

[88] F. Bošković, J. Zhu, K. Chen, and U. F. Keyser, *Monitoring G-Quadruplex Formation with DNA Carriers and Solid-State Nanopores*, Nano Lett. **19**, 7996 (2019).

[89] D. Fologea, E. Brandin, J. Uplinger, D. Branton, and J. Li, *DNA Conformation and Base Number Simultaneously Determined in a Nanopore*, Electrophoresis **28**, 3186 (2007).

[90] Q. Chen, S. Diao, and C. Wu, *How Does a Supercoiled DNA Chain Pass through a Small Conical Glass Pore?*, Soft Matter **8**, 5451 (2012).

[91] E. Krueger et al., *Modeling and Analysis of Intercalant Effects on Circular DNA Conformation*, ACS Nano **10**, 8910 (2016).

[92] W. Yang, L. Restrepo-Pérez, M. Bengtson, S. J. Heerema, A. Birnie, J. Van Der Torre, and C. Dekker, *Detection of CRISPR-DCas9 on DNA with Solid-State Nanopores*, Nano Lett. **18**, 6469 (2018).

[93] R. D. Bulushev, S. Marion, E. Petrova, S. J. Davis, S. J. Maerkl, and A. Radenovic, *Single Molecule Localization and Discrimination of DNA-Protein Complexes by Controlled Translocation Through Nanocapillaries*, Nano Lett. **16**, 7882 (2016).

[94] H. Kaur, S. Nandivada, M. C. Acharjee, D. S. McNabb, and J. Li, *Estimating RNA Polymerase Protein Binding Sites on λ DNA Using Solid-State Nanopores*, ACS Sensors **4**, 100 (2019).

[95] M. Raveendran, A. R. Leach, T. Hopes, J. L. Aspden, and P. Actis, *Ribosome Fingerprinting with a Solid-State Nanopore*, ACS Sensors **5**, 3533 (2020).

[96] A. Squires, E. Atas, and A. Meller, *Nanopore Sensing of Individual Transcription Factors Bound to DNA*, Sci. Rep. **5**, 1 (2015).

[97] V. Tabard-Cossa, M. Wiggin, D. Trivedi, N. N. Jetha, J. R. Dwyer, and A. Marziali, *Single-Molecule Bonds Characterized by Solid-State Nanopore Force*

Spectroscopy, ACS Nano **3**, 3009 (2009).

[98] B. Dorvel, G. Sigalov, Q. Zhao, J. Comer, V. Dimitrov, U. Mirsaidov, A. Aksimentiev, and G. Timp, *Analyzing the Forces Binding a Restriction Endonuclease to DNA Using a Synthetic Nanopore*, Nucleic Acids Res. **37**, 4170 (2009).

[99] A. Ivankin, S. Carson, S. R. M. Kinney, and M. Wanunu, *Fast, Label-Free Force Spectroscopy of Histone–DNA Interactions in Individual Nucleosomes Using Nanopores*, J. Am. Chem. Soc. **135**, 15350 (2013).

[100] M. Langecker, A. Ivankin, S. Carson, S. R. M. Kinney, F. C. Simmel, and M. Wanunu, *Nanopores Suggest a Negligible Influence of CpG Methylation on Nucleosome Packaging and Stability*, Nano Lett. **15**, 783 (2015).

[101] G. V Soni and C. Dekker, *Detection of Nucleosomal Substructures Using Solid-State Nanopores*, Nano Lett. **12**, 3180 (2012).

[102] N. B. Grover, J. Naaman, S. Ben-Sasson, and F. Doljanski, *Electrical Sizing of Particles in Suspensions: I. Theory*, Biophys. J. **9**, 1398 (1969).

[103] N. B. Grover, J. Naaman, S. Ben-Sasson, F. Doljanski, and E. Nadav, *Electrical Sizing of Particles in Suspensions: II. Experiments with Rigid Spheres*, Biophys. J. **9**, 1415 (1969).

[104] J. C. Maxwell, *A Treatise on Electricity & Magnetism (Volume 1)*, Third edit (Dover Publications, n.d.).

[105] N. B. Grover, J. Naaman, S. Ben-Sasson, and F. Doljanski, *Electrical Sizing of Particles in Suspensions: III. Rigid Spheroids and Red Blood Cells*, Biophys. J. **12**, 1099 (1972).

[106] R. W. DeBlois and C. P. Bean, *Counting and Sizing of Submicron Particles by the Resistive Pulse Technique*, Rev. Sci. Instrum. **41**, 909 (1970).

[107] A. J. Storm, J. H. Chen, X. S. Ling, H. W. Zandbergen, and C. Dekker, *Fabrication of Solid-State Nanopores with Single-Nanometre Precision*, Nature Materials.

[108] H. Kwok, K. Briggs, and V. Tabard-Cossa, *Nanopore Fabrication by Controlled*

Dielectric Breakdown, PLoS One **9**, (2014).

[109] J. E. Hall, *Access Resistance of a Small Circular Pore*, J. Gen. Physiol. **66**, 531 (1975).

[110] S. W. Kowalczyk, A. Y. Grosberg, Y. Rabin, and C. Dekker, *Modeling the Conductance and DNA Blockade of Solid-State Nanopores*, Nanotechnology **22**, 315101 (2011).

[111] D. S. Talaga and J. Li, *Single-Molecule Protein Unfolding in Solid State Nanopores*, J. Am. Chem. Soc. **131**, 9287 (2009).

[112] E. C. Yusko, J. M. Johnson, S. Majd, P. Prangkio, R. C. Rollings, J. Li, J. Yang, and M. Mayer, *Controlling Protein Translocation through Nanopores with Bio-Inspired Fluid Walls*, Nat. Nanotechnol. **6**, 253 (2011).

[113] E. C. Yusko, P. Prangkio, D. Sept, R. C. Rollings, J. Li, and M. Mayer, *Single-Particle Characterization of Aβ Oligomers in Solution*, ACS Nano **6,** 5909 (2012).

[114] E. C. Yusko et al., *Real-Time Shape Approximation and Fingerprinting of Single Proteins Using a Nanopore*, Nat. Nanotechnol. **12**, 360 (2017).

[115] L. J. Steinbock, J. F. Steinbock, and A. Radenovic, *Controllable Shrinking and Shaping of Glass Nanocapillaries under Electron Irradiation*, Nano Lett. **13**, 1717 (2013).

[116] W. Li, N. A. W. Bell, S. Hernández-Ainsa, V. V. Thacker, A. M. Thackray, R. Bujdoso, and U. F. Keyser, *Single Protein Molecule Detection by Glass Nanopores*, ACS Nano **7**, 4129 (2013).

[117] A. J. Storm, J. H. Chen, H. W. Zandbergen, and C. Dekker, *Translocation of Double-Strand DNA through a Silicon Oxide Nanopore*, Phys. Rev. E **71**, 051903 (2005).

[118] D. Fologea, M. Gershow, B. Ledden, D. S. McNabb, J. A. Golovchenko, and J. Li, *Detecting Single Stranded DNA with a Solid State Nanopore*, Nano Lett. **5**, 1905 (2005).

[119] P. R. Cook, *The Organization of Replication and Transcription*, Science.

[120] C. Lavelle, *Pack, Unpack, Bend, Twist, Pull, Push: The Physical Side of Gene Expression*, Curr. Opin. Genet. Dev. **25**, 74 (2014).

[121] L. Baranello, D. Levens, A. Gupta, and F. Kouzine, *The Importance of Being Supercoiled: How DNA Mechanics Regulate Dynamic Processes*, Biochim. Biophys. Acta - Gene Regul. Mech. **1819**, 632 (2012).

[122] R. Cortini, M. Barbi, B. R. Caré, C. Lavelle, A. Lesne, J. Mozziconacci, and J. M. Victor, *The Physics of Epigenetics*, Rev. Mod. Phys. **88**, 025002 (2016).

[123] J. G. Brahms, S. Brahms, Y. Ohara, and V. Vagner, *Activation and Inhibition of Transcription by Supercoiling*, J. Mol. Biol **181**, 455 (1985).

[124] G. Snounou and A. D. B. Malcolm, *Supercoiling and the Mechanism of Restriction Endonucleases*, Eur. J. Biochem. **138**, 275 (1984).

[125] S. E. Halford, N. P. Johnson, and J. Grinsted, *The Reactions of the EcoRI and Other Restriction Endonucleases*, Biochem. J. **179**, 353 (1979).

[126] T. R. Strick, V. Croquette, and D. Bensimon, *Single-Molecule Analysis of DNA Uncoiling by a Type II Topoisomerase*, Nature **404**, 901 (2000).

[127] D. A. Koster, V. Croquette, C. Dekker, S. Shuman, and N. H. Dekker, *Friction and Torque Govern the Relaxation of DNA Supercoils by Eukaryotic Topoisomerase IB*, Nature **434**, 671 (2005).

[128] J. Gore, Z. Bryant, M. D. Stone, M. Nöllmann, N. R. Cozzarelli, and C. Bustamante, *Mechanochemical Analysis of DNA Gyrase Using Rotor Bead Tracking*, Nature **439**, 100 (2006).

[129] A. Gansen, K. Tóth, N. Schwarz, and J. Langowski, *Structural Variability of Nucleosomes Detected by Single-Pair Förster Resonance Energy Transfer: Histone Acetylation, Sequence Variation, and Salt Effects*, J. Phys. Chem. B **113**, 2604 (2009).

[130] N. P. Hazan, T. E. Tomov, R. Tsukanov, M. Liber, Y. Berger, R. Masoud, K. Toth, J. Langowski, and E. Nir, *Nucleosome Core Particle Disassembly and Assembly Kinetics Studied Using Single-Molecule Fluorescence*, Biophys. J. **109**, 1676 (2015).

[131] S. S. Zakharova, W. Jesse, C. Backendorf, S. U. Egelhaaf, A. Lapp, and J. R. C. Van Der Maarel, *Dimensions of Plectonemically Supercoiled DNA*, Biophys. J. **83**, 1106 (2002).

[132] E. C. Yusko, J. M. Johnson, S. Majd, P. Prangkio, R. C. Rollings, J. Li, J. Yang, and M. Mayer, *Controlling Protein Translocation through Nanopores with Bio-Inspired Fluid Walls*, Nat. Nanotechnol. **6**, 253 (2011).

[133] L. J. Steinbock, S. Krishnan, R. D. Bulushev, S. Borgeaud, M. Blokesch, L. Feletti, and A. Radenovic, *Probing the Size of Proteins with Glass Nanopores*, Nanoscale **6**, 14380 (2014).

[134] B. M. Reinhard, J. M. Yassif, P. Vach, and J. Liphardt, *Plasmon Rulers as Dynamic Molecular Rulers in Enzymology*, in *Methods in Enzymology*, Vol. 475 (Academic Press Inc., 2010), pp. 175–198.

[135] S. E. Halford, *Hopping, Jumping and Looping by Restriction Enzymes*, in *Biochemical Society Transactions*, Vol. 29 (2001), pp. 363–373.

[136] Y. Nechemia-Arbely et al., *Human Centromeric CENP-A Chromatin Is a Homotypic, Octameric Nucleosome at All Cell Cycle Points*, J. Cell Biol. **216**, 607 (2017).

[137] R. Hu, C. Liu, W. Lu, G. Wei, D. Yu, W. Li, P. Chen, G. Li, and Q. Zhao, *Probing the Effect of Ubiquitinated Histone on Mononucleosomes by Translocation Dynamics Study through Solid-State Nanopores*, Nano Lett. **22**, 888 (2022).

[138] S. K. Maheshwaram, K. Sreenivasa, and G. V. Soni, *Fingerprinting Branches on Supercoiled Plasmid DNA Using Quartz Nanocapillaries*, Nanoscale **13**, 320 (2021).

[139] U. F. Keyser, B. N. Koeleman, S. Van Dorp, D. Krapf, R. M. M. Smeets, S. G. Lemay, N. H. Dekker, and C. Dekker, *Direct Force Measurements on DNA in a Solid-State Nanopore*, Nat. Phys. **2**, 473 (2006).

[140] R. M. M. Smeets, U. F. Keyser, D. Krapf, M.-Y. Y. Wu, N. H. Dekker, and C. Dekker, *Salt Dependence of Ion Transport and DNA Translocation through Solid-State Nanopores*, Nano Lett. **6**, 89 (2006).

[141] N. Laohakunakorn, Electrokinetic Phenomena in Nanopore Transport, Ph. D. Thesis, University of Cambridge, 2015.

[142] B. D. Brower-Toland, C. L. Smith, R. C. Yeh, J. T. Lis, C. L. Peterson, and M. D. Wang, *Mechanical Disruption of Individual Nucleosomes Reveals a Reversible Multistage Release of DNA*, Proc. Natl. Acad. Sci. U. S. A. **99**, 1960 (2002).

[143] R. Vlijm, J. S. J. Smitshuijzen, A. Lusser, and C. Dekker, *NAP1-Assisted Nucleosome Assembly on DNA Measured in Real Time by Single-Molecule Magnetic Tweezers*, PLoS One **7**, (2012).

[144] O. Ordu, A. Lusser, and N. H. Dekker, *DNA Sequence Is a Major Determinant of Tetrasome Dynamics*, Biophys. J. **117**, 2217 (2019).

[145] Y. Arimura, H. Tachiwana, T. Oda, M. Sato, and H. Kurumizaka, *Structural Analysis of the Hexasome, Lacking One Histone H2A/H2B Dimer from the Conventional Nucleosome*, Biochemistry **51**, 3302 (2012).

[146] T. J. Richmond, J. T. Finch, B. Rushton, D. Rhodes, and A. Klug, *Structure of the Nucleosome Core Particle at 7 Å Resolution*, Nature **311**, 532 (1984).

[147] K. Yoda, S. Ando, S. Morishita, K. Houmura, K. Hashimoto, K. Takeyasu, and T. Okazaki, *Human Centromere Protein A (CENP-A) Can Replace Histone H3 in Nucleosome Reconstitution in Vitro*, Proc. Natl. Acad. Sci. U. S. A. **97**, 7266 (2000).

[148] S. Pisano, D. Leoni, A. Galati, D. Rhodes, M. Savino, and S. Cacchione, *The Human Telomeric Protein HTRF1 Induces Telomere-Specific Nucleosome Mobility*, Nucleic Acids Res. **38**, 2247 (2010).

[149] K. M. Krzemien, M. Beckers, S. Quack, and J. Michaelis, *Atomic Force Microscopy of Chromatin Arrays Reveal Non-Monotonic Salt Dependence of Array Compaction in Solution*, PLoS One **12**, (2017).

[150] G. C. Runger, D. M. Hawkins, and D. H. Olwell, *Cumulative Sum Charts and Charting for Quality Improvement* (Springer Verlag, 1998).

[151] C. Raillon, P. Granjon, M. Graf, L. J. Steinbock, and A. Radenovic, *Fast and Automatic Processing of Multi-Level Events in Nanopore Translocation*

Experiments, Nanoscale **4**, 4916 (2012).

[152] N. Ermann, N. Hanikel, V. Wang, K. Chen, N. E. Weckman, and U. F. Keyser, *Promoting Single-File DNA Translocations through Nanopores Using Electro-Osmotic Flow*, J. Chem. Phys. **149**, (2018).

[153] M. Nekrasov, B. Wild, and J. Müller, *Nucleosome Binding and Histone Methyltransferase Activity of Drosophila PRC2*, EMBO Rep. **6**, 348 (2005).

[154] K. L. Diehl, E. J. Ge, D. N. Weinberg, K. S. Jani, C. D. Allis, and T. W. Muir, *PRC2 Engages a Bivalent H3K27M-H3K27me3 Dinucleosome Inhibitor*, Proc. Natl. Acad. Sci. U. S. A. **116**, 22152 (2019).

[155] X. Wang, R. D. Paucek, A. R. Gooding, Z. Z. Brown, E. J. Ge, T. W. Muir, and T. R. Cech, *Molecular Analysis of PRC2 Recruitment to DNA in Chromatin and Its Inhibition by RNA*, Nat. Struct. Mol. Biol. **24**, 1028 (2017).

[156] P. R. Heenan, X. Wang, A. R. Gooding, T. R. Cech, and T. T. Perkins, *Bending and Looping of Long DNA by Polycomb Repressive Complex 2 Revealed by AFM Imaging in Liquid*, Nucleic Acids Res. **48**, 2969 (2020).

[157] X. Lin, R. Leicher, S. Liu, and B. Zhang, *Cooperative DNA Looping by PRC2 Complexes*, (2021).

[158] R. Leicher, E. J. Ge, X. Lin, M. J. Reynolds, W. Xie, T. Walz, B. Zhang, T. W. Muir, and S. Liu, *Single-Molecule and in Silico Dissection of the Interaction between Polycomb Repressive Complex 2 and Chromatin*, Proc. Natl. Acad. Sci. U. S. A. **117**, 30465 (2020).

[159] E. Shema, D. Jones, N. Shoresh, L. Donohue, O. Ram, and B. E. Bernstein, *Single-Molecule Decoding of Combinatorially Modified Nucleosomes*, Science, **352**, 717 (2016).

[160] K. Chen, F. Gularek, B. Liu, E. Weinhold, and U. F. Keyser, *Electrical DNA Sequence Mapping Using Oligodeoxynucleotide Labels and Nanopores*, ACS Nano **15**, 2679 (2021).

[161] A. Rand, P. Zimny, R. Nagel, C. Telang, J. Mollison, A. Bruns, E. Leff, W. Reisner, and W. B. Dunbar, *Electronic Mapping of a Bacterial Genome with Dual Solid-State Nanopores and Active Single-Molecule Control*, ACS Nano

(2022).

[162] S. D. Chandradoss, A. C. Haagsma, Y. K. Lee, J.-H. Hwang, J.-M. Nam, and C. Joo, *Surface Passivation for Single-Molecule Protein Studies*, J. Vis. Exp. 1 (2014).

[163] G. I. Bell, *Models for the Specific Adhesion of Cells to Cells*, Science (80-.). **200**, 618 (1978).

[164] E. Evans and K. Ritchie, *Dynamic Strength of Molecular Adhesion Bonds*, Biophys. J. **72**, 1541 (1997).

Appendix 1: Stiffness measurements using AFM

Introduction

AFM is an ideal instrument to perform force spectroscopy as it can accurately measure tiny deflection in the cantilever with nanometer accuracy. This feature is exploited to measure or apply the forces to measure the mechanical properties of the surfaces or the binding strength of the molecules. Here, we discuss the results obtained on measurements of elastic modulus of PDMS which is a soft material and is routinely used in biophysical measurements due to its biocompatibility. It is also used for experiments where the cells are grown on substrates with controlled stiffness as the mechanical properties of PDMS can be manipulated by optimizing the ratio of the polymer and the crosslinking agent. In this case, we have prepared samples with polymer and crosslinking agent in the ratios 5:1 and 10:1 with the latter being softer.

Experimental methods

Sylgard 184 (Dow) contains two liquid components which are PDMS polymer and crosslinking agent. In order to prepare two samples with different mechanical properties, we added polymer to crosslinking agent in the weight ratios of 5:1 and 10:1 in a glass petridish. The solution is then mixed thoroughly and placed in desiccator for 1 hr to remove air bubbles trapped in the viscous solution. Sometimes, there are large air bubbles, they were removed by venting and turning on the vacuum again, and this process is repeated till all the bubbles are removed. Once the solution is clear, the petridishes are placed in a hot air oven at 100 °C for 1 hr for curing. When the PDMS is cured, a small slice is cut using surgical knife and the side that was in contact with the glass surface and remained clean is now secured on AFM sample stage and facing the cantilever (Contact-G from Budget Sensors, nominal spring constant of 0.2 N/m and half-cone angle (θ) of 10°) measurements are recorded.

Experimental principle and data analysis

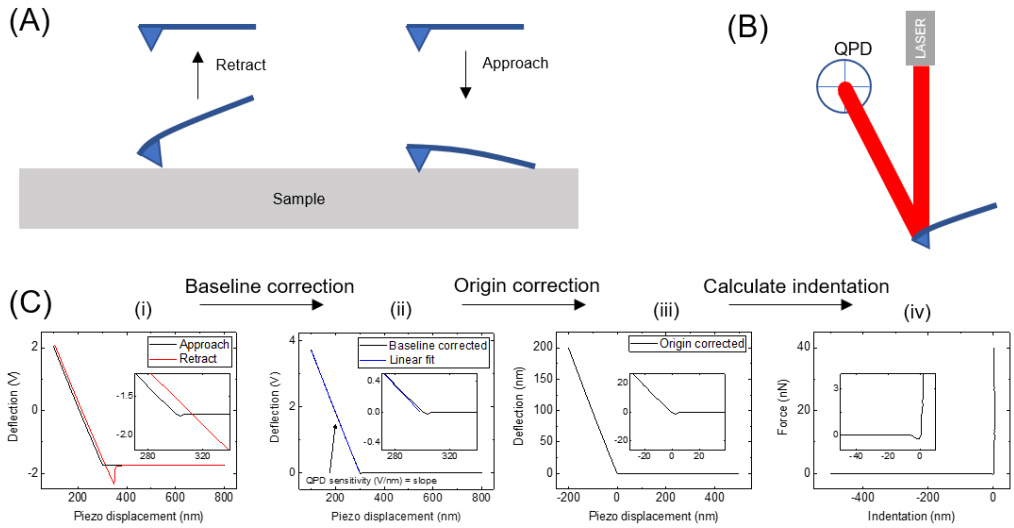


Figure A1.1: Force-distance curves on a substrate.

(A) Schematic showing the approach and retraction of AFM tip from the surface. (B) Schematic showing the reflection of laser onto the QPD to detect tiny changes in the bending of cantilever. (C) Step-by-step procedure to obtain F vs δ demonstrated on an F-x curve recorded on glass surface. (i) The curve generated by the AFM system with Piezo displacement on x-axis and Deflection voltage on y-axis. (ii) Approach curve is considered and baseline corrected to set the free cantilever deflection to zero. The QPD sensitivity is estimated from the slope of this curve upon contacting the surface. (iii) The y-axis is converted to deflection in nm and the curve is moved along x-axis to make the force zero when the piezo displacement is zero. (iv) The x-axis is converted to tip-sample separation (indentation) and the y-axis is converted to force applied. The insets of all the plots show zooms of the curve at the contact point.

In force spectroscopy experiments, the tip approaches the surface and continues to indent it by pressing against it, this increase in the applied force also results in bending of the cantilever as shown in Fig A1.1A. The cantilever bending is recorded as the deflection in the laser beam incident on the QPD measured from the difference in voltage between the top and bottom quadrants (Fig A1.1B). On the hard surfaces such as glass coverslip, the force-distance (F-x) curve is recorded and found that the curve is linear upon pressing the cantilever tip on the surface (Fig A1.1C). This is because the cantilever deflects by the amount equal to the piezo-displacement beyond the contact point as the hard surface cannot be indented. The slope of this curve gives the deflection sensitivity of the QPD (V/nm) as it quantifies the change in voltage (V) generated due to the deflection in cantilever by 1 nm. We have used a custom-written LabVIEW code to automate the analysis of F-x curves to estimate the mechanical properties of the substrate. To measure the indentation in the surface from the F-x curve, a series of steps need to be followed.

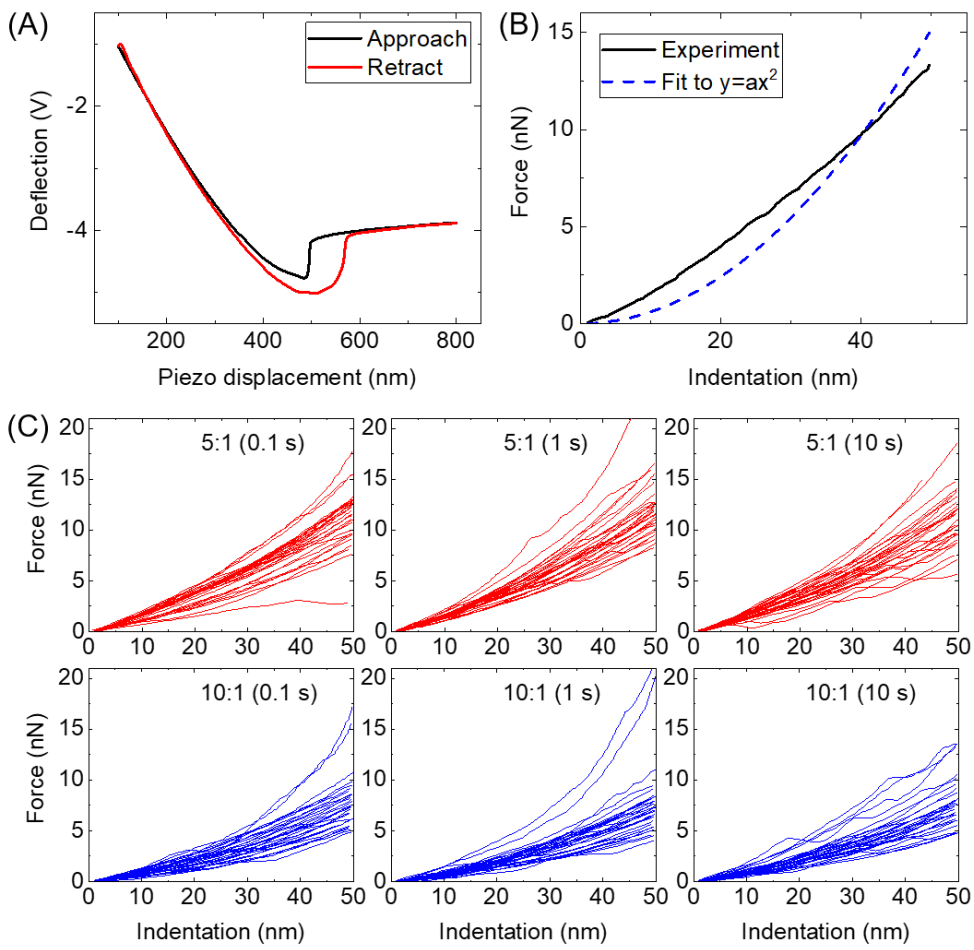


Figure A1.2: Force-distance curves on a substrate.

(A) A typical force-distance curve while approaching and retracting from the PDMS block. (B) The approach curve is processed as shown in Fig A1.1C and the force-indentation curve is obtained. It is fitted with the elasticity model for using a cantilever with a conical tip. (C) All the force-indentation curves obtained at different cycle durations of 0.1 s, 1 s and 10 s respectively. Red and Blue curves represent the 5:1 and 10:1 PDMS samples respectively.

In a given sweep cycle, F-x curves are recorded while the cantilever approaches and retracts from the surface but further analysis is performed only on the approach curves. The curve is processed in the following steps:

(ii) Baseline correction is done to make the deflection zero when the cantilever is far away from the surface. This is done by fitting the last 25 datapoints of the curve with a straight line and the fit is extrapolated to the full x-range of the curve and it is subtracted from the full curve. This results in the curve shown in step (ii) of Fig A1.1C. The slope of the linear region of the curve gave a QPD sensitivity of 0.019 V/nm.

(iii) The transition from the flat region to linearly increasing region happens at a piezo-displacement of about 300 nm which is supposed to be the contact point of cantilever

with the surface (origin). This is obtained by finding the location of the curve where the linear region starts reducing below 0 V while going from left to right. The x-axis is corrected to set the contact point at 0 nm.

(iv) Now, the F-x curve is converted to force-indentation curve by calculating the indentation as

$$\text{Indentation (nm)} = -(\text{Piezo displacement (nm)} + \text{Deflection (nm)})$$

Then the force vs indentation is plotted. In the case of glass surface, we can see that the cantilever experiences zero force till it comes into contact with the surface, and then it cannot indent into the surface even when it experiences high forces, as expected for a rigid surface.

Elastic modulus of PDMS

A small piece of PDMS is placed on the glass coverslip and held tight with the clips on the sample stage of AFM. The F-x curves are recorded three times at each of the cycle durations of 0.1 s, 1 s and 10 s. Cycle duration is defined as the total time taken for approach-retract cycle for the set total distance of piezo-movement. This is repeated at ten different locations on the surface. The tip is moved between the different spots on the sample surface by using the micrometer attached to the sample stage. A typical F-x curve obtained during approach and retract cycles on the PDMS block is shown in Fig A1.2A having non-linear region which is quite distinct from the curves obtained on the glass surface. The approach curves are processed similar to the case of glass surface and this resulted in the curve shown in Fig A1.2B. The data is fitted with the Sneddon model for a conical tip to extract the elastic model of PDMS,

$$F = \frac{2}{\pi}\frac{E}{1-\nu^2}\delta^2 \tan\theta$$

Here $\nu = 0.5$ is the Poisson's ratio of PDMS, θ is the half-cone angle of the tip, E is the elastic modulus, F is the force applied and δ is the indentation of the sample. All the F- δ curves obtained for both 5:1 and 10:1 PDMS samples under various conditions are shown in Fig A1.2C.

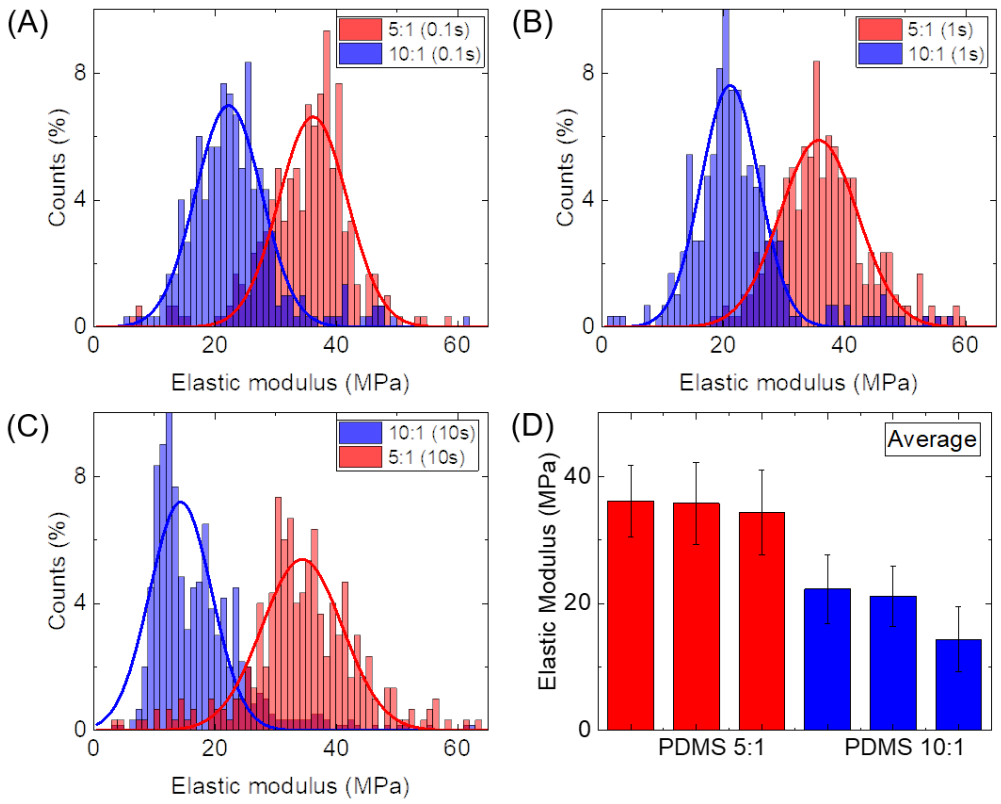


Figure A1.3: Statistical analysis of PDMS elastic modulus.

(A)-(C) compare the histograms of elastic modulus obtained for 5:1 and 10:1 PDMS samples at various sweep cycle durations of 0.1 s, 1 s and 10 s respectively. (D) Bar plot shows the mean and error bars of elastic modulus obtained for both the samples at all sweep cycle durations.

The elastic modulus value is extracted from each of these curves by fitting the model. The histograms of elastic modulus are plotted and comparison is made between the 5:1 and 10:1 samples in Fig A1.3A-C for curves recorded at 0.1 s, 1 s and 10 s respectively. The elastic modulus of 5:1 is higher compared to that of 10:1 for any cycle duration. Fig A1.3D shows the mean and error bars in the measurement of elastic modulus of PDMS. It is clear that the 5:1 PDMS is higher compared to that of 10:1 PDMS. This is expected as the increase in amount of curing agent increases the cross-linking between the PDMS polymers thereby resulting in the increased stiffness of the polymer.

Appendix 2: AFM force spectroscopy on streptavidin-biotin

Single-molecule force spectroscopy measurements help us understand the kinetics of bond rupture between molecules of interest, in this case we have chosen to study the binding between streptavidin (STV) and biotin. The interaction between STV (60 kDa protein) and biotin (244 Da ligand) is considered to form one of the strongest non-covalent bonds. Streptavidin has four binding sites, with one biotin allowed to bind at each site. From X-ray crystallography studies this interaction was found to be a combination of hydrogen bonding and Van der Waals forces. We employ atomic force microscope to perform the measurements on STV-biotin binding strength.

Experimental methods

Cantilever tips:

Firstly, perform piranha cleaning of the tips (MLCT-D cantilevers with nominal spring constant of 30 pN/nm from Bruker) for 30 min followed by washing 3X in water. For this prepare the piranha solution in a glass petridish which is placed on the hot plate at 80 °C and place the cantilevers dipped into this solution. Incubate the tips in 1 M KOH for 30 min and wash it 3X with water. Then wash it 1X with methanol. The surface would be hydrophilic at this stage. Now prepare 3 % APTES solution (100 ml methanol + 5 ml acetic acid + 3 ml APTES (Sigma-Aldrich)) in a glass petridish and carefully dip the cantilevers in this solution, leave them for 30 min and then wash 3X in methanol. Dry the tips under nitrogen gas. Incubate the tips in 2.5 % glutaraldehyde for 3 hrs followed by washing 4 X with water. Incubate the tips in 0.17 mg/ml streptavidin solution for 2 hrs and wash it 2X in water. In order to use smaller quantities of glutaraldehyde and streptavidin solutions, 1.5 ml Eppendorf tube caps are filled with these solutions and one cantilever is dipped into each tube.

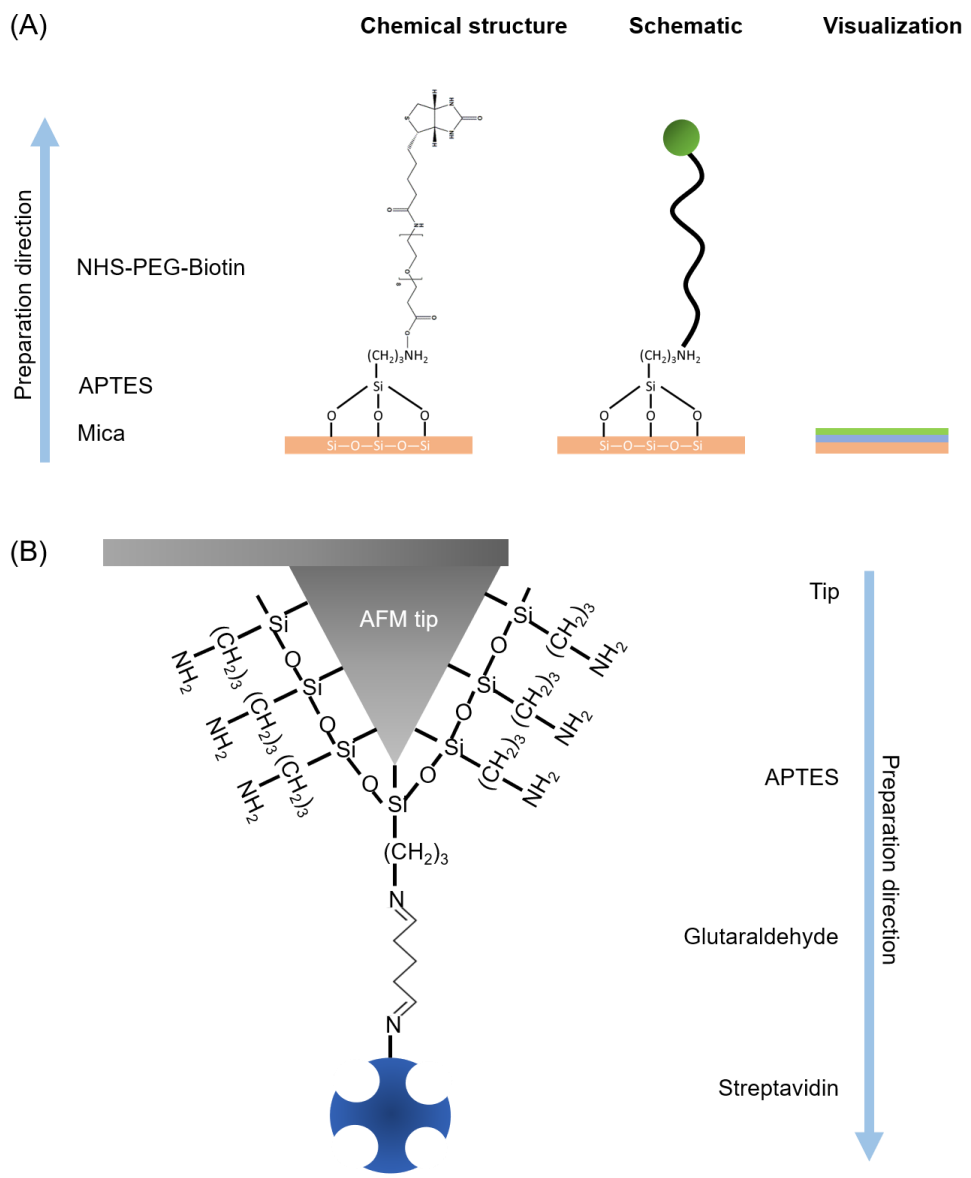


Figure A2.1: Schematics showing the step-by-step procedure for coating the mica and silicon tip surfaces.

(A) Schematic showing the steps involved in coating the mica surface with APTES followed by NHS-PEG biotin. (B) Schematic showing the steps involved in coating the AFM tip with APTES, glutaraldehyde and streptavidin.

Mica sheets:

Cleave the mica sheets using scotch tape and incubate them in 1 M KOH for 30 min to make the surface hydrophilic. Wash them 3X in water followed by washing 1X in methanol. Incubate the sheets in 3 % APTES solution for 30 min followed by washing 3X in methanol and dry them under nitrogen gas. Washing the cantilevers is done by sucking the solution by Vacusip from the edges slowly making sure the cantilevers remain in their position. Then methanol is added slowly into the petridish again making

sure the cantilevers are moved away due to the flow and this process is repeated 3X. Prepare a solution of 2 mg Biotin-PEG5000-NHS (Cat# 757802, Sigma) + 80 mg mPEG5000-NHS (Cat# 85973, Sigma) in 640 µl 0.1 M $NaHCO_3$ (freshly prepared) and centrifuge at 15000 rpm for 1 min to remove air bubbles. Add 20 µl of this solution on each mica sheet and store them in humid box (pipette tip storage box filled with water underneath the tip rack) for 5 hrs in a dark place at 4 °C. Wash them 3X in water before performing the experiment. Here, washing is done by filling a petridish with water and dipping the cantilevers and the solution is made to circulate by slowly vortexing the petridish. Then the petridish is replaced with fresh water and the procedure is repeated 3X.

STV-biotin force spectroscopy

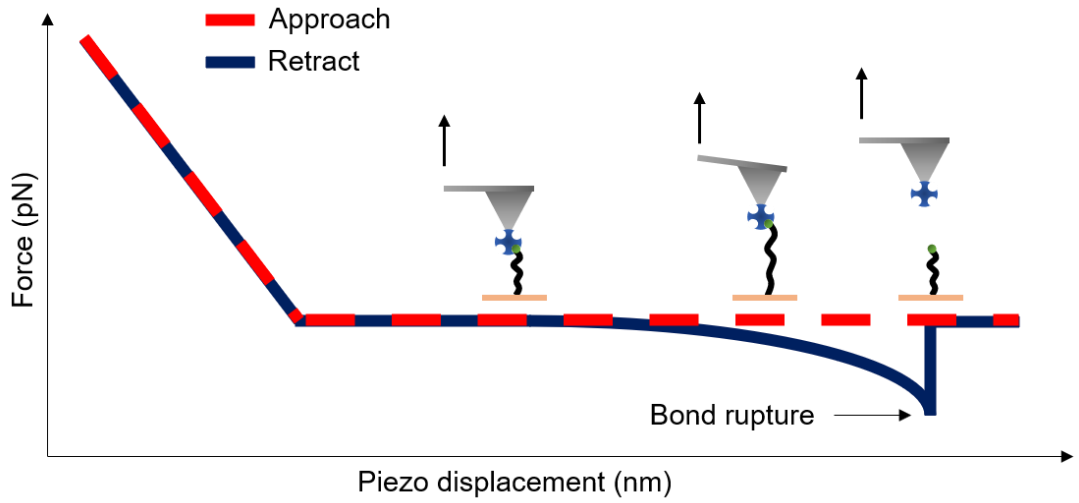


Figure A2.2: Schematics showing the working principle of AFM force spectroscopy

Force spectroscopy is used to measure the bond strength between streptavidin and biotin molecules. The streptavidin is immobilized on the tip surface and the PEG-biotin is coated on the mica surface. The cantilever is brought into contact with the mica surface where the molecules on either side are allowed to interact and then the tip is pulled away leading to the bond rupture. Figure shows the behavior of an F-x curve during this experiment.

To study the interaction between single molecules of interest, we attach STV molecules on a sharp tip and the biotin molecules with a PEG-linker on the mica surface with click chemistry approaches [162] as explained in the methods section (Fig A2.1). The working principle of the AFM force spectroscopy for estimating the bond strength is shown in a

schematic in Fig A2.2. When the tip approaches the surface, the molecules present on either side get an opportunity to bind to each other. Now when the tip retracts, the successful binding of STV and biotin results in the stretching of the PEG linker covalently bound to the biotin. The force-distance (F-x) curve follows the expected worm-like chain model for the polymer extension under force. Once the applied force exceeds the interaction strength of STV-biotin bond, the bond ruptures resulting in a sudden change in the force experienced by the cantilever and it freely moves experiencing zero force. The force at which the bond ruptures is calculated which signifies the bond strength.

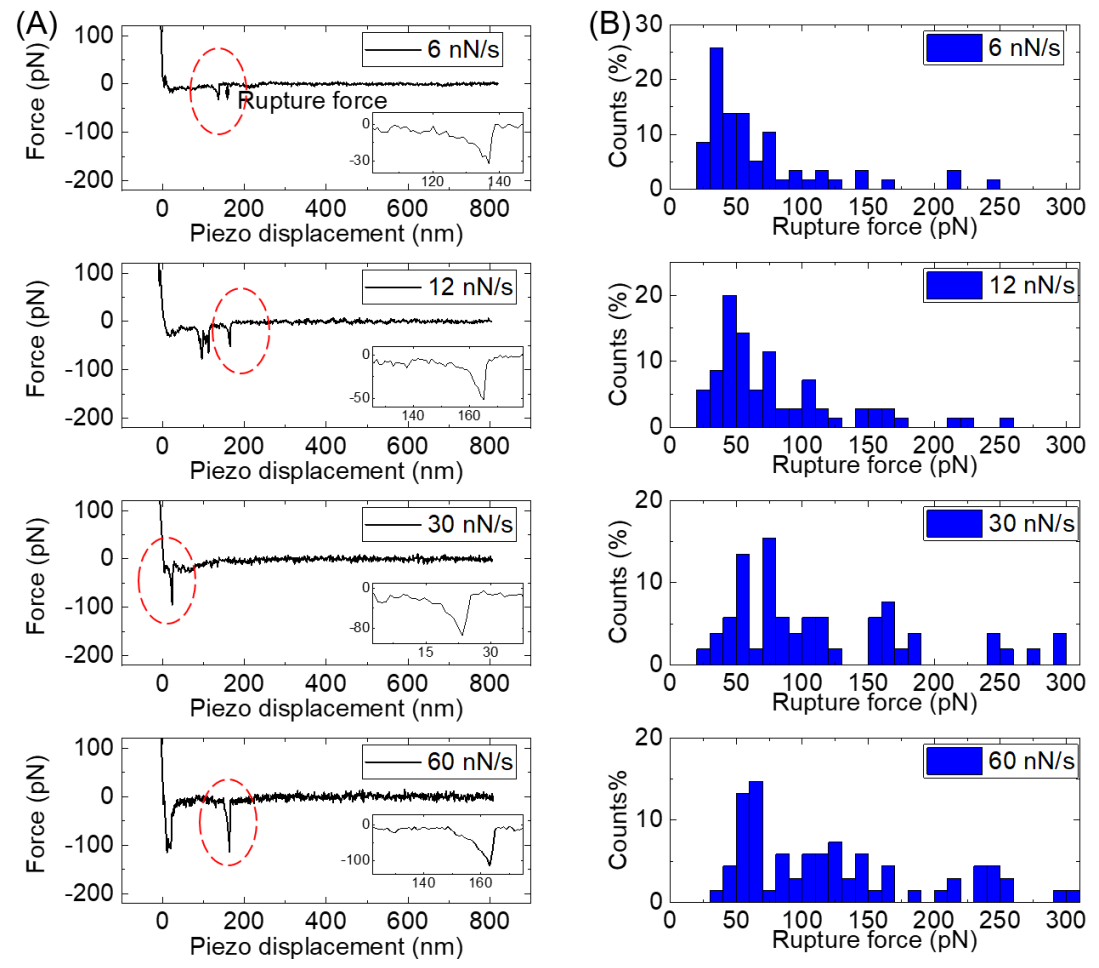


Figure A2.3: Histograms of rupture forces at various loading rates.

(A) & (B) show the representative force-distance curves and rupture force histograms at various loading rates of 6, 12, 30 and 60 nN/s respectively recorded using MLCT-D cantilevers. Insets in (A) show the zoom of the ruptures marked in red dashed circles.

The mica surface is placed in a liquid cell filled with 1X Phosphate-buffered saline (PBS) solution at pH 7.4. The tip is placed about 1000 nm away from the surface and moved in cycles of downward and upward movement in a given duration. This duration is called

the sweeping time which was set to 1 s, 2 s, 5 s or 10 s to generate varying loading rates. We recorded 2048 datapoints for every sweep done by the cantilever. The F-x curves are recorded at multiple locations on the mica surface at various loading rates of 6, 12, 30 and 60 nN/s. The F-x curves are initially baseline corrected and origin corrected and the deflection voltage on y-axis is converted to applied force as explained in the Appendix 1. The representative F-x curves at various loading rates are shown in Fig A2.3A. We can see that the F-x curves has spikes below the baseline level due to the interaction between the tip and the mica surface. The magnitude of peak height is called the rupture force. The histograms of the rupture force are plotted as Fig A2.3B for different loading rates of 6, 12, 30 and 60 nN/s. As expected from the Bell-Evans model [163,164], the data shows that the rupture force increases with the loading rate. Performing such measurements could provide us an insight into the energy landscape of the interacting molecules.

Appendix 3: Additional information on Cas9-DNA binding

MATERIALS AND METHODS

The Cas9 Nuclease, *S. pyogenes* (MO386S) for the experiments was purchased from New England Biolabs (NEB) and the target DNA sequence used in this study was chosen on the supercoiled plasmid pGEM-3z/601 (3025 bp) (gift from Cees Dekker lab, TU Delft). This plasmid was transformed into, amplified and purified from *E. coli DH5α* bacterial strain grown at 37 °C overnight at 180 rpm under ampicillin antibiotic pressure. Nicked circular and linear forms of this plasmid were generated by nicking the plasmid with Nt.BspQ1 (NEB, R0644S) and digesting with ScaI (NEB, R3122S) or NotI (NEB R0189S) enzymes, respectively. The target sequence on the plasmid used in this study is: 5'GGCACCGGGATTCTCCAGGG3'. The corresponding CRISPR RNA (crRNA) and *trans*-activating crRNA (tracrRNA) were purchased from Integrated DNA Technology (IDT, USA) and were resuspended in nuclease free buffer (30 mM HEPES, 100 mM potassium acetate, pH 7.5, provided by manufacturer) to make 100 µM stock of each RNA. The 1 kb DNA ladder (N3232S) purchased from NEB was used as size standard for all the gel electrophoresis experiments. Note that the Cas9 and the duplex guide RNA are labelled in short forms as C9 and R in all the gel images.

Duplex RNA formation:

The crRNA and tracrRNA were mixed in nuclease free buffer at an equimolar concentration to prepare a final duplex guide RNA concentration of 1 µM. The mixture was heated to 90 °C for 30 sec followed by slow cooling to room temperature (1 hr) to make the crRNA-tracrRNA duplex (RNA).

Cas9-RNA and DNA binding reaction:

The RNA was loaded on to the Cas9 enzyme by mixing them in 1:1 molar ratio at 50 nM concentration and incubating for 15 minutes at 37 °C in cleavage buffer (20 mM Tris-HCl, 100 mM KCl, 5 mM $MgCl_2$, 5 % glycerol, 1 mM DTT, pH 7.5). Binding to target was achieved by adding the target DNA to the loaded Cas9 in the molar ratio of 1:50:50

(DNA: Cas9: RNA at 1 nM:50 nM:50 nM) and incubating for 1 hr at 37 °C. We use the reaction of *apo*-Cas9 (Cas9 without RNA) with DNA as a control.

Time-dependent binding kinetics:

After the binding reaction, samples were removed at respective time points t = 0 s, 4 s,10 s, 30 s, 40 s, 50 s, 1 min, 2 min, 3 min, 10 min, 30 min and 60 min into separate tubes containing 25 mM EDTA and incubated on ice for 30 min to stop the reaction.

Gel analysis:

The reaction results were visualized on 1.5 % agarose gel (ran at 90 V for 90 min in 1X TAE buffer), stained in SyBr Gold and imaged using a UV Transilluminator. The quantification of band intensities was done in ImageJ according to the procedure given in methods section of supplementary file. Full-length raw images of all the gels quantified in this study are included in the supplementary information file. The fitting of the data was performed using OriginLab software.

AFM imaging:

For AFM imaging, mica sheets (71853-15, Muscovite Mica V4 grade, Electron Microscopy Sciences) were first cleaned using methanol and dried under nitrogen. Freshly cleaved mica substrate was incubated for 2 min at room temperature with a 25 µl drop of 0.0005% Poly-L-Lysine (PLL, P4832, Sigma) solution in MilliQ water. Washed and dried PLL coated mica (PLL-mica) was then used for sample deposition. The samples (DNA or Cas9-DNA complexes) were diluted in 1X TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8) to a final concentration of 1 ng/µl and mixed well. 25 µl of the prepared sample was dropped on PLL-mica and incubated for 10 min at room temperature. The substrate was then washed with MilliQ water and desiccated for 10 min to dry and then imaged. Molecular Imaging AFM (Pico Plus - Pico Scan 3000) was operated in tapping mode in air using TAP190-G probes (190 kHz resonant frequency, 48 N/m spring constant, Budget Sensors) for all images. The system was controlled using PicoView 1.14 software. The imaging was performed over a region of 2 x 2 μm^2 with a scan speed of 4 µm/s (line scanning rate of 1 Hz) and a resolution of 512 x 512 px^2 using the high-resolution scanner of 10 µm range.

Appendix 4: Additional information on sc-DNA data

Pore ID	D_P (nm)	G (nS)	ΔG_1 (nS)	ΔG_2 (nS)	Δt (ms)
P01	68	147.71	0.2		0.34
P02	76	207.04	0.3		0.53
P03	93	235.85	0.2		0.55
P04	77	242.13	0.2		0.30
P05	85	201.21	0.3		0.32
P06	74	255.75	0.2		0.41
P07	81	138.31	0.4	0.8	0.28
P08	50	174.83	0.3	0.4	0.32
P09	54	183.82	0.3	0.7	0.40
P10	18	81.43	1.9	3.8	0.15
P11	21	101.42	1.1	2.2	0.31
P12	21	84.75	1.6	3.2	0.20
P13	22	83.40	1.75	3.4	0.27
P14	17	87.34	2.0	4.0	0.20
P15	19	78.31	1.4	2.7	0.16
P16	21	87.64	1.3	2.5	0.17
P17	16	50.35	1.8	3.6	0.21
P18	20	97.94	1.0	1.9	0.18
P19	19	84.53	2.05	4.0	0.17
P20	22	100.60	1.4	2.6	0.14
P21	18	81.04	1.5	3.0	0.20
P22	30	110.25	0.6	1.0	0.26
P23	23	100.50	0.8	1.5	0.22
P24	13	58.69	2.8	5.6	0.20
P25	17	66.14	1.85	3.7	0.21
P26	18	77.46	1.9	3.8	0.15
P27	17	60.75	0.9	1.7	0.16
P28	16	64.85	1.4	2.8	0.17
P29	18	77.82	1.3	2.6	0.19
P30	39	159.24	0.4	0.9	0.22
P31	19	68.45	1.1	2.1	0.21
P32	17	61.80	1.2	2.2	0.19
P33	14	54.59	2.1	4.3	0.13
P34	18	48.10	1.1	2.1	0.20
P35	20	54.95	0.9	1.7	0.18
P36	22	88.97	1.5	2.9	0.15
P37	19	62.07	2.0	3.9	0.14

Table A4.1: Summary of sc-DNA translocation data

The data was obtained from a total of 37 nanopores as shown in Fig 3.3A (inset).

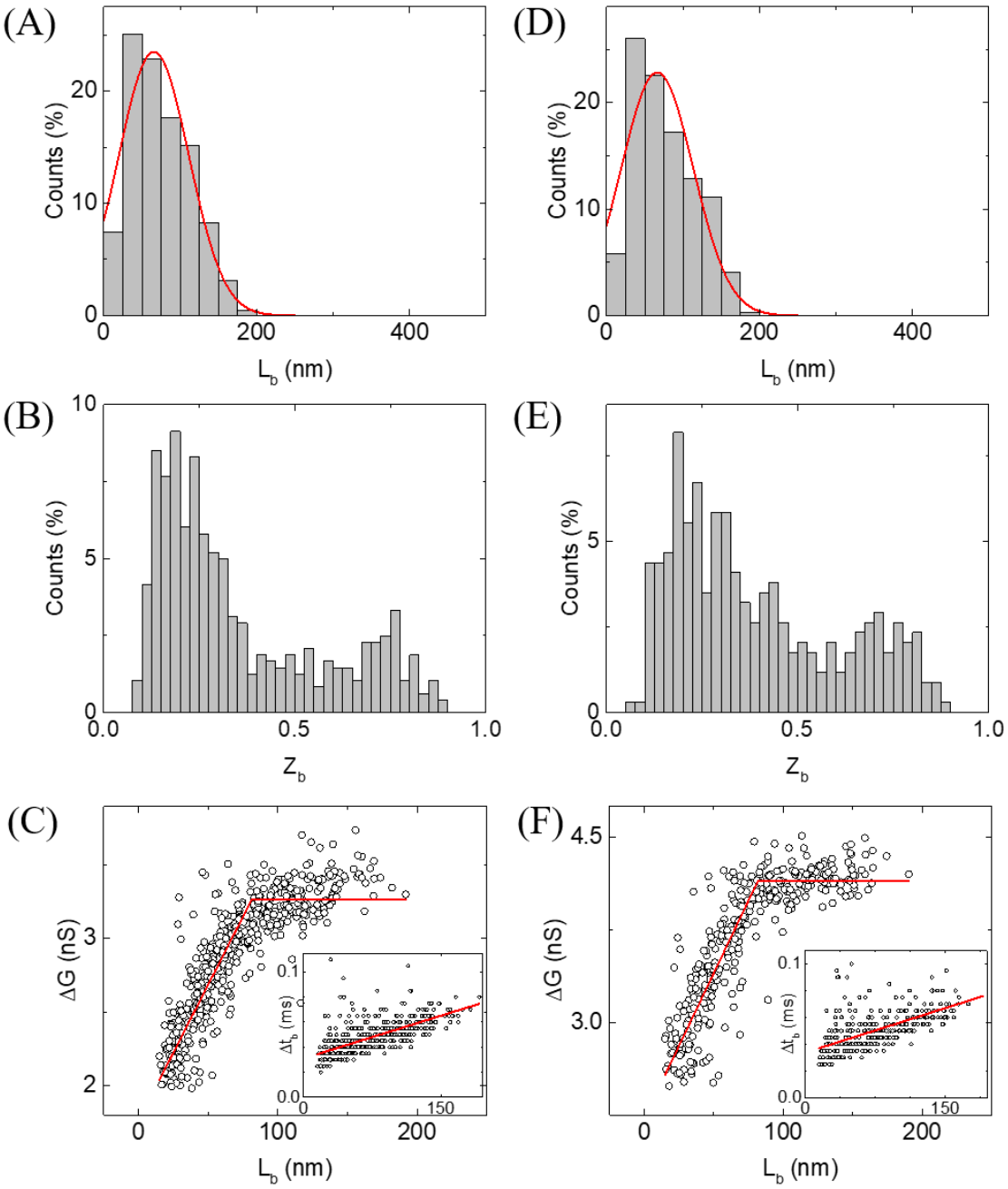


Figure A4.1: Reproducibility of scDNA branch analysis (P12, P19)

Analysis of branches in sc-DNA as measured in nanopores P12 (n = 1192), P19 (n = 745)

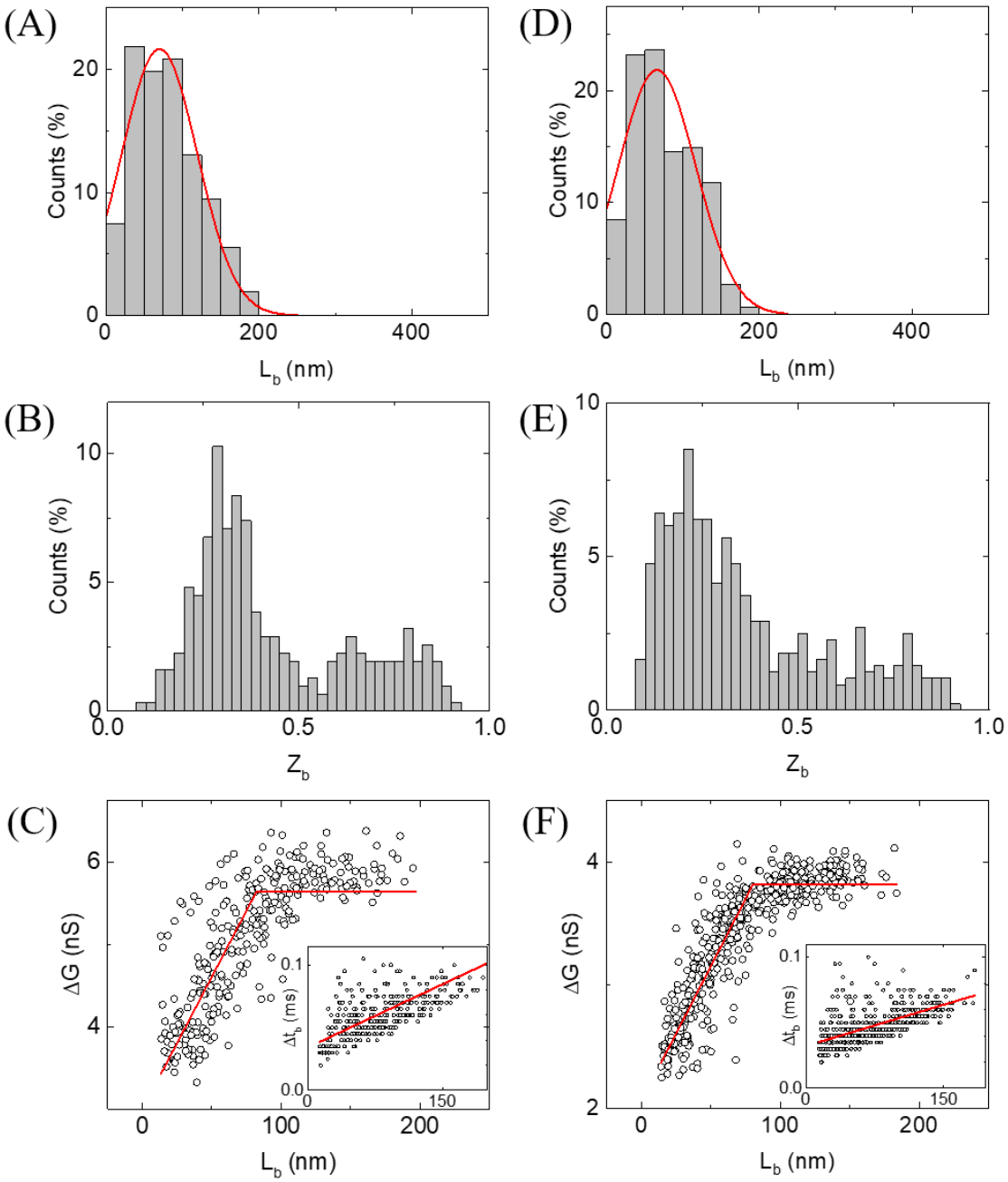


Figure A4.2: Reproducibility of scDNA branch analysis (P24, P26)

Analysis of branches in sc-DNA as measured in nanopores P24 (n = 752), P26 (n = 1008)

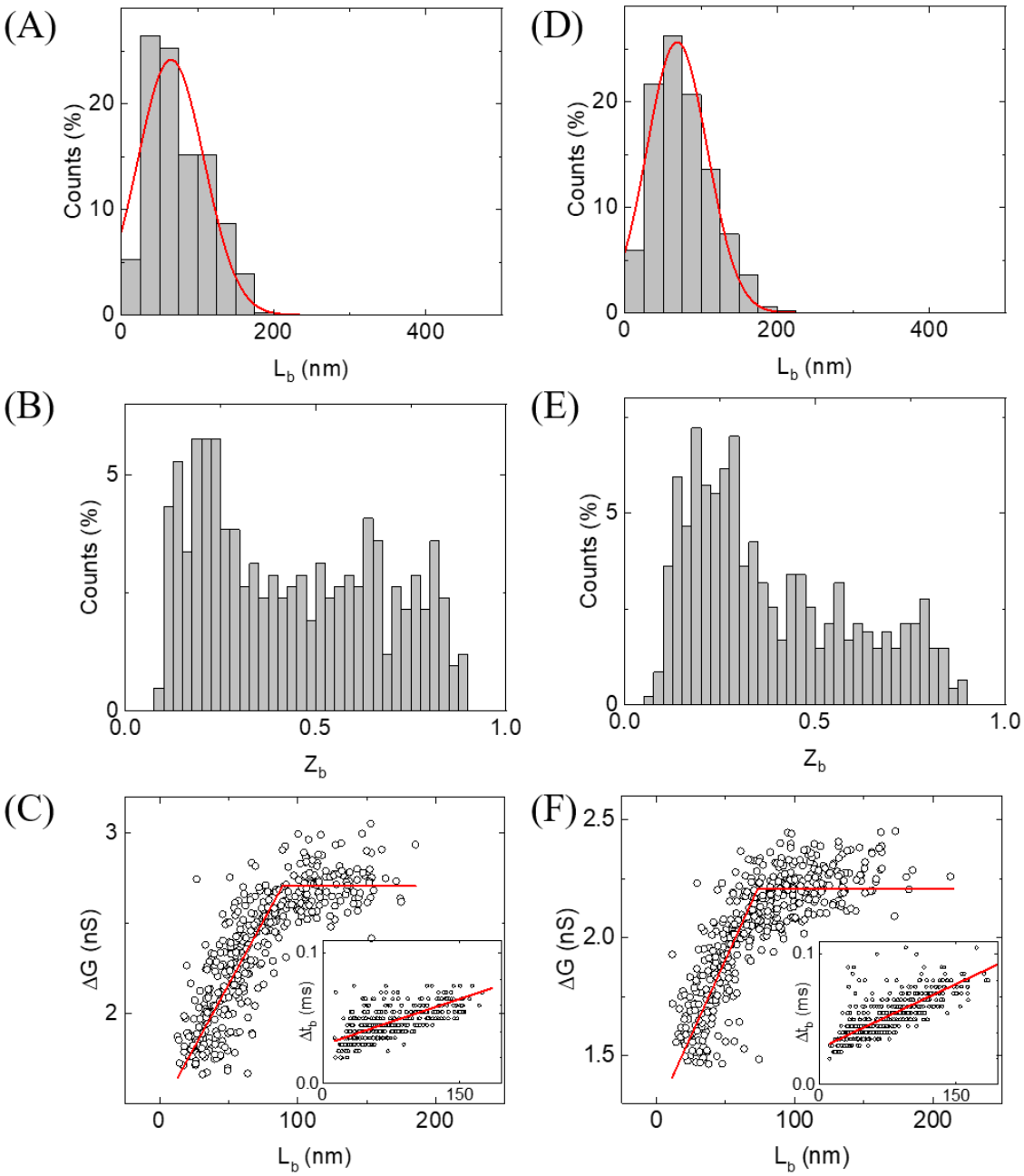


Figure A4.3: Reproducibility of scDNA branch analysis (P29, P31)

Analysis of branches in sc-DNA as measured in nanopores P29 (n = 1249), P31 (n = 960).

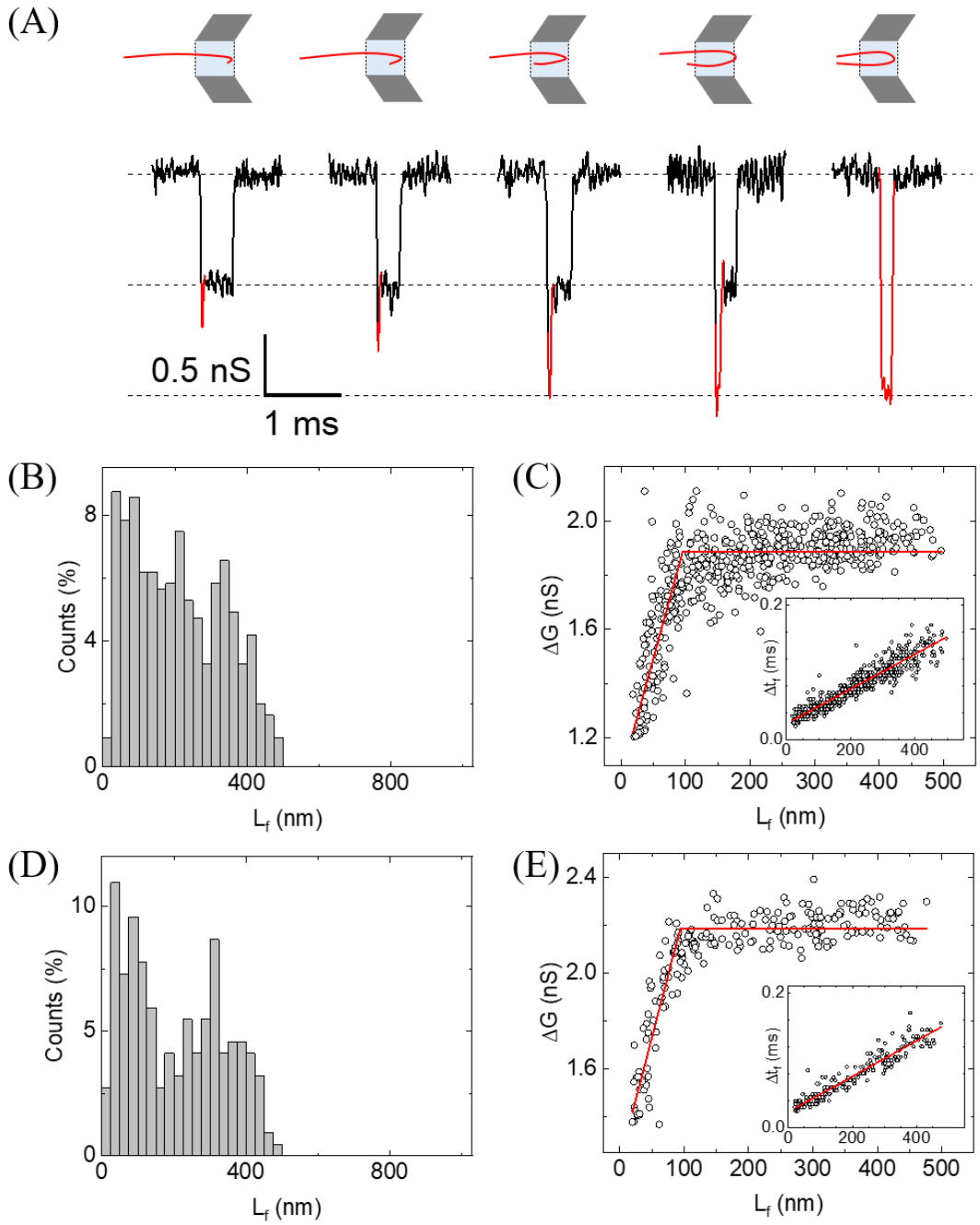


Figure A4.4: Analysis of spikes performed on 3 kb linear DNA TL data

(A) Events due to various lengths of folding from P37 (left to right: 25, 50, 100, 250, 500 nm). (B) & (D) Histograms of fold length for P37 and P33. (C) & (E) Conductance drop as function of fold length. Insets: Fold dwell time as a function of fold length for the respective pores.

Appendix 5: Reproducibility of Mononucleosome data

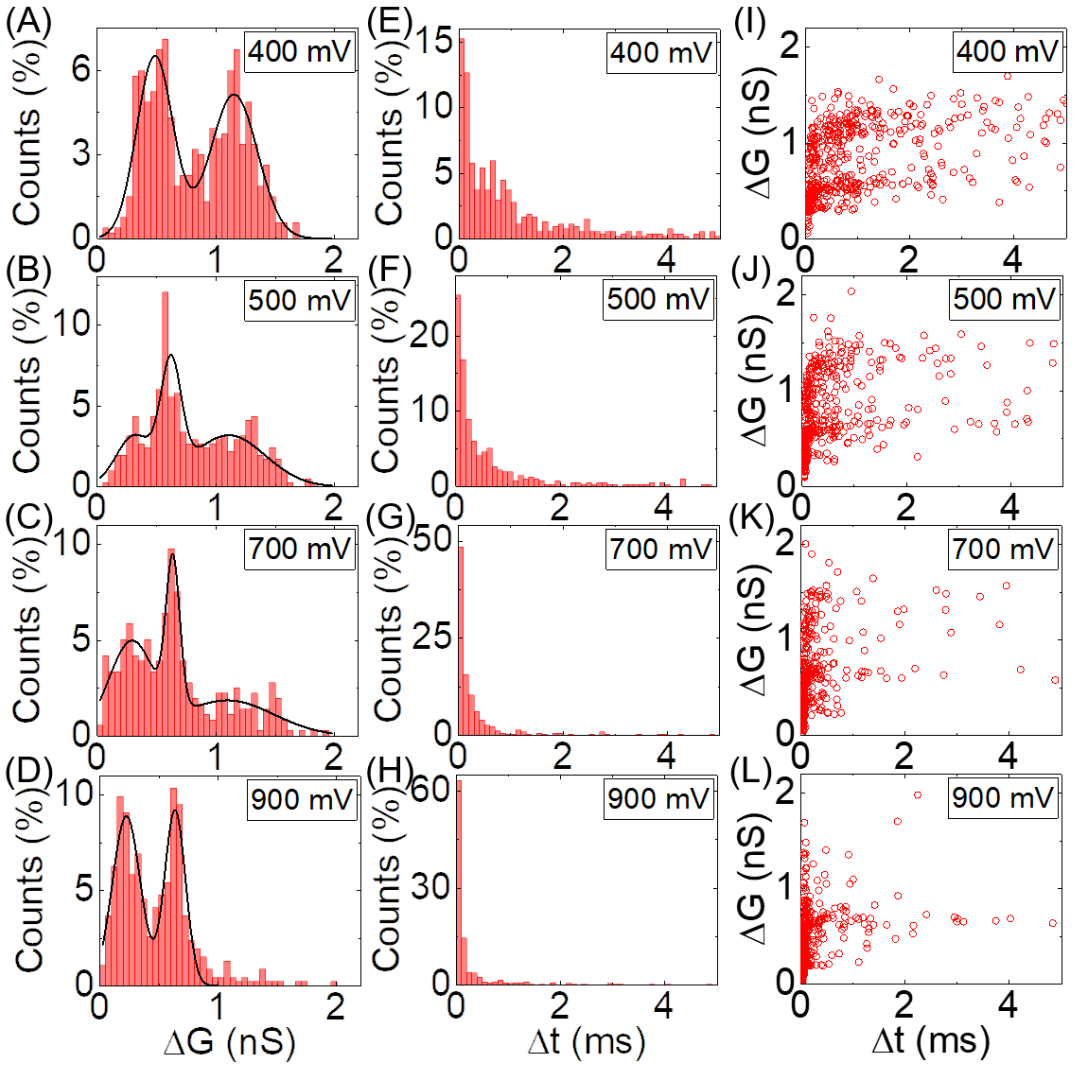


Figure A5.1: Statistical analysis of nucleosome translocation.

(A)-(D) ΔG histograms for nucleosome translocation at 400, 500, 700 and 900 mV respectively. Figures (A)-(D) are reproduced from Fig 4.3 for clarity. (E)-(H) Δt histograms for nucleosome translocation at 400, 500, 700 and 900 mV respectively. (I)-(L) ΔG-Δt scatter plots for nucleosome translocation at 400, 500, 700 and 900 mV respectively. All the nucleosome data shown here is obtained using Pore-1 (18 nm).

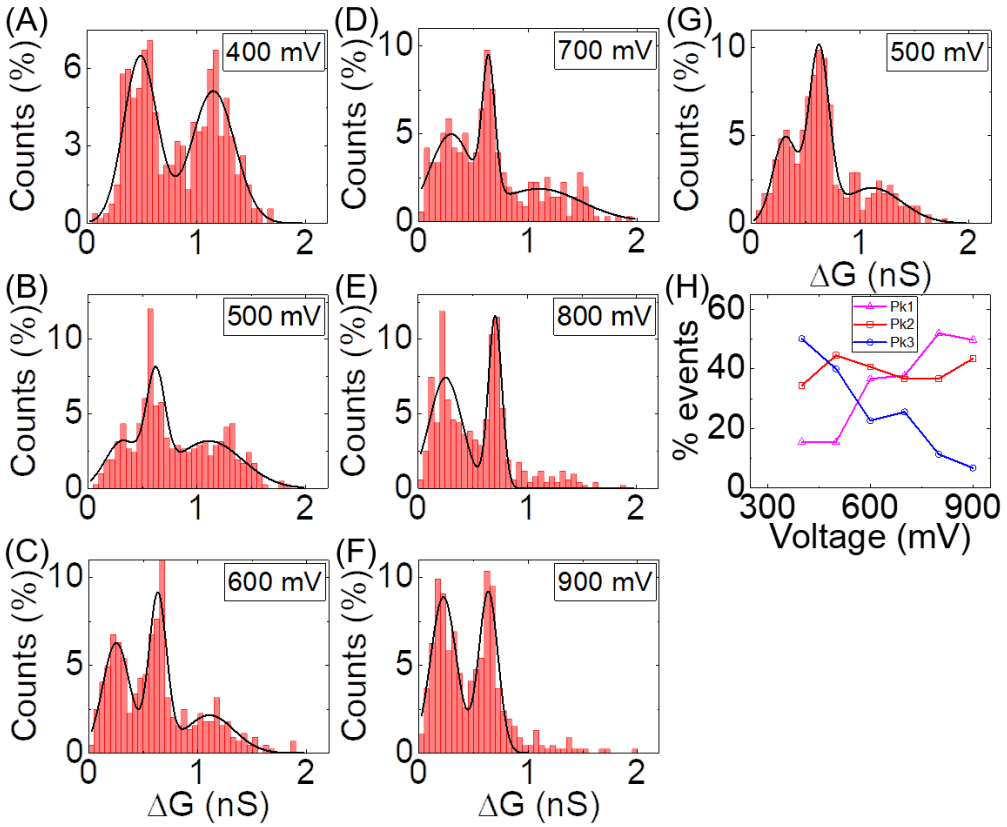


Figure A5.2: Full dataset corresponding to Fig 4.3

(A)-(G) ΔG histograms for nucleosome translocation at 400 (N = 538), 500 (N = 417), 600 (N = 447), 700 (N = 360), 800 (N = 523), 900 mV (N = 464) and 500mV (2nd round, N = 416) respectively. (H) Quantification of % events in each peak in ΔG histogram at various voltages shown in (A)-(F). Data points in stars represent the data at 500 mV (2nd round). Figures (A), (B), (D) and (F) are reproduced from Fig 4.3 for clarity. Data was obtained using Pore-1 (18 nm) and Sample 2 of nucleosomes. The experiments were performed at different voltages in the order 400, 500, 700, 900, 600, 800, 500 mV respectively.

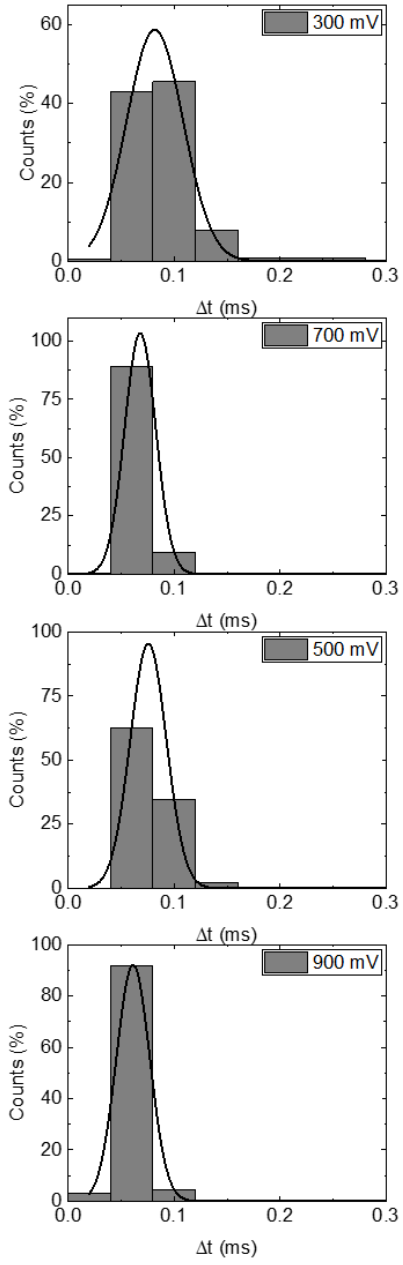


Figure A5.3: Statistical analysis of dwell times of 5 kb DNA as a function of voltage

Data shown here corresponds to Pore-1.

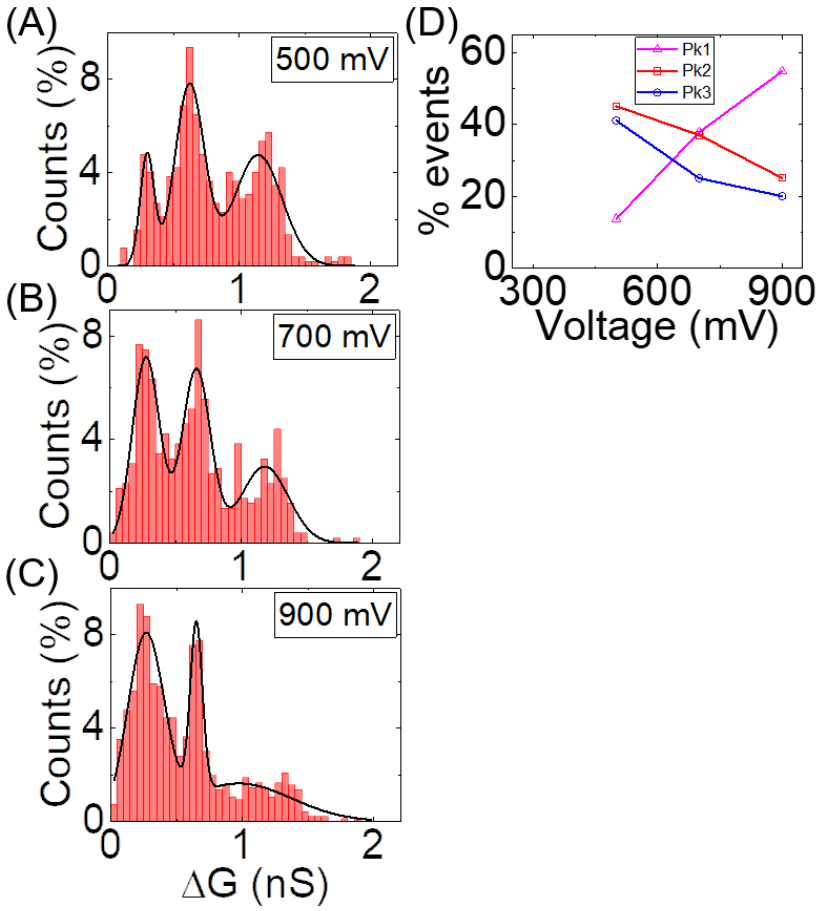


Figure A5.4: Statistical analysis of nucleosome (Sample 2) translocation (to show data reproducibility).

(A) – (C) ΔG histograms for nucleosomes at 500 (N = 525), 700 (N = 523) and 900 mV (N = 968) respectively in Pore-2 (20 nm) and in the order 500, 900, 700 mV respectively.

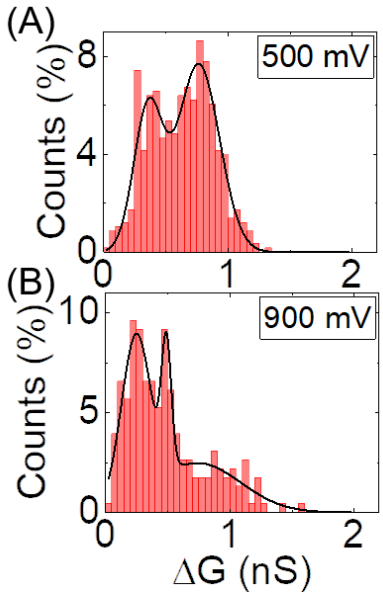


Figure A5.5: Statistical analysis of nucleosome (Sample 2) translocation (to show data reproducibility).

(A) – (B) ΔG histograms for nucleosomes at 500 (N = 579) and 900 mV (N = 229) respectively in Pore-7 (19 nm) and in the order 500 and 900 mV respectively.

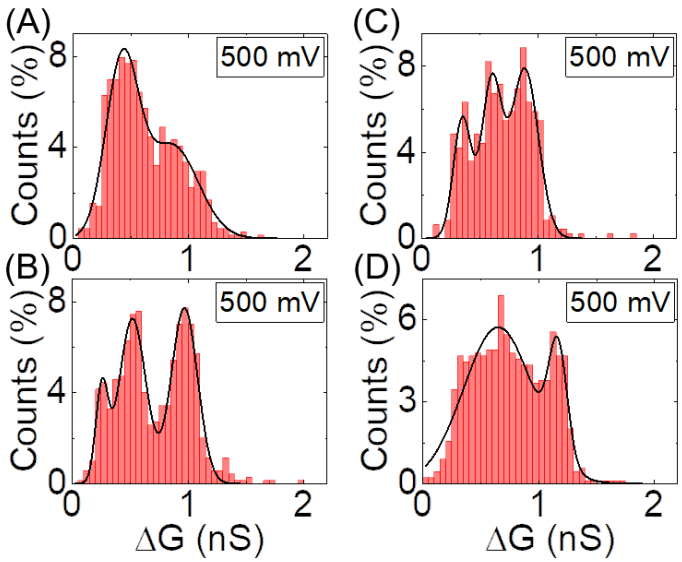


Figure A5.6: Statistical analysis of nucleosome translocation (to show data reproducibility) with independent sample preparation.

(A) – (D) ΔG histograms for nucleosomes at 500 mV in Pore-3 (19 nm, N = 717), Pore-4 (24 nm, N = 700), Pore-5 (23 nm, N = 476) and Pore-6 (23 nm, N = 899) respectively.

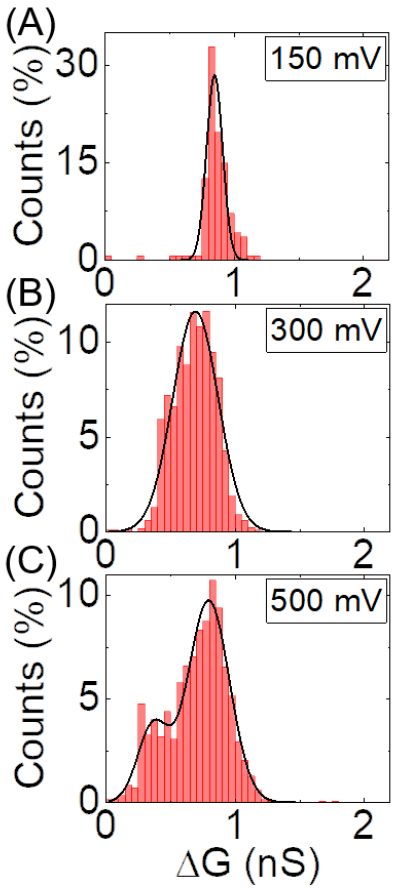


Figure A5.7: Statistical analysis of 147 bp nucleosome translocation.

(A) – (C) ΔG histograms for nucleosomes at 150 (N = 169), 300 (N = 1334) and 500 mV (N = 1577) respectively in Pore-9 (18 nm) and in the order 500, 300 and 150 mV respectively.

Appendix 6: Shape factor calculation

The shape factors of oblate spheroids are estimated as follows [114]:

$$n_{\parallel} = \frac{1}{1-m^2}\left(1 - \frac{m}{\sqrt{(1-m^2)}}\cos^{-1}(m)\right) \qquad \text{(A6.1)}$$

$$n_{\perp} = \frac{1 - n_{\parallel}}{2}$$

$$\gamma_{\parallel} = \frac{1}{1 - n_{\parallel}}$$

$$\gamma_{\perp} = \frac{1}{1 - n_{\perp}}$$

where m = A/B < 1. A and B are minor and major radii of the oblate spheroid. For nucleosome, A = 6 and B = 10. Using eqn (A6.1), we obtain $\gamma_{\perp}$=1.36 and $\gamma_{\parallel}$=1.91. The scenario shown in Fig 6.1A is more probable than that of Fig 6.1B as the free DNA which has higher charge density would be pulled towards the nanopore and hence we consider γ_{nucl}=$\gamma_{\perp}$=1.36.

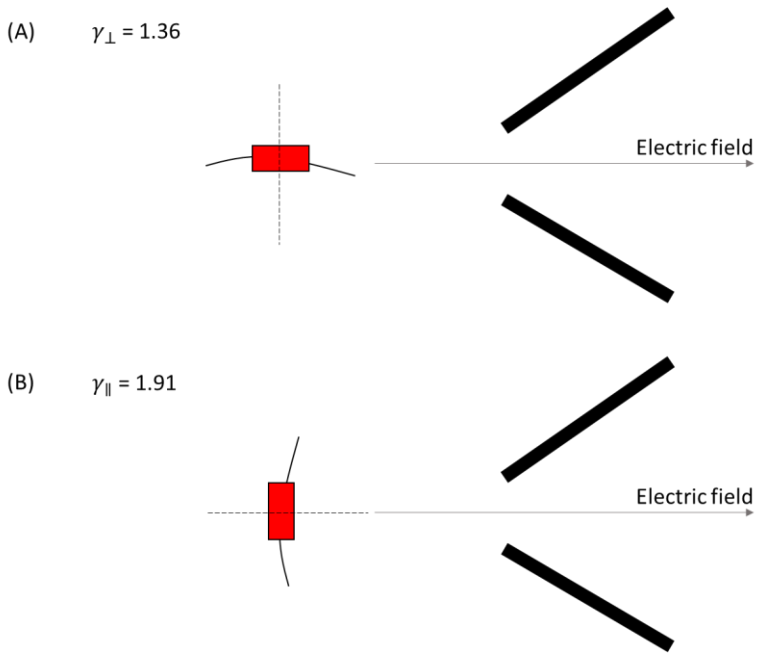


Figure A6.1: Demonstration of possible orientations of nucleosome through nanopore.

(A) & (B) The cylindrical particle passing through the nanopore with its axis perpendicular and parallel respectively to the translocation direction.

Appendix 7: Reproducibility of 12mer data

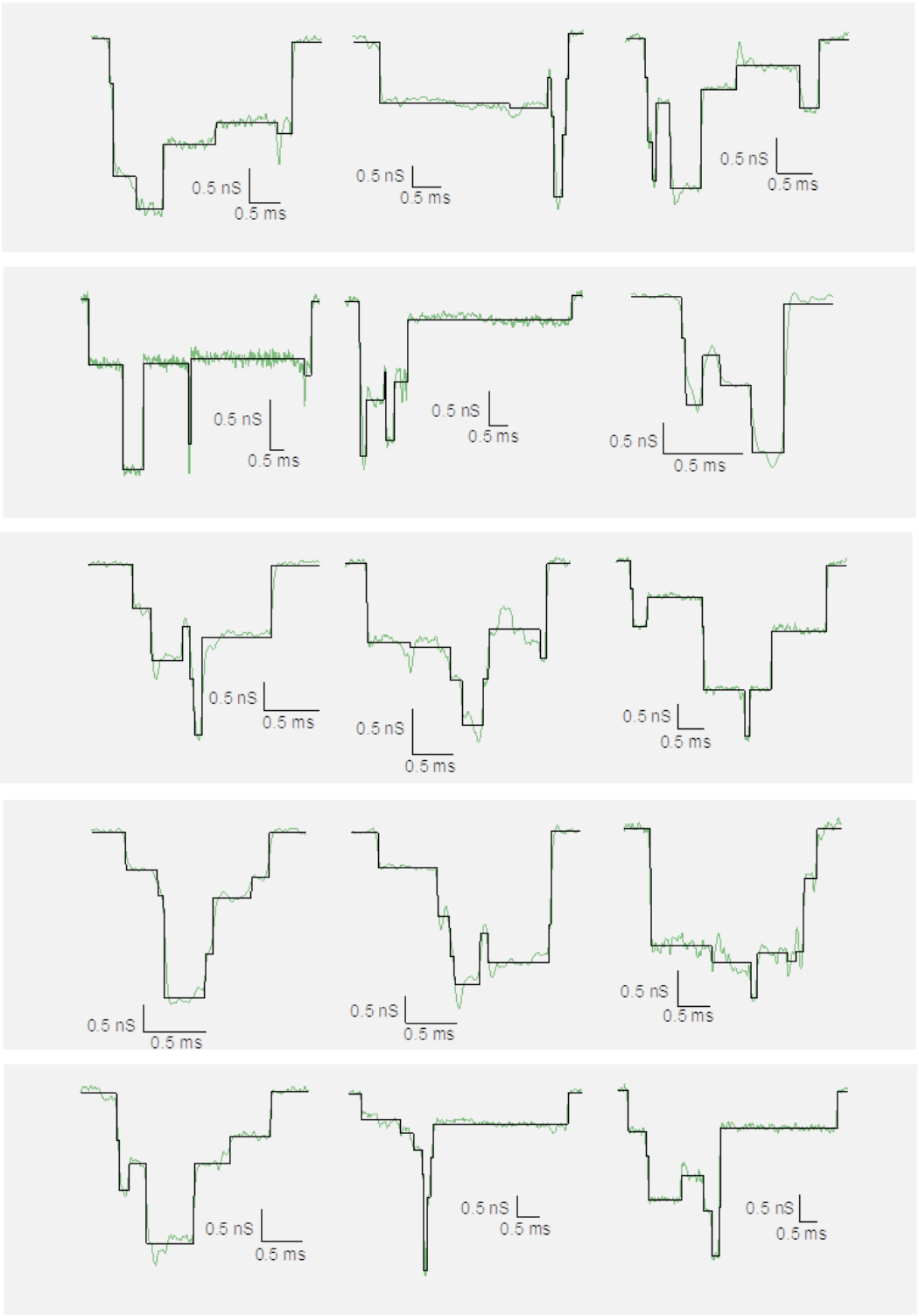


Figure A7.1: Library of 12mer event.

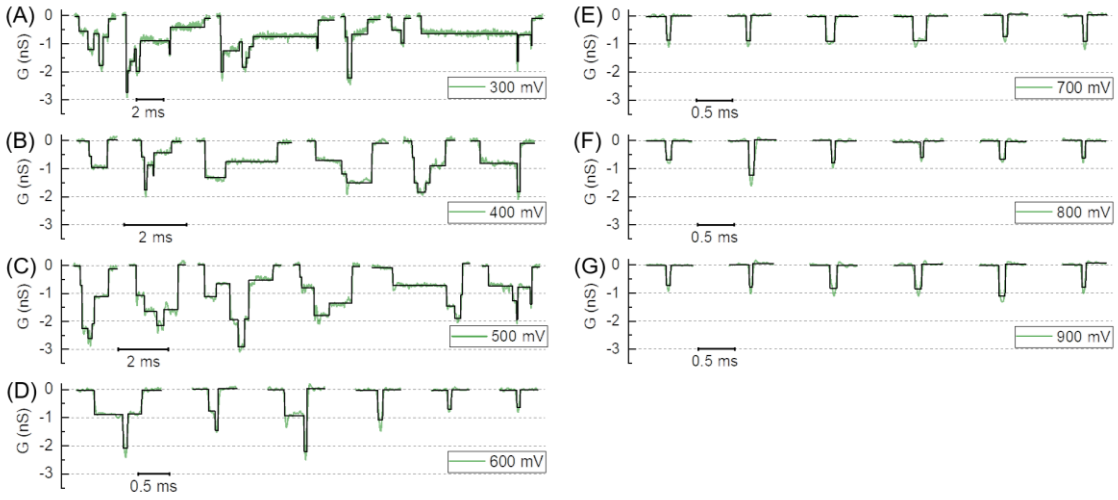


Figure A7.2: Representative events at a complete voltage range.

(A)-(G) Representative events along with their level fits for array translocation at 300, 400, 500, 600, 700, 800 and 900 mV respectively. Data recorded using Pore-10. This is a complete voltage range for the data shown in Fig 3.

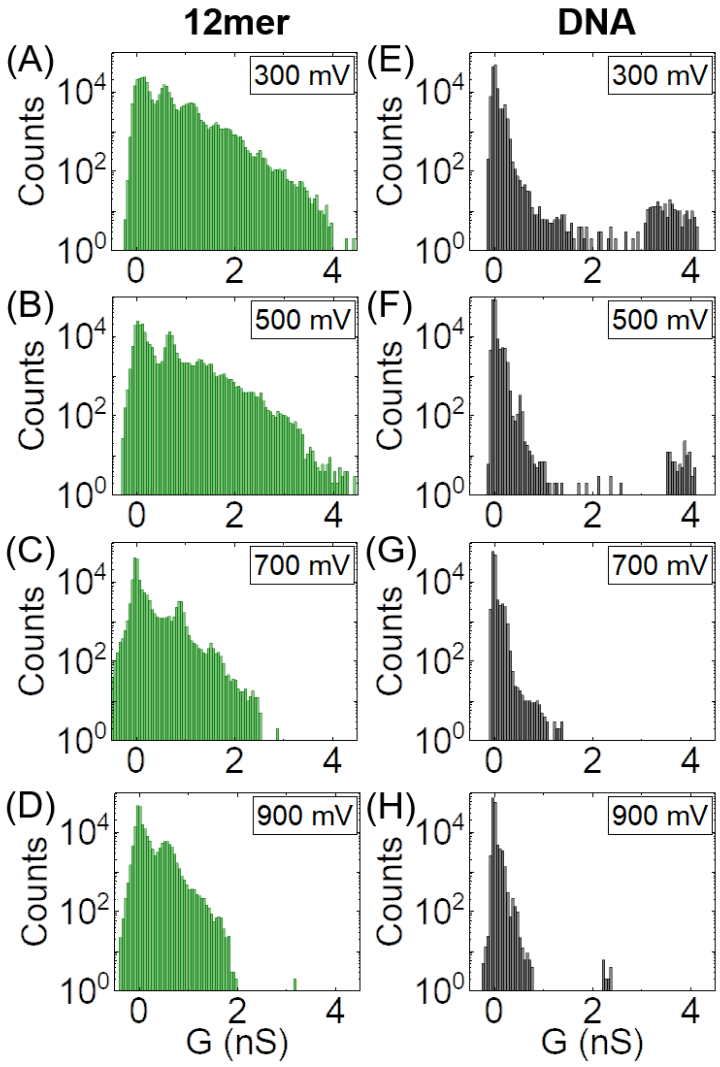


Figure A7.3: Statistical analysis of nanopore conductance (G) for nucleosome array translocation.

(A)-(D) shows the G histograms for array translocation at 300 (N = 564), 500 (N = 534), 700 (N = 929) and 900 mV (N = 1063) respectively in 23 nm pore (Pore-10). (E) – (H) shows the G histograms for control 5 kb linear DNA at 300 (N = 967), 500 (N = 1579), 700 (N = 983) and 900 mV (N = 1204) in 18 nm pore (Pore-1).

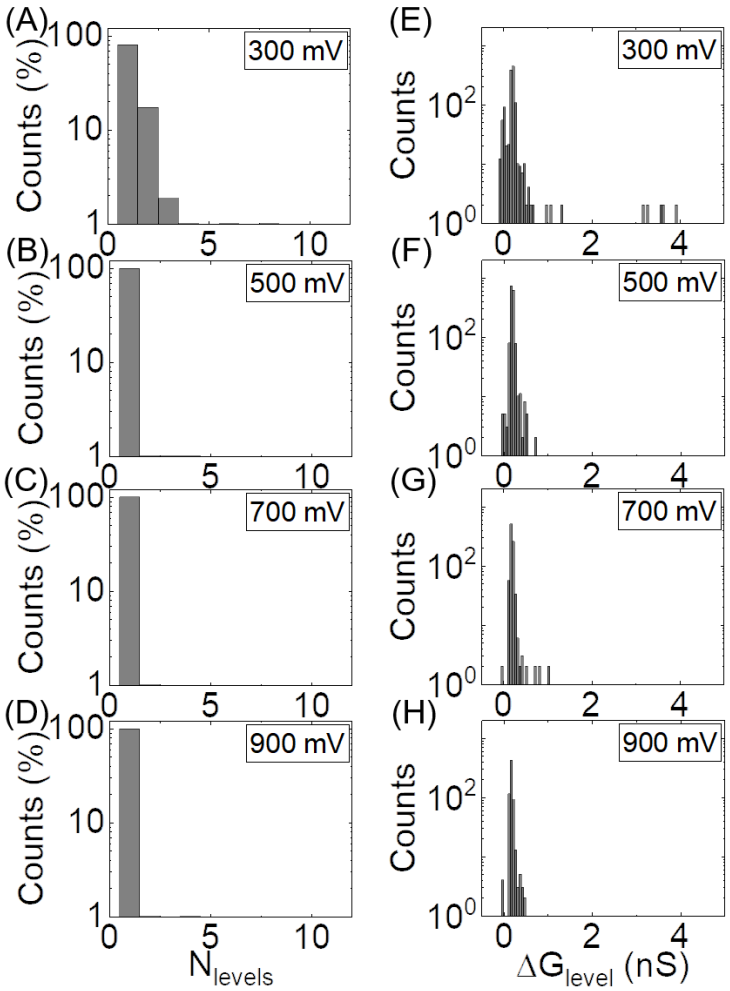


Figure A7.4: Statistical analysis of number of levels per event (N_{levels}) and level-depths (ΔG_{level}) in 5 kb DNA translocation events.

(A)-(D) and (E)-(H) show the N_{levels} and ΔG_{level} histograms for control 5 kb linear DNA measured in the same pore (Pore-1) and at the same voltages, 300 mV (N = 967), 500 mV (N = 1579), 700 mV (N = 983) and 900 mV (N = 1204).

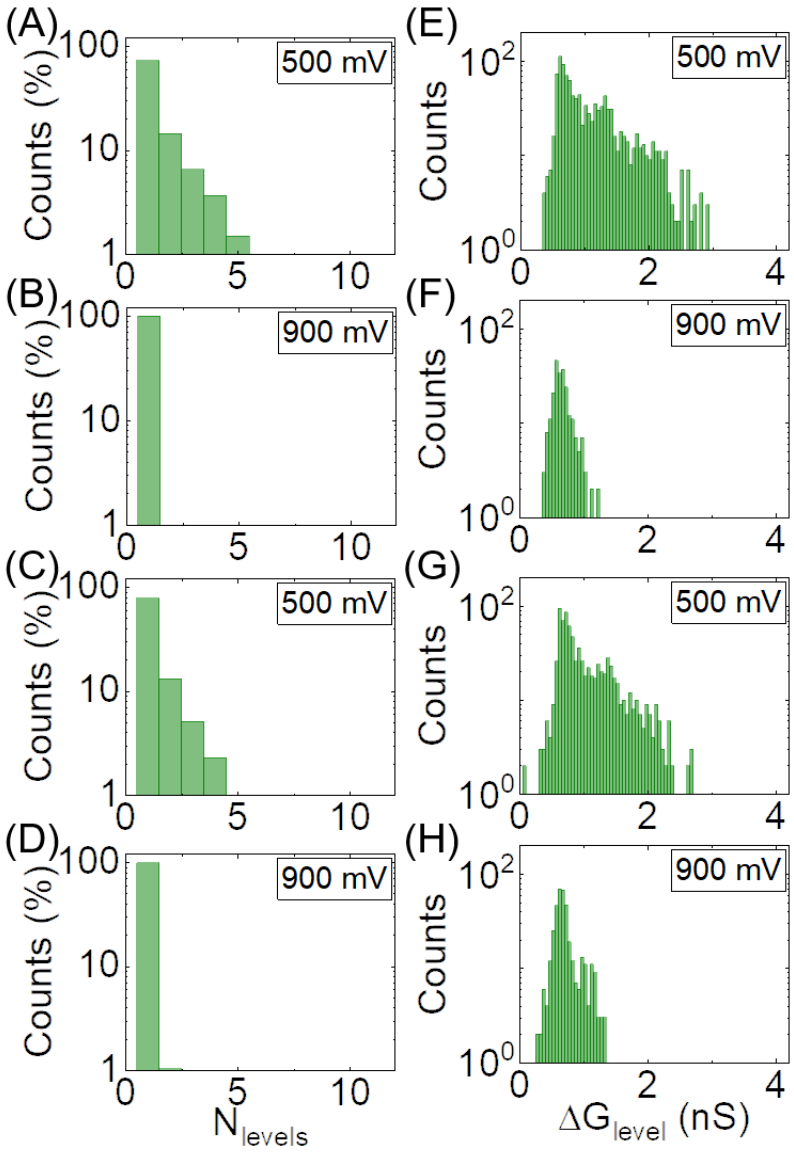


Figure A7.5: Statistical analysis of array translocation (to show data reproducibility).

Plots are shown in the order the experiment was performed. Experiment was performed using Pore-11 (18 nm) and in the order 500 (N = 871), 900 (N = 995), 500 (N = 746) and 900 mV (N = 1059).

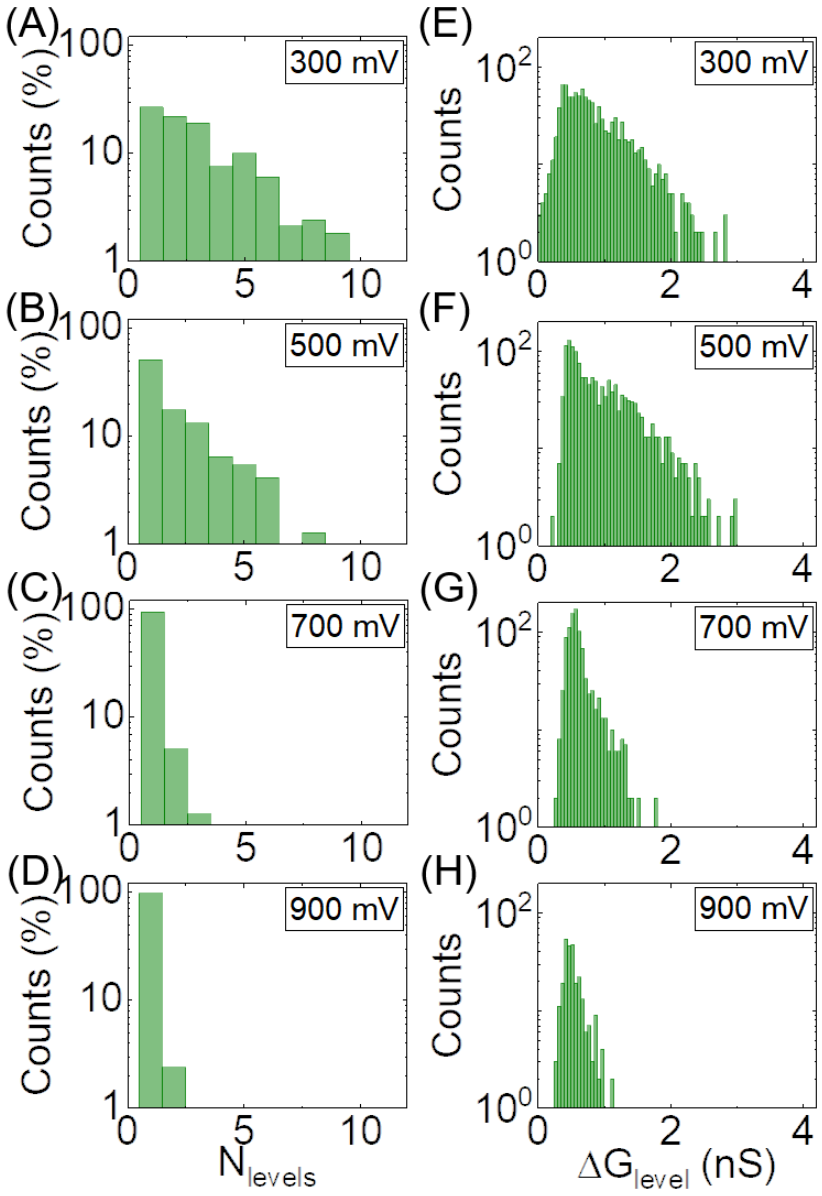


Figure A7.6: Statistical analysis of array translocation (to show data reproducibility).

Experiment was performed using Pore-12 (20 nm) and in the order 500 (N = 701), 900 (N = 422), 300 (N = 331) and 700 mV (N = 1186) respectively.

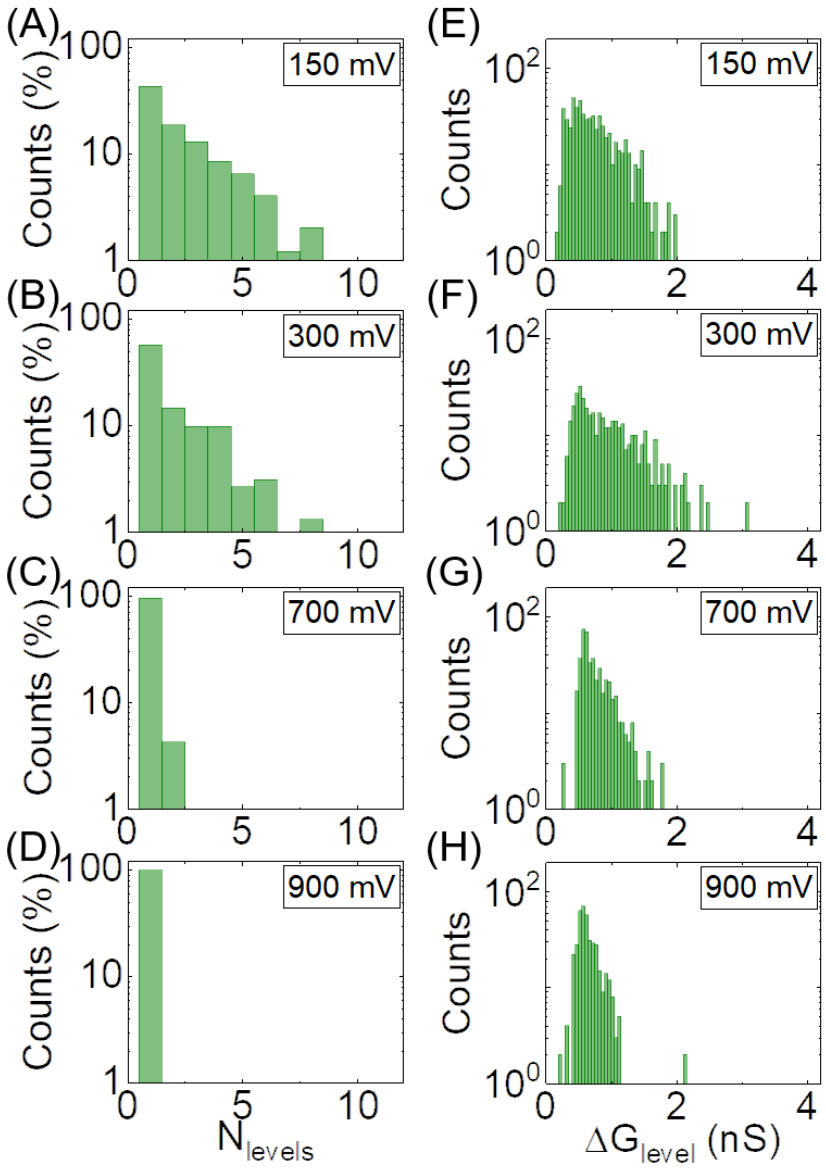


Figure A7.7: Statistical analysis of array translocation (to show data reproducibility).

Experiment was performed using Pore-13 (19 nm) and in the order 300 (N = 225), 900 (N = 943), 150 (N = 246) and 700 mV (N = 730) respectively.

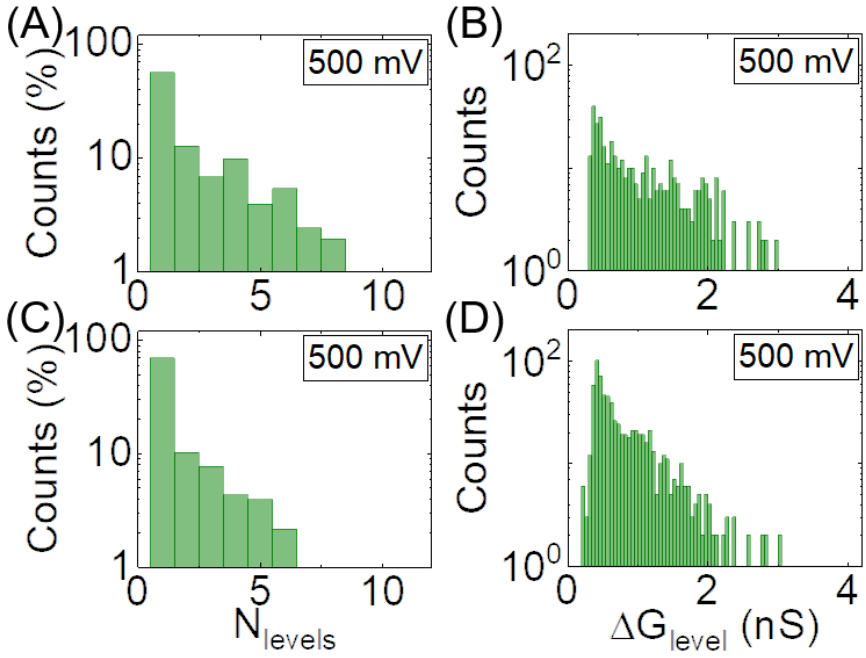


Figure A7.8: Statistical analysis of array translocation (to show data reproducibility).

Experiment was performed using independently prepared sample of 12-mer at 500 mV. (A) – (B) Data collected on Pore-14 (22 nm, N = 204) and (C) – (D) Data collected on Pore-15 (21 nm, N = 507).

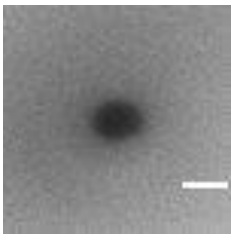
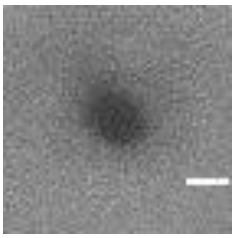
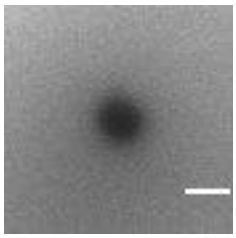
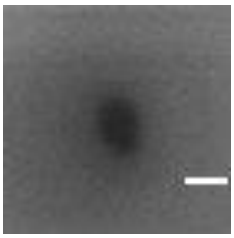
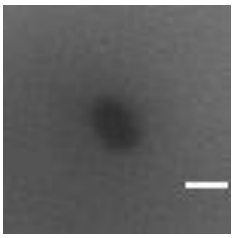
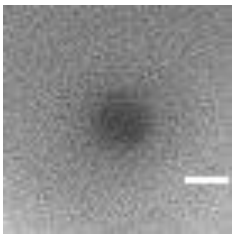
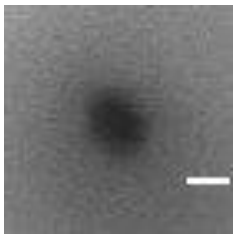
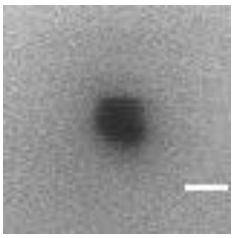
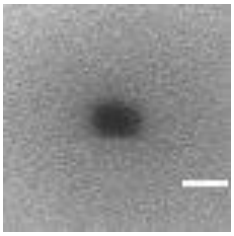
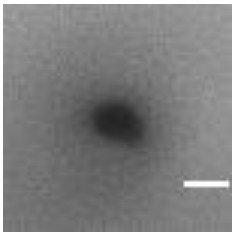
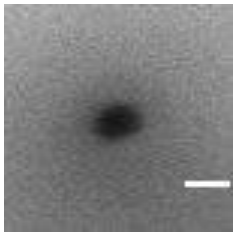
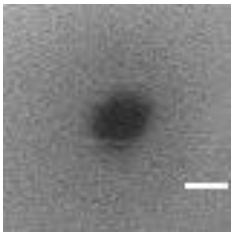
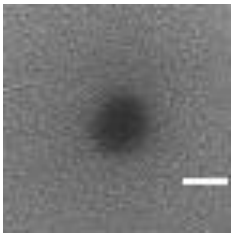
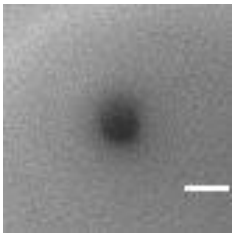
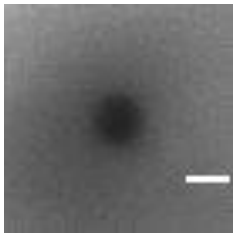
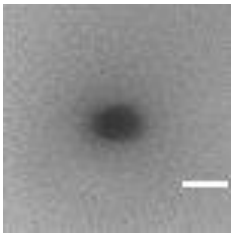
Pore-1: 18 nm	Pore-5: 23 nm	Pore-9: 18 nm	Pore-13: 19 nm
			
Pore-2: 20 nm	Pore-6: 23 nm	Pore-10: 23 nm	Pore-14: 22 nm
			
Pore-3: 19 nm	Pore-7: 19 nm	Pore-11: 18 nm	Pore-15: 21 nm
			
Pore-4: 24 nm	Pore-8: 18 nm	Pore-12: 20 nm	Pore-16: 18 nm
			

Table A7.1: Library of SEM images of nanopores

The images shown here has scale bars of 20 nm.

Synopsis

Thesis Title: **Understanding Physical Mechanisms in Chromatin Folding**

INTRODUCTION

This thesis highlights our investigations on structural forms of DNA polymer and nucleosome arrays. Along with gel electrophoresis and AFM imaging, we primarily use the single-molecule resistive pulse technique (Nanopore platform) to study these structural conformations. This thesis is broadly divided into three parts. In the first part, we measure supercoiled conformations of single DNA plasmids. The supercoiled DNA can assume a variety of interwound branched structures, called plectonemes, which form due to the interplay between twisting and bending energy of the DNA [1,2]. Plectonemes with branches protruding out of the superhelical axis were initially observed by electron microscope [1] and later studied in detail using AFM [3,4] under physiological conditions. Supercoiling which acts as a topological constraint on DNA is a key control mechanism in regulating transcription, replication, and compaction of DNA in both prokaryotes and eukaryotes [1,2]. In this work, we demonstrate the nanopore platform as a label-free technique to quantitatively measure a variety of branched DNA structures in aqueous solutions at nanometer resolution. We develop a detailed event charge deficit (ECD) based analysis to quantify the DNA lengths in these supercoiled branches. This analysis also provided us a method to determine the internal geometrical parameters of the nanopore. Finally, we use the nanopore measurements to study the kinetics of enzyme-dependent conformational changes in supercoiled DNA. Our work extracts a wealth of information about the branched sub-structures in supercoiled DNA demonstrating facets of the nanopore platform useful in the study of the dynamic conformational changes of DNA at high resolution.

In the second part, we study the folding of linear DNA polymers in solution. We measure DNA folds during the translocation process as a function of multiple geometric constraints and DNA length. We report variation in translocation dynamics in forward and reverse translocation as well as on the folding statistics of DNA during these

reversible translocations. A modified version of the ECD analysis was employed for the estimation of fold lengths in linear DNA as well as the effective length of the nanopores.

In the final part of the thesis, we build on the lessons learnt from the free DNA conformational structures and apply them to the nucleosome fibers, a fundamental unit of chromosomes. The chromatin inside a cell is very dynamic due to the interaction with various enzymes during the cellular processes. The chromatin is known to exist in the form of beads on a string where the bead is a nucleosome, the basic unit of chromatin [5]. The nucleosomes are highly dynamic as the genetic machinery needs to access the underlying DNA sequence [6]. In this work, we contrast structural features of free DNA and mononucleosomes with nucleosome fibers (12-mer nucleosome arrays). They are initially analyzed using gel electrophoresis and later the single-molecule observations were done using AFM and finally using nanopores. Finally, we study the interaction of DNA with chromatin architectural proteins (CAP) which are responsible for the maintenance of DNA and chromatin architecture during its condensation or decondensation [7,8]. We close the final chapter with a study on the condensation of DNA and nucleosomes controlled by CAPs.

In Summary, the research work presented in this thesis is divided into the following parts:
1. Quantification of supercoiled DNA branches.
2. Quantification of folds in linear DNA.
3. Detection of folds in nucleosome arrays.
4. Observation of the ACP protein interaction with DNA and nucleosomes.

FINGERPRINTING BRANCHES ON SUPERCOILED PLASMID DNA:

Supercoiling which acts as a topological constraint on DNA is a key control mechanism in regulating transcription, replication, and compaction of DNA in both prokaryotes and eukaryotes [1,2]. Fluorescence and AFM based single-molecule studies have revealed that the plectoneme formation depends on the underlying sequence as it governs the local DNA curvature [9]. The interwinding and branching can juxtapose multiple distant sites which are important for processes such as enhancer-promoter interaction [10]. The negative supercoiling favours the local untwisting of DNA thereby affecting the activity of transcription [11] and restriction digestion [12]. The studies on the kinetics of

interconversion between the linear, circular, and supercoiled plasmid conformations [13] need to be carried out as the DNA inside a cell is highly dynamic due to its interaction with various types of enzymes during the processes like nucleosome formation, recombination, viral genome integration, etc. Mechanisms involved in the maintenance of supercoiling, the kinetics of modulating the supercoiling of DNA by topoisomerases and gyrases have been studied using single-molecule techniques like magnetic tweezers [14] and rotor bead assays [15]. However, there is a dearth of single-molecule measurements required for quantification of the native branched structures and enzyme-dependent conformational changes in DNA supercoiling. In this part of the thesis, we observed the various forms of plasmid which are linear, circular, and supercoiled DNA. Using nanopore and AFM, we have observed the branches on the supercoiled DNA (Fig1).

RESISTIVE PULSE TECHNIQUE USING NANOPORES:

The nanopore platform has recently emerged as an excellent platform for detecting various polymers. The nanopore device works on the principle of resistive pulse technique developed by Coulter [16]. Here the nanopore acts as a channel between two reservoirs filled with electrolyte. When a bias voltage is applied across the nanopore, open pore current is measured due to the movement of electrolyte ions. For the duration an analyte particle translocates through the nanopore, it displaces ions in the nanopore resulting in conductance drop. These electrical events (conductance drops) contain details about the size and shape of the particle. Various types of nanopores are used which are made from silicon oxide [17], silicon nitride [18], carbon nanotubes [19], graphene [20], and glass capillaries [21–23]. They have been used to detect molecules like DNA, RNA, proteins, and protein-DNA complexes [24–26]. The application of nanopores in the detection of conformational changes in proteins is also demonstrated [27–29]. They have been used for measuring the kinetics of biomolecular reactions [29] and to study conformations of DNA in the distribution of knot conformations [30,31], G-quadruplexes [32], and supercoiled DNA [33,34].

All the custom-made nanopores are characterized by optical microscopy and SEM imaging to measure the taper length and tip geometry of the pore respectively. The nanopores were made with glass capillaries and most of the samples were prepared in the lab. LabVIEW codes were used for acquiring the translocation data as well as to analyze the various parameters of the experiment.

CHARACTERIZATION OF NANOPORE AND TRANSLOCATION OF VARIOUS FORMS OF PLASMID:

We demonstrate the capability of our quartz nanopores in detecting multiple DNA conformations using a linearized 48.5 kb long λ-DNA translocating through a nanopore of diameter 20 nm at 500 mV bias voltage. Next, we compare different forms of a 3 kb plasmid DNA which is in linear (lin-), circular (cir-), and supercoiled (sc-) conformations. We show a comparison of the ECD values for the three forms of the plasmid DNA and find their values close to each other. From the above observations, it is clear that the quartz capillary based nanopore platform can successfully distinguish between different plasmid conformations.

QUANTIFICATION OF BRANCHES IN SUPERCOILED DNA USING ECD ANALYSIS:

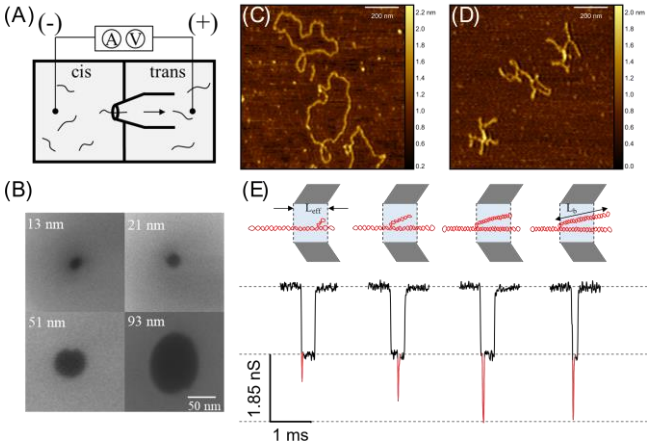


Figure 1: AFM imaging and Nanopore analysis of 3kb plasmid DNA. (A) Schematic of the fluid cell with nanocapillary mounted on a Teflon flow cell. (B) The representative head-on SEM images of nanopores with sizes ranging from 13 nm to 93 nm, used in this study. (C) and (D) are AFM images of circular and supercoiled forms of plasmid DNA respectively. (E) Typical events measured with sc-DNA showing different branch lengths translocating through a 17 nmpore (left to right: 25, 50, 100, 150 nm length of the branched region). Here the sample is driven from outside of the pore to inside (left to right movement). Red traces show the current spikes due to branched regions of the sc-DNA sample. Schematics of possible DNA conformations corresponding to each event is shown on the top.

We quantify the branches in sc-DNA using ECD analysis of the translocation events. Fig1E shows the typical translocation events of sc-DNA through a 17 nm pore. The ECD of an event corresponds to the total amount of ions displaced by the molecule which depends on the size (length, in the case of a linear polymer) of the molecule and is independent of its conformation during translocation. [33] We use this analysis to quantify the linear lengths (L_b) and positions (Z_b) of the branches in sc-DNA, for all events that display the branch signatures. From the above analysis, we understand that sc-DNA in its natural form exists in both unbranched and branched structures with a wide range of sizes.

ENZYME DEPENDENT CHANGES IN PLASMID CONFORMATIONS:

Next, we perform quantification of NdeI dependent linearization of sc-DNA using our nanopore platform. The supercoiled plasmid was digested with NdeI enzyme for various durations of 0, 5, 10, 30 & 60 min and then the samples were analyzed by both gel electrophoresis and nanopore platform. We quantified the amount of lin-DNA in the samples with increasing reaction times to be 3.8%, 22.4%, 41.5%, and 55.2% respectively. The values of rate constants were found to be 0.13 ± 0.07 and 0.12 ± 0.02 min^{-1} from nanopore and gel data respectively which are in excellent agreement with each other. This work establishes the nanopore platform to study thermal and enzyme-dependent conformational changes in biomolecules at the single-molecule level.

MEASUREMENT OF DNA FOLDS USING NANOPORE PLATFORM

In the next chapter, we explore the thermally generated conformational folds of linear DNA. The bending ability of bare DNA alone has interesting consequences in gene regulation in both prokaryotes and eukaryotes and allows for its compaction in all organisms along with viruses. It was reported that the translocation frequency increases with an increase in the length of DNA [35]. Another interesting experiment studied the role of translocation direction in the translocation dynamics and it was observed that the molecule translocation is much slower when it moves out of the capillary. [36]

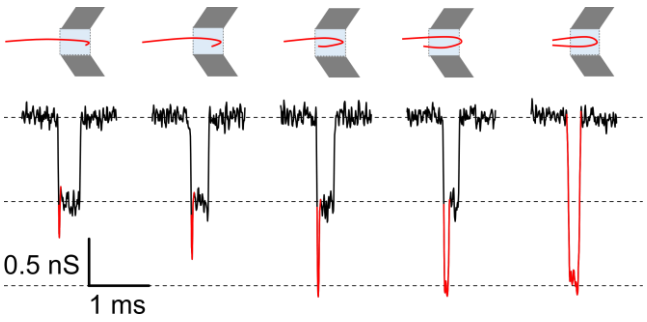


Figure 2: Nanopore analysis of linear 3kb DNA. (A) Representative events due to various lengths of DNA folding as recorded from a 19 nm pore (left to right: 25, 50, 100, 250, 500 nm length of the folded region). Schematics of possible DNA conformations corresponding to each event is shown on the top. The experiment was performed in 4M LiCl at 300 mV with the molecules translocating from the fluid well into the capillary (forward direction).

Here we perform a detailed analysis of folding lengths of linear DNA using the nanopore device (Fig2) to understand its dependence on the translocation direction. The linearized DNA exists in the form of a randomly coiled globule but becomes unfolded while passing through the nanopore. Due to the random orientation of the polymer as it reaches the pore

mouth, it can result in a single or multi-file translocation event. We have prepared 3kb, 5kb, and 10kb linearized DNA from their respective plasmids. We observe that the percentage of multi-file events is proportional to the length of DNA. The electrical signatures corresponding to linear DNA contain information about the internal geometry of the pore. This can be further used for the volumetric analysis of biomolecular structures.

NANOPORE ANALYSIS OF NUCLEOSOMES (1mer, 3-5mers & 12mers):

The chromatin has been studied using techniques like TEM [37], AFM [38], and fluorescence microscopy [39]. There are various factors that affect the chromatin structure like, the salt concentration, nucleosome interactions, and its interactions with enzymes. It was shown that chromatin exists in open form when the divalent ions are low in concentration and becomes folded or compact at a high concentration as shown by SAXS [40], ultracentrifugation [41], electron microscopy [42]. This thesis deals with the label and immobilization free study of nucleosome arrays using nanopore. The AFM images (Fig3A-D) show the bright spots in the case of 146 bp Nucl but a visible extra DNA can be seen around the 344 bp Nucl which is expected due to its long length. In the case of poly-nucleosomes, we can see the multiple nucleosomes on a single DNA molecule.

COMPARISON OF DNA, MONONUCLEOSOMES, AND 12mer ARRAY USING NANOPORE: We initially analyzed mononucleosomes on the nanopore and observed that they interact with the nanopore surface resulting in the long-tailed distribution in the dwell times. We also analyzed commercially available short nucleosome arrays which are clearly distinct from mononucleosome events and have higher conductance drops with longer event durations. We prepared 12-mer nucleosome arrays and characterized the sample by gel electrophoresis and AFM. The representative translocation events (Fig3E-G) are shown for DNA, mononucleosomes, and 12-mer arrays on the same pore which shows their structural differences clearly in the electrical signatures. We have written a custom LabVIEW code to quantify the number of levels in each event. On each pore we have linear DNA as a calibration sample, this allows us to calculate the molecular sizes of structures corresponding to each level.

OBSERVATION OF THE ACP PROTEIN INTERACTION WITH DNA AND NUCLEOSOMES

The genetic material inside each of the cell types is the same but their expression patterns are completely different and it is regulated by enzymes. One of the enzymes involved in silencing the genes by modification of chromatin is PRC2. The PRC2 enzyme is a complex of multiple proteins and is involved in gene silencing by spreading the methylation marks from one nucleosome to the other. There have been many reports on the binding of PRC2 binding to nucleosomes and bare DNA, mainly based on gel electrophoresis. The affinity of PRC2 to bare DNA has been shown to be much higher than nucleosome arrays. It is known to bridge neighbouring nucleosomes that may result in the compaction of chromatin [43]. A recent paper based on an optical tweezer experiment reported its interaction with modified and unmodified nucleosome arrays [44]. Due to the various possible binding modes of PRC2 onto the substrate, they result in a variety of folded structures in the case of both DNA and nucleosome arrays.

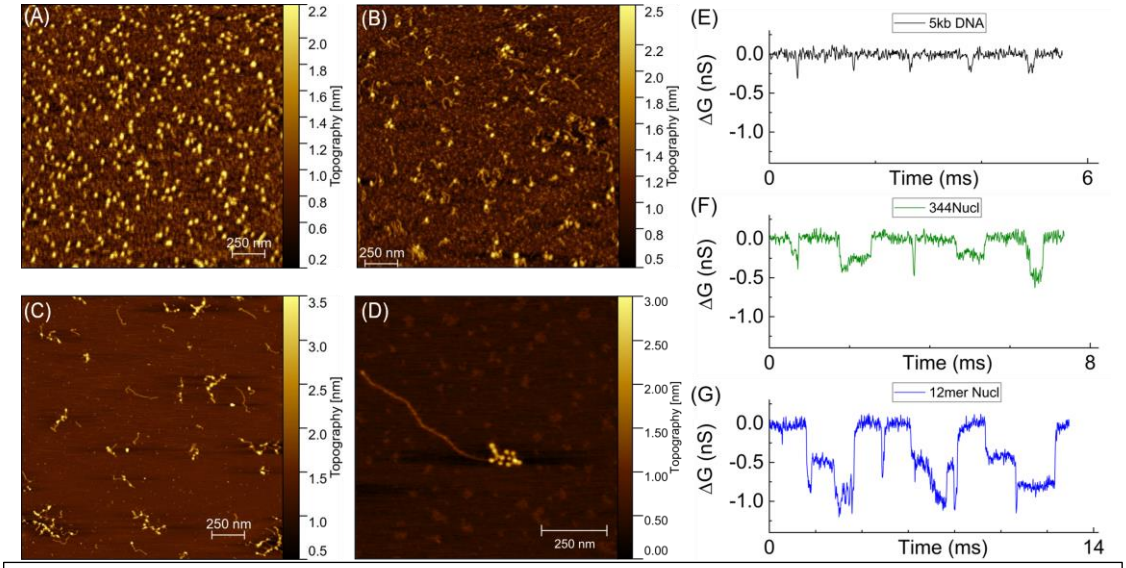


Figure3: AFM imaging and Nanopore analysis of various types of nucleosomes. (A) Mononucleosomes on 146 bp DNA, (B) Mononucleosomes on 344 bp DNA, (C) 3-5mer nucleosome arrays, (D) 12mer nucleosome arrays. (E)-(G) show the representative events for DNA, mononucleosomes and 12-mer nucleosome arrays, respectively. on a 22 nm pore at 500 mV in 0.5 M KCl.

The results presented in the previous chapters established that nanopore is a promising tool for quantitative analysis of conformational changes. In the final part of this thesis, we performed a detailed analysis of the PRC2 induced changes in conformations of DNA and nucleosome arrays. We performed titrations of DNA and nucleosome arrays with various concentrations of PRC2 and found the optimum conditions for their binding. From the nanopore experiments, we observe that PRC2 bound DNA produces greater ΔG and Δt values compared to bare DNA. The length of the linear DNA used here is 5kb.

The data clearly shows a very good contrast between PRC2 bound DNA and the bare DNA. The shape of the events corresponding to bare DNA is symmetric triangular whereas the PRC2 bound DNA are asymmetric. The possibility of this asymmetry stemming from the non-uniform distribution of PRC2 along the length of DNA is discussed and quantitative parameters are derived from these binding studies. Nanopore results along with gel electrophoresis and AFM analysis help us understand the structural details of these complexes.

SUMMARY:

In this thesis, we explored various aspects involved in the compaction of genetic material. In the first part, we quantified the branching of supercoiled DNA that helped us understand the native confirmations of bacterial plasmids. In the next part, we studied the dependence of folds in linear DNA as a function of translocation direction. The ECD analysis of the branched or folded conformations of DNA allowed us to estimate the internal geometrical parameters of the nanopore. Later the nanopores allowed us to quantitatively understand the folding of model chromatin system which are nucleosome arrays. Finally, we analyzed the binding of PRC2 to DNA and nucleosomes using various experimental techniques. The studies presented in the thesis provide us a better understanding of the structural features of DNA and chromatin.

REFERENCES (for Synopsis only):

[1] J. Vinograd, J. Lebowitz, R. Radloff, R. Watson, and P. Laipis, *The Twisted Circular Form of Polyoma Viral DNA.*, Proc. Natl. Acad. Sci. U. S. A. **53**, 1104 (1965).

[2] Andrew D. Bates and Anthony Maxwell, *DNA Topology* (Oxford University Press, 2005).

[3] Y. L. Lyubchenko and L. S. Shlyakhtenko, *Visualization of Supercoiled DNA with Atomic Force Microscopy in Situ*, Proc. Natl. Acad. Sci. U. S. A. **94**, 496 (1997).

[4] L. Loeff, S. J. J. Brouns, and C. Joo, *Repetitive DNA Reeling by the Cascade-Cas3 Complex in Nucleotide Unwinding Steps.*, Mol. Cell **70**, 385 (2018).

[5] C. L. F. Woodcock, J. P. Safer, and J. E. Stanchfield, *Structural Repeating Units in Chromatin. I. Evidence for Their General Occurrence*, Exp. Cell Res. **97**, 101 (1976).

[6] K. Luger and J. C. Hansen, *Nucleosome and Chromatin Fiber Dynamics*, Curr. Opin. Struct. Biol. **15**, 188 (2005).

[7] S. J. McBryant, V. H. Adams, and J. C. Hansen, *Chromatin Architectural Proteins*, Chromosom. Res. **14**, 39 (2006).

[8] P. R. Heenan, X. Wang, A. R. Gooding, T. R. Cech, and T. T. Perkins, *Bending and Looping of Long DNA by Polycomb Repressive Complex 2 Revealed by AFM Imaging in Liquid*, Nucleic Acids Res. **48**, 2969 (2020).

[9] S. Hyun Kim, M. Ganji, E. Kim, J. van der Torre, E. Abbondanzieri, and C. Dekker, *DNA Sequence Encodes the Position of DNA Supercoils*, Elife **7**, e36557 (2018).

[10] Y. Liu, V. Bondarenko, A. Ninfa, and V. M. Studitsky, *DNA Supercoiling Allows Enhancer Action over a Large Distance*, Proc. Natl. Acad. Sci. U. S. A. **98**, 14883 (2001).

[11] J. G. Brahms, S. Brahms, Y. Ohara, and V. Vagner, *Activation and Inhibition of Transcription by Supercoiling*, J. Mol. Biol **181**, 455 (1985).

[12] G. Snounou and A. D. B. Malcolm, *Supercoiling and the Mechanism of Restriction Endonucleases*, Eur. J. Biochem. **138**, 275 (1984).

[13] S. E. Halford, N. P. Johnson, and J. Grinsted, *The Reactions of the EcoRI and Other Restriction Endonucleases*, Biochem. J. **179**, 353 (1979).

[14] T. R. Strick, V. Croquette, and D. Bensimon, *Single-Molecule Analysis of DNA Uncoiling by a Type II Topoisomerase*, Nature **404**, 901 (2000).

[15] J. Gore, Z. Bryant, M. D. Stone, M. Nöllmann, N. R. Cozzarelli, and C. Bustamante, *Mechanochemical Analysis of DNA Gyrase Using Rotor Bead Tracking*, Nature **439**, 100 (2006).

[16] W. H. Coulter, *Means for Counting Particles Suspended in a Fluid. US Patent, 2656508*, (1953).

[17] A. J. Storm, C. Storm, J. Chen, H. Zandbergen, J. F. Joanny, and C. Dekker, *Fast DNA Translocation through a Solid-State Nanopore*, Nano Lett. **5**, 1193 (2005).

[18] J. Li, M. Gershow, D. Stein, E. Brandin, and J. A. Golovchenko, *DNA Molecules and Configurations in a Solid-State Nanopore Microscope*, Nat. Mater. **2**, 611 (2003).

[19] L. Liu, C. Yang, K. Zhao, J. Li, and H. C. Wu, *Ultrashort Single-Walled Carbon Nanotubes in a Lipid Bilayer as a New Nanopore Sensor*, Nat. Commun. **4**, 2989 (2013).

[20] G. F. Schneider, S. W. Kowalczyk, V. E. Calado, G. Pandraud, H. W. Zandbergen, L. M. K. Vandersypen, and C. Dekker, *DNA Translocation through Graphene Nanopores*, Nano Lett. **10**, 3163 (2010).

[21] L. J. Steinbock, R. D. Bulushev, S. Krishnan, C. Raillon, and A. Radenovic, *DNA Translocation through Low-Noise Glass Nanopores*, ACS Nano **7**, 11255 (2013).

[22] L. J. Steinbock, O. Otto, C. Chimerel, J. Gornall, and U. F. Keyser, *Detecting DNA Folding with Nanocapillaries*, Nano Lett. **10**, 2493 (2010).

[23] J. A. Bafna and G. V. Soni, *Fabrication of Low Noise Borosilicate Glass Nanopores for Single Molecule Sensing*, PLoS One **11**, e0157399 (2016).

[24] D. Branton et al., *The Potential and Challenges of Nanopore Sequencing*, Nature Biotechnology.

[25] M. Wanunu, *Nanopores: A Journey towards DNA Sequencing*, Physics of Life Reviews.

[26] G. V Soni and C. Dekker, *Detection of Nucleosomal Substructures Using Solid-State Nanopores*, Nano Lett. **12**, 3180 (2012).

[27] P. Waduge, R. Hu, P. Bandarkar, H. Yamazaki, B. Cressiot, Q. Zhao, P. C. Whitford, and M. Wanunu, *Nanopore-Based Measurements of Protein Size, Fluctuations, and Conformational Changes*, ACS Nano **11**, 5706 (2017).

[28] H. Chae, D. K. Kwak, M. K. Lee, S. W. Chi, and K. B. Kim, *Solid-State Nanopore Analysis on Conformation Change of P53TAD-MDM2 Fusion Protein Induced by Protein-Protein Interaction*, Nanoscale **10**, 17227 (2018).

[29] X. Wang, M. D. Wilkinson, X. Lin, R. Ren, K. R. Willison, A. P. Ivanov, J. Baum, and J. B. Edel, *Single-Molecule Nanopore Sensing of Actin Dynamics and Drug Binding*, Chem. Sci. **11**, 970 (2020).

[30] C. Plesa, D. Verschueren, S. Pud, J. Van Der Torre, J. W. Ruitenberg, M. J. Witteveen, M. P. Jonsson, A. Y. Grosberg, Y. Rabin, and C. Dekker, *Direct Observation of DNA Knots Using a Solid-State Nanopore*, Nat. Nanotechnol. **11**, 1093 (2016).

[31] R. Kumar Sharma, I. Agrawal, L. Dai, P. S. Doyle, and S. Garaj, *Complex DNA Knots Detected with a Nanopore Sensor*, Nat. Commun. **10**, 4473 (2019).

[32] F. Bošković, J. Zhu, K. Chen, and U. F. Keyser, *Monitoring G-Quadruplex Formation with DNA Carriers and Solid-State Nanopores,* Nano Lett. **19**, 7996 (2019).

[33] D. Fologea, E. Brandin, J. Uplinger, D. Branton, and J. Li, *DNA Conformation and Base Number Simultaneously Determined in a Nanopore*, Electrophoresis **28**, 3186 (2007).

[34] E. Krueger et al., *Modeling and Analysis of Intercalant Effects on Circular DNA Conformation*, ACS Nano **10**, 8910 (2016).

[35] N. A. W. Bell, M. Muthukumar, and U. F. Keyser, *Translocation Frequency of Double-Stranded DNA through a Solid-State Nanopore*, Phys. Rev. E **93**, (2016).

[36] N. A. W. Bell, K. Chen, S. Ghosal, M. Ricci, and U. F. Keyser, *Asymmetric Dynamics of DNA Entering and Exiting a Strongly Confining Nanopore*, Nat. Commun. **8**, (2017).

[37] J. T. Finch, M. Noll, and R. D. Kornberg, *Electron Microscopy of Defined Lengths of Chromatin*, Proc. Natl. Acad. Sci. U. S. A. **72**, 3320 (1975).

[38] K. M. Krzemien, M. Beckers, S. Quack, and J. Michaelis, *Atomic Force Microscopy of Chromatin Arrays Reveal Non-Monotonic Salt Dependence of Array Compaction in Solution*, PLoS One **12**, 1 (2017).

[39] A. K. Estandarte, S. Botchway, C. Lynch, M. Yusuf, and I. Robinson, *The Use of DAPI Fluorescence Lifetime Imaging for Investigating Chromatin Condensation in Human Chromosomes*, Sci. Rep. **6**, 1 (2016).

[40] S. Mangenot, A. Leforestier, P. Vachette, D. Durand, and F. Livolant, *Salt-*

Induced Conformation and Interaction Changes of Nucleosome Core Particles, Biophys. J. **82**, 345 (2002).

[41] N. Korolev, A. Allahverdi, Y. Yang, Y. Fan, A. P. Lyubartsev, and L. Nordenskiöld, *Electrostatic Origin of Salt-Induced Nucleosome Array Compaction*, Biophys. J. **99**, 1896 (2010).

[42] S. A. Grigoryev, G. Arya, S. Correll, C. L. Woodcock, and T. Schlick, *Evidence for Heteromorphic Chromatin Fibers from Analysis of Nucleosome Interactions*, Proc. Natl. Acad. Sci. **106**, 13317 (2009).

[43] S. Poepsel, V. Kasinath, and E. Nogales, *Cryo-EM Structures of PRC2 Simultaneously Engaged with Two Functionally Distinct Nucleosomes*, Nat. Struct. Mol. Biol. **25**, 154 (2018).

[44] R. Leicher, E. J. Ge, X. Lin, M. J. Reynolds, W. Xie, T. Walz, B. Zhang, T. W. Muir, and S. Liu, *Single-Molecule and in Silico Dissection of the Interaction between Polycomb Repressive Complex 2 and Chromatin*, Proc. Natl. Acad. Sci. U. S. A. **117**, 30465 (2020).

Synopsis submitted: 30-04-2021

Name of the Student: MAHESHWARAM SUMANTH KUMAR

M. Sumanth

Signature of the Student

Name of the Supervisor: Dr. GAUTAM V. SONI

Signature of the Supervisor