ABSTRACT

MING LIU. Protein-induced DNA Conformation Changes under Nanoconfinement (Under the direction of Dr. Robert Riehn and Dr. Hong Wang).

Two essential components of a living cell are DNA and proteins. DNA carries genetic information, and proteins are the central functional units for various cell activities. Protein-DNA interactions are required for DNA to fulfill its function. This thesis investigates protein-DNA interactions, especially the protein-induced DNA conformation changes, using a nanofluidic system. In the scope of this thesis, we focused on several proteins from different protein systems, namely the MBD family, the shelterin complex, and the cohesin complex. Combined with Atomic Force Microscope (AFM) for higher spatial resolution, we demonstrated a distinct difference in affecting DNA conformation between two proteins from the MBD family. We found that ATTO 565-labeled MBD2 is a good candidate as a staining agent for epigenetic mapping. We also investigated a ternary protein system with TRF1, TIN2 from the shelterin complex, and SA1 from the cohesin complex. We find that TRF1, TIN2, and SA1 work synergistically to cause compaction of the DNA substrate. By altering the exposure sequence, we suggest that all three proteins can be understood as an additive effect of TRF1-TIN2 and SA1-based compaction that proceed from independent DNA binding.

Protein-induced DNA Conformation Changes under Nanoconfinement

by
Ming Liu

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Physics

Raleigh, North Carolina
2022

APPROVED BY:

Robert Riehn Committee Chair	Hong Wang
Mary Elting	Albert Keung

DEDICATION

To my dear parents

BIOGRAPHY

I was born in a remote county in Xinyu, Jiangxi. I attended Xinyu No.4 Middle School from 2005 to 2011. After the National College Entrance Examination, I attended the University of Science and Technology of China and obtained a bachelor's degree in applied physics in 2015. I started my Ph.D. program at North Carolina State University in 2016 and obtained the Ph.D. degree in 2022.

ACKNOWLEDGMENTS

First, I want to express my greatest gratitude to my advisor Dr. Robert Riehn and my co-advisor, Dr. Hong Wang, for all their support and guidance during my Ph.D. program. I thank them for sending their help in need in various aspects of my life. I thank them for their guidance in helping me out when I was struggling and confused during these years. Their broad knowledge and insightful, detailed support for research enabled me to explore my interest. Without their help, I could not get through the Ph.D. program.

I would like to thank Dr. Mary Elting and Dr. Albert Keung for taking the time to be my dissertation committee. I thank them for their helpful comments on my research work. I am also grateful to the people in the physics department. I thank Dr. Keith Weninger for attending my prelim presentation and giving helpful comments during these years. I thank Rhonda for her help in graduation affairs.

I must thank my lab mates for their invaluable help in learning research skills. I thank Dr. Hai Pan for his help in various biochemistry skills. I thank Dr. Saroj Dangi for his help in nanofabrication, and thank Dr. Parminder Kaur for her help in AFM. I also thank my lab mates Abbas and Kevin for their encouragement.

Last but never the least, I am forever grateful to my dear parents for all their sacrifices and all they gave to me. I would not be able to make any achievements without them. I also thank my friends for their encouragement and support throughout the whole Ph.D. program and make it an wonderful journey in my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 INTRODUCTION TO PROTEIN-DNA INTERACTION

In 1957, Francis Crick stated a Central Dogma for biology (Crick 1970), which interprets the information transfer pathway from nucleotide sequences to amino acid sequences. The gene expression pathway, which is from DNA to mRNA and mRNA to protein, is also referred to as gene transcription and translation, respectively. However, genes are not expressed all the time without limitations in live cells. This regulation of gene expression is usually controlled by proteins or RNA, which results in the suppression or promotion of gene expression. Significantly during gene transcription, the DNA molecule can be modified by proteins to alter gene expression. A well-known example is the lac operon which regulates gene expression by looping DNA (Razo-Mejia et al. 2014). This relation between gene expression and protein-DNA interactions points to the importance of studying the interaction between DNA and proteins.

Despite the critical role in gene expression, protein-DNA interactions are also involved in other phenomena, such as chromosome structure, DNA looping, and chromatin cohesion. The binding of the protein to DNA can be specific or nonspecific. There are nonspecific protein-DNA complexes like the formation of nucleosomes, which are DNA wrapped around a small histone complex (Luger et al. 1997). Some proteins contain DNA binding domains, which can specifically bind to DNA structures or recognize specific DNA sequences. For instance, the MBD family, mentioned in later content, can specifically bind onto methylated CpG sites and act as a gene regulator (Mahmood and Rabbani 2019). The binding of these proteins usually causes psychological or physical changes. For instance, restriction enzymes (endonucleases) can target specific DNA sequences and cleave them (Smith 1979). The MeCP2 from the MBD family can

loop DNA molecules and cause compaction. By studying the protein-induced DNA conformation changes, we can better understand the influence of proteins on DNA.

Researchers have developed various methods to investigate protein-DNA interactions. We will mention their details and principles in the next chapter. Due to the complexity of protein-DNA interactions, several methods may be involved simultaneously to understand the mechanism entirely. The binding of proteins to DNA can cause conformation changes to DNA molecules, like the histone, which requires understanding the physics of the interaction. This thesis will mainly discuss the methods that we have used for our protein systems. This chapter will first focus on the physics and biology background for DNA-protein interaction.

1.1 Physics properties of DNA and worm-like chain model

To understand the physics of DNA-protein interaction, we start with the DNA molecule. In 1953, Francis Crick and James Watson first proposed the double helix structure of DNA based on X-ray diffraction (Watson and Crick 1953). It contains two strands of a polynucleotide, which are connected by the hydrogen bonds of four bases, Adenine (A), Thymine (T), Guanine (G), and Cytosine (C). Its phosphate backbone makes the DNA appear negatively charged inside water. The DNA double helix diameter is around 2 nm, and the distance between the base pairs is around 0.34 nm. However, due to the thermal dynamics, large DNA molecules are not forming a specific shape in the solution. Despite the double helix structure of DNA, researchers usually simplify it as a negatively charged polymer chain when investigating DNA conformation changes, like looping and wrapping, which usually happen on a much larger scale. Depending on whether the elasticity of DNA is taken into account, DNA can be simplified into two models,

random walk and worm-like chain. Textbooks summarize these properties well, and we will follow them in the next three subsections (Strobl 2007; Phillips et al. 2012).

1.1.1 Random Walk

This model considers the DNA molecule as a freely jointed chain, also called an ideal chain. DNA is simplified as a series of rods connected by hinges. The length of the rod is the Kuhn length l_k, which is around 300 bp for DNA molecules. Then obviously, we will have:

$$L = Nl_k \tag{1}$$

Here, L is the contour length of the DNA molecule, and the N is the number of segments. To get the size of the DNA molecule, we start with the end-to-end vector:

$$\vec{R} = \sum_{i=1}^{N} \overrightarrow{x_i} \tag{2}$$

Here the $\overrightarrow{x_i}$ is the segment vector. First, we ignore the interaction between DNA segments, which means that the rods are not excluding each other. Then the segments can walk in any direction. However, because of the thermal fluctuation of the DNA molecule, $\vec{R}$ is changing all the time, which means a statistical result is more valuable. The mean and variance of the $\vec{R}$:

$$\langle \vec{R} \rangle = \left\langle \sum_{i=1}^{N} \overrightarrow{x_i} \right\rangle = \sum_{i=1}^{N} \langle \overrightarrow{x_i} \rangle \ = 0 \tag{3}$$

$$\langle \vec{R}^2 \rangle = \left\langle \sum_{i=1}^{N} \sum_{j=1}^{N} \overrightarrow{x_i} \overrightarrow{x_j} \right\rangle = \sum_{i=1}^{N} \langle \overrightarrow{x_i}^2 \rangle + \left\langle \sum_{i=1}^{N} \sum_{j \neq i}^{N} \overrightarrow{x_i} \overrightarrow{x_j} \right\rangle = N {l_k}^2 \tag{4}$$

The end-to-end distance:

$$\sqrt{\langle R^2 \rangle} = \sqrt{N} l_k \tag{5}$$

is one way to estimate the DNA size, which implies that the size of DNA increases proportional to $\sqrt{N}$. In a three-dimensional random walk, the distribution function of the end-to-end distance of DNA is:

$$f(\vec{R}, N) = \left(\frac{3}{2\pi N l_k^2}\right)^{\frac{3}{2}} \exp\left(-\frac{3R^2}{2N l_k^2}\right) \tag{6}$$

Another way to describe the size of DNA is the radius of gyration:

$$R_g^2 = \frac{1}{N}\sum_{i=1}^{N} \langle |\vec{r_i} - \vec{r_c}|^2 \rangle \tag{7}$$

Here the $\vec{r_c}$is the position of the center. As for the ideal chain, the relation between the radius of gyration and the end-to-end distance is:

$$\langle R_g^2 \rangle = \frac{\langle R^2 \rangle}{6} \tag{8}$$

1.1.2 Worm-like Chain

If we consider DNA's elasticity, then the DNA can be treated as a beam that can be bent or stretched. Due to the thermal fluctuation, the radius of curvature $R(s)$ along the DNA depends on its position s. Then, the local bending energy should be:

$$E_{local} = \frac{E}{2}\int_{\partial\Omega} \frac{z^2}{R(s)^2}\, dA \tag{9}$$

Here the E is Young's modulus, z is the distance to the neutral axis where there is no extension, and the $\partial\Omega$ is the cross-section area. If we define the geometric momentum as:

$$I = \int_{\partial\Omega} z^2 dA \tag{10}$$

Then:

$$E_{local} = \frac{EI}{2R(s)^2} \tag{11}$$

The bending energy is the integral of local bending energy over the whole molecule:

$$E_{bend} = \frac{EI}{2}\int_0^L \frac{1}{R(s)^2}\, ds \tag{12}$$

Although the bending energy increases with the DNA length, the entropy is also increasing. This competition between the energy and the entropy creates a threshold length when the DNA is roughly rigid, which is the persistence length. The definition of persistence length is the decay rate of the tangent correlation function, i.e.

$$\langle \vec{t}(s) \cdot \vec{t}(0) \rangle = \exp\left(-\frac{s}{l_p}\right) \tag{13}$$

Here the $\vec{t}$ is the unit tangent vector along the DNA molecule and l_p is the persistence length. To get the persistence in terms of DNA physical properties, we can think about a short piece of DNA, in which case, $s << l_p$. In this case, we can treat $R(s)$ as a constant R along the length s. Due to the thermal fluctuation, this short DNA can bend in various directions, and the probability follows Boltzmann distribution, i.e.:

$$Z = \int_0^{2\pi}\int_0^{\pi} d\theta\, sin\theta \exp\left(-\frac{E_{bend}(\theta)}{k_B T}\right)\, d\varphi = \int_0^{2\pi}\int_0^{\pi} d\theta sin\theta \exp\left(-\frac{EI}{2k_B T s}\theta^2\right) d\varphi \tag{14}$$

Here, the Z is the partition function, and the bending angle $\theta = s/R$. Also, for $\theta << 1$:

$$\langle \vec{t}(s) \cdot \vec{t}(0) \rangle = \exp\left(-\frac{s}{l_p}\right) = \langle \cos(\theta) \rangle = \left\langle 1 - \frac{\theta^2}{2} \right\rangle \tag{15}$$

Use the partition function to get the average:

$$\left\langle 1 - \frac{\theta^2}{2} \right\rangle = 1 - \frac{1}{2Z}\int_0^{2\pi}\int_0^{\pi} d\theta\theta^2 sin\theta \exp\left(-\frac{EI}{2k_B T s}\theta^2\right) d\varphi = 1 - \frac{k_B T}{EI}s = \exp\left(-\frac{s}{l_p}\right) \tag{16}$$

Since $s << l_p$, the persistence length is:

$$l_p = \frac{EI}{k_B T} \tag{17}$$

Also,

$$E_{loop} = \frac{l_p k_B T}{2} \int_0^L \frac{1}{R(s)^2}\, ds \tag{18}$$

For DNA molecules, the persistence length is around 50 nm. It is half of the Kuhn length l_k in the random walk section.

Another thing showing the competition between the entropy and the energy is the cyclization of DNA molecules. The main problem of cyclization for long DNA molecules is for the DNA ends to find each other. In this case, the entropy rules because the microstates of DNA ends get close is not a lot. In other words, the end-to-end distance $R < \delta$. The probability of getting a small end-to-end distance:

$$p = \int_0^{\delta} 4\pi R^2 f(\vec{R}, N) dR = \int_0^{\delta} 4\pi R^2 \left(\frac{3}{2\pi N l_k^2}\right)^{\frac{3}{2}} \exp\left(-\frac{3R^2}{2N{l_k}^2}\right) dR \tag{19}$$

Here we use the random walk model for long-distance. For cyclization, we have $\delta \ll l_k$, so the probability is approximated as:

$$p \cong \left(\frac{6}{\pi N^3}\right)^{\frac{1}{2}} \left(\frac{\delta}{l_k}\right) \tag{20}$$

We can see that $p \propto N^{-\frac{3}{2}}$. Now, if we combine entropy and the bending energy (18), the free energy of getting loops is:

$$\Delta G = \Delta E_{loop} - k_B T ln p \cong k_B T \left(\frac{2\pi^2 l_p}{N l_k} + \frac{3}{2} ln N + \text{Const}\right) \tag{21}$$

Here we simplified the problem by considering the DNA is forming one single loop with a radius R. We can see that the free energy first decreases and then increase when N increase, which implies the existence of N that can give minimum free energy. Researchers used the ligase to ligate the complementary DNA ends (Shore and Baldwin 1983). The J-factor is defined as:

$$J = \frac{K_c}{K_a} \tag{22}$$

It is a measurement of DNA cyclization. Here, the K_c is the equilibrium constant for cyclization and K_a is the equilibrium constant for the association of two DNA molecules with complementary DNA ends. The experiment results showed a maximum peak for J-factor at around 500 bp DNA length, indicating that 500 bp DNA has the highest looping probability (Shore and Baldwin 1983).

1.1.3 Excluded Volume

In previous models, we did not include the interactions between the monomer 'rods' of the DNA molecule. It is feasible in most cases when DNA extends in solutions. However, it was obvious that different monomers could not show up in the same place in a real chain, and this effect plays an essential role under nanoconfinement for DNA extension. More about excluded volume in a nanofluidic device will be covered in a later chapter. Here we are going to give some background. The excluded volume interaction can result in the expansion of DNA molecules, also called expanded or self-avoiding chains. However, the entropy rules in large molecules, preventing the chain from expanding. Based on the model built by Flory (de Gennes and Witten 1980; Reisner, Pedersen, and Austin 2012) in three dimensions, the self-exclusion energy is proportional to the square of monomer concentration $c = \frac{N}{R^3}$, and can be approximated as:

$$F_{ex} \cong \frac{1}{2} k_B T \chi c^2 R^3 = \frac{\frac{1}{2} k_B T \chi N^2}{R^3} \quad (23)$$

Here the excluded volume $\chi = \omega l_k^2$ (ω, l_k are the width and length of monomer). The entropy from the random walk (19) can be approximated as:

$$T\Delta S \cong \frac{3 k_B T R^2}{2 l_k^2 N} + const \quad (24)$$

The total free energy is the combination of entropy and exclusion energy. The minimum of the free energy can be achieved when the effects of entropy and exclusion are comparable. Details will be in the later chapter. Here we give the result. The radius of the molecule at minimum free energy is:

$$R_F \cong (\omega l_p)^{\frac{1}{5}} L^{\frac{3}{5}} \quad (25)$$

Here, R_F is the Flory radius of the molecule, L is the contour length of the molecule, and l_p is the persistence length and ω is the effective width. The self-exclusion effect causes a change in the scaling law of the end-to-end distance, which is:

$$\sqrt{\langle R^2 \rangle} = R_F \propto N^{\frac{3}{5}} \quad (26)$$

It is different from the ideal chain result $\sqrt{\langle R^2 \rangle} \propto N^{1/2}$. Here the N is the number of segments the $\vec{R}$ is the end-to-end vector based on the self-exclusion model. The importance of the self-exclusion effect depends on the effective width of the monomer ω and the total length of the polymer L (Reisner, Pedersen, and Austin 2012). The excluded volume starts to matter when the free energy of excluded volume is larger than the magnitude of $1 k_B T$, that is:

$$F_{ex} \cong \frac{\frac{1}{2} k_B T \chi N^2}{R_F^3} \sim k_B T \rightarrow L_{**} = \frac{\alpha_{**} l_p^3}{\omega^2} \quad (27)$$

Here the L_{**} is the threshold length when the excluded volume starts to matter. The α_{**} is a numerical factor. Experimental results show the $L_{**}\sim 10kbp$ (Valle et al. 2005). A more general scaling for dimension d is:

$$\sqrt{\langle R^2 \rangle} \propto N^v, v = \frac{3}{2+d} \tag{28}$$

The v is called the Flory exponent, and for $d = 4$ the chain is ideal.

1.2 Structure and mechanism of DNA binding domains

We have discussed the physical properties of DNA molecules in the last section. DNA molecules alone can form some specific shapes, like supercoils, because of the interplay of entropy and elastic energy. However, in the presence of proteins, the conformation of DNA can be quite different from bare DNA. A famous example is a histone (Luger et al. 1997) and the lac operon. In this case, understanding the mechanism of DNA binding and the structure of the DNA binding domain is essential to understanding DNA-protein interaction.

1.2.1 Structure of DNA binding domain

Knowledge of the protein structure is essential to understanding the binding between a protein and DNA. The protein structure contains four levels. The primary structure of a protein is the sequence of the amino acid chain. There are 20 kinds of amino acids, which are coded by the DNA sequence. Due to the different properties of the side chains, these 20 amino acids sort into five kinds: nonpolar hydrophobic, polar uncharged, polar basic, polar acidic, and exceptional cases. Hydrophobic amino acids are more likely buried inside the protein, while hydrophilic amino acids are more likely to be found on the protein surface. The secondary structure is the small, structured fragment mainly consisting of α-helix and β-sheet. The structure can also be

missing in some regions, like intrinsically disordered regions. Hydrogen bonds between the amino acids create these two structures. The secondary structures are linked into a tertiary structure, which is the large-scale 3D structure of a protein. The fourth level of structure are functional protein complexes (Stormo 2013).

The interaction mechanism between DNA and protein can be divided into two kinds, specific or nonspecific. The specific interaction requires the recognition of a unique sequence or special shapes of DNA molecules. The specific structure of DNA binding domains helps proteins target the specific sequence or unique shapes. Depending on the secondary structure in the DNA binding domain, DNA binding proteins are divided into four main kinds: mainly α, mainly β, mixed α/β, and multi-domain proteins (Rohs et al. 2010). I will focus on the structure in the following content based on two reviews (Rohs et al. 2010; Stormo 2013).

- **Mainly α**

Mainly α refers to those protein domains that interact with DNA mainly through α-helix. α-helix is the most frequently used structure for DNA sequence recognition. One example is the helix-turn-helix motif (HTH). This motif contains one helix binding on the major groove and specifically interacts with the DNA sequence. The other helix binds the backbone, which can stabilize the binding. One famous example is the Cro repressor(Anderson et al. 1981), which symmetrically binds onto DNA as a dimer and has an HTH motif to target the DNA specifically, as shown in Fig. 1.1 A. The binding of HTH and DNA can vary for different proteins. Some of the variants may have an additional β-sheet, which is also called ‘winged HTH’. The additional β-sheet creates additional DNA binding in its minor groove. In some cases, the helix interacts

with the minor groove while the β-sheet interacts with the major groove, like RFX1 (Gajiwala et al. 2000). Helix-loop-helix and leucine zipper motifs can also be sorted into mainly α. Both of them are characterized by a long α-helix that can specifically bind with DNA in the major groove. The helix-loop-helix motif has a long α-helix and a short α-helix connected by a loop. The leucine zipper (Fig. 1.1 B) is named by its periodic Leu repeats on the motif. Both of them usually dimerized and worked as a coiled-coil structure for functioning.

- **Mainly $\boldsymbol{\beta}$**

Although most of the specific recognitions are based on the interaction between the α-helix and DNA major groove, in some protein domains, the β-sheet can also target the specific sequence. An important example is the TATA box-binding protein (TBP) (J. L. Kim, Nikolov, and Burley 1993), as shown in Fig. 1.1 C. Its β-sheet interacts with the minor groove on the bended DNA.

- **Mixed $\boldsymbol{\alpha/\beta}$**

Some DNA binding domains may have both the α-helix and β-sheet involved. One example are the zinc finger proteins (Klug 2010), as shown in Fig. 1.1 D. The zinc finger domain consists of one α-helix and a paired β-sheet, with a zinc ion coordinated by cysteine and histidine between to stabilize the structure. The α-helix binds to the major groove and creates a specific recognition of DNA sequences. Zinc finger proteins can have several zinc fingers for their specific recognition. Another example is the ribbon-helix-helix (RHH) motif (Fig. 1.1 E) which has a paired β-sheet and two α-helix. The specific recognition is based on the interaction between the major groove and the β-sheet of the RHH motif.

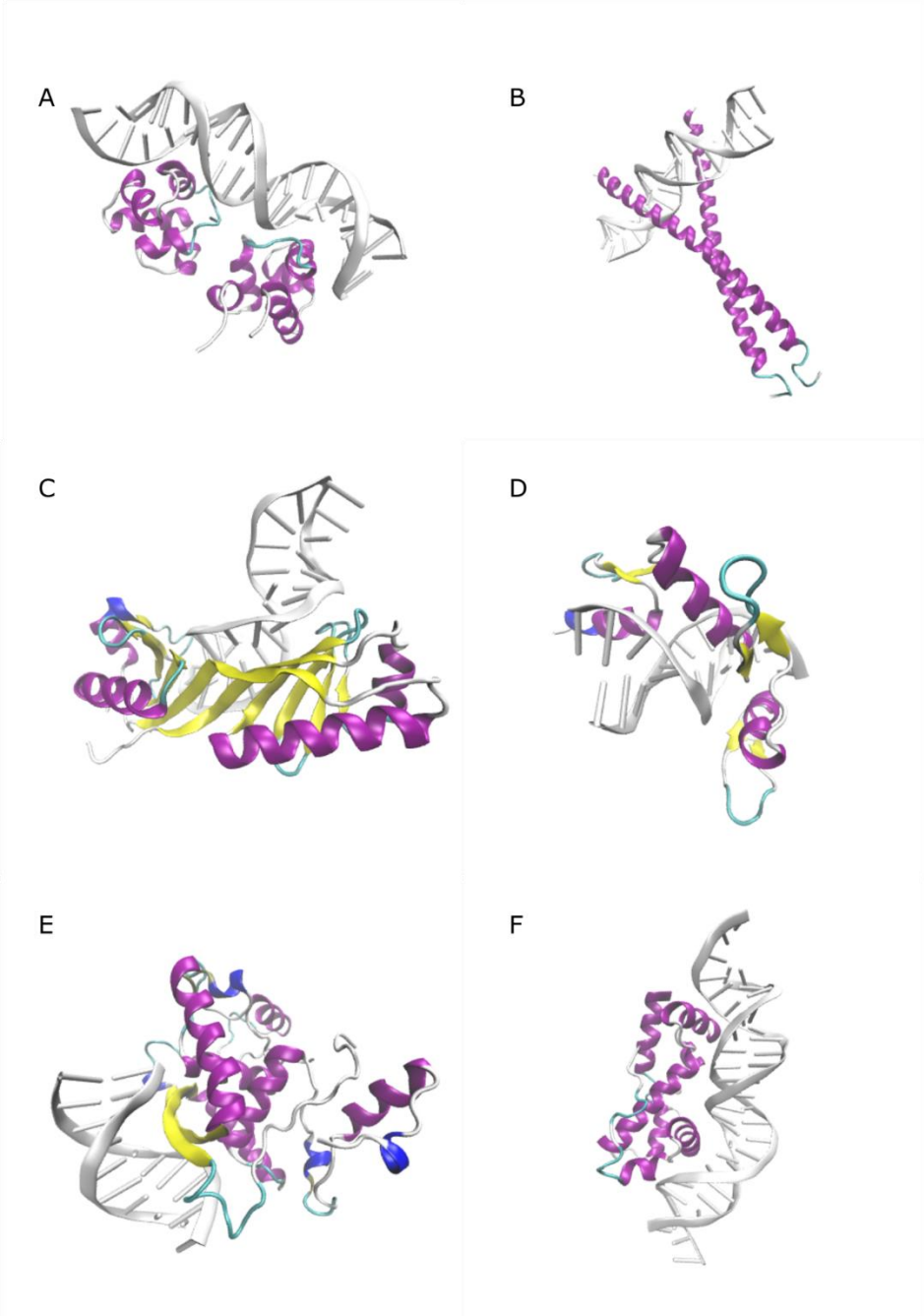

Figure 1.1 Structure of different DNA binding domains interacting with DNA. (A) HTH motif of Cro repressor (PDB 3CRO). (B) leucine zipper structure of Gcn4 (PDB 1DGC). (C) ***β***-sheet domain of TBP (PDB 1YTB). (D) zinc finger of zif268 (PDB 1AAY). (E) RHH motif of MetJ (PDB 1CMA). (F) two HTH of MarA (PDB 1BL0). Figures were created by VMD, and the source files were from Protein Data Bank.

- **Multi-domain proteins**

We have already discussed different motifs or domains and their interaction with the DNA molecule. However, some proteins have several domains that target different places along the DNA molecule. Like the MarA (Rhee et al. 1998), it contains two HTH motifs that can bind specifically to two significant grooves.

1.2.2 Mechanism of DNA binding

DNA recognition can divide into two groups, the recognition of specific DNA sequences and the recognition of specific DNA shapes. The recognition of a specific sequence depends on the hydrogen bond interactions or hydrophilic interactions between the proteins and bases. Different bases in the major groove show different hydrogen bond donors and acceptors patterns. The specific recognition comes from forming a hydrogen bond between the side chain of amino acids and the bases. This formation of hydrogen bonds can be directly connected or mediated by water molecules in between. Furthermore, hydrophobic interactions are often involved. Hydrophobic interactions mainly help in distinguishing cytosine (C) and thymine (T). It is also essential in binding single-strand DNA because of the exposure of the hydrophilic residues.

This recognition of bases also happens in the minor groove. However, hydrogen bond donors and acceptor patterns are indistinguishable in the minor groove between AT and TA or CG and GC, as shown in Fig 1.2. Thus, there is an insufficient basis for specific recognition, and other recognition mechanisms, like shape recognition, are needed for achieving specificity. The hydrophobic interaction in the minor groove may involve a change in DNA shapes, like TBP we mentioned before.

However, the specificity usually depends not only on the recognition of bases. For specific recognition, the recognition of DNA shapes is an alternative way. Local DNA shapes refer to a local structural change, including minor groove narrowing and DNA kinks (Rohs et al. 2010). Global DNA shapes refer to changes on a larger scale, like DNA bending (Starr, Hoopes, and Hawley 1995). The difference in shape can also be sequence-dependent because the energy costs of bending are generally sequence-dependent. Although the DNA we usually refer to is B-DNA, there are also A-DNA and Z-DNA, which have different double helix structures. Proteins also target these different DNA structures for recognition.

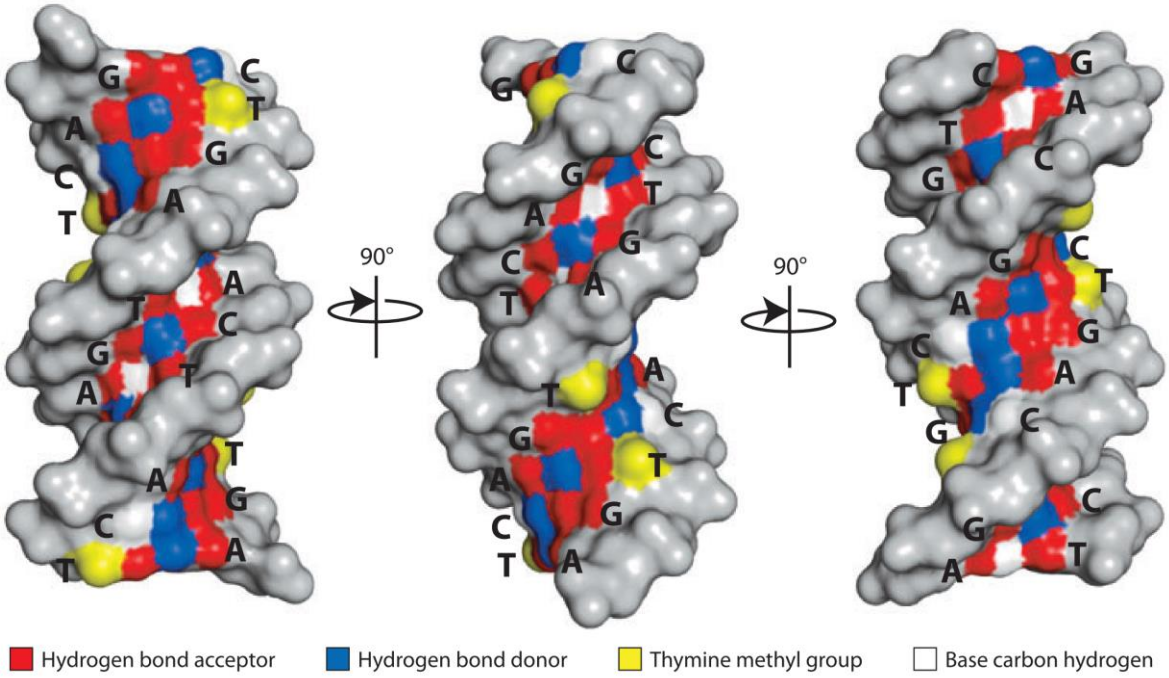

Figure 1.2 Pattern of hydrogen bond acceptors and donors on DNA. The figure was taken from (Rohs et al. 2010). The patterns of donors and acceptors of AT and TA or CG and GC are distinguishable in major grooves but indistinguishable in minor grooves.

1.2.3 Cooperativity and allostery

Protein binding affinity can sometimes be influenced by the proteins already bound to the DNA. The cooperativity of the protein binding causes this phenomenon. Unlike the cases in which the protein acts independently, the protein binding affinity can increase or decrease in the presence of bound proteins. This positive or negative correlation is called positive cooperativity or negative cooperativity. A famous model is the cooperativity behavior of oxygen binding to the hemoglobin (Bohr, Hasselbalch, and Krogh 1904). The hemoglobin can bind four oxygen molecules, and the binding is dependent.

When simple ligand-receptor model, the reaction can be described as:

$$L + R \rightleftharpoons LR \tag{29}$$

Furthermore, the dissociation constant is expressed as:

$$K_d = \frac{[L][R]}{[LR]} \tag{30}$$

In the case when receptors have several binding sites, the dissociation constant is:

$$K_d = \frac{[L]^n[R]}{[L^nR]} \tag{31}$$

The probability of receptors that are bonded by ligands, which is the Hill function (Phillips et al. 2012):

$$p = \frac{[L^nR]}{[R] + [L^nR]} = \frac{[L]^n}{K_d + [L]^n} \tag{32}$$

Here, n is the Hill coefficient. We usually fit the data in the experiment to get the Hill coefficient. However, the Hill coefficient is usually not an accurate estimation of the number of receptor binding sites. It is an estimation of how many ligands effectively interact with a receptor.

The Monod-Wyman-Changeus (MWC) model is a distinct classical model for cooperative binding (Monod, Wyman, and Changeux 1965), which induced a concept of allostery in biology. An essential part of this model is that the protein has two states, tense (T) and relax (R). Each of the two states can change to the other state, and the ratio depends on each state's energy. However, if the binding of ligands prefers the state with higher energy, the protein has a higher probability of staying in the high energy state with the increasing concentration of ligands. There are also some other models of cooperativity. The Pauling model is based on the assumption that there is a paired interaction between the ligands and Adair's model also includes the interaction between three or more (Adair, Bock, and Field 1925; Pauling 1935).

With the development of the field in protein structure, researchers have realized that protein structure is dynamic instead of static. Inspired by thermal dynamics, allostery was described as conformation selection by changing the probability distribution and is an intrinsic property of the protein (J. Liu and Nussinov 2016). Methods of investigating allostery include nuclear magnetic resonance (NMR) (Grutsch, Brüschweiler, and Tollinger 2016) and fluorescence resonance energy transfer (FRET) (Ramanoudjame et al. 2006).

1.3 Dynamics of DNA and protein binding

Molecules like proteins diffuse around inside solutions due to collision with water molecules. This is called Brownian motion and follows:

$$\langle x \rangle^2 = 2Dt \tag{33}$$

D is the diffusion constant. The diffusion rate depends on the solution's viscosity and the particle's shape. By simplifying the molecules into a sphere, we can estimate the diffusion constant:

$$D = \frac{k_B T}{6\pi\eta r} \tag{34}$$

This formula is also known as Einstein's Formula. Here the r is the radius of the protein. For water or buffer, the viscosity $\eta \sim 10^{-3}\ Pa \cdot s$, and the radius of a typical DNA-binding protein is several nanometers. Thus, we can obtain a diffusion constant $D \sim 10^2 \mu m^2/s$. The association rate per unit concentration for proteins binding onto a specific target DNA site through 3D diffusion was evaluated (Halford 2004):

$$k = 4\pi D a \tag{35}$$

Here a is the size of target sites, which is usually smaller than the size of protein r. This rate is around $10^8/(M \cdot s)$. This estimation assumes the protein finds the target only through 3D diffusion searching.

1.3.1 Facilitated diffusion

However, researchers have noticed that there are cases when the association rate of protein to its target sites on DNA is larger than the 3D diffusion limits, like in the case of the Lac repressor (Riggs, Bourgeois, and Cohn 1970). This experimental result implies another mechanism involved in protein binding, named facilitated diffusion (Halford 2004). The critical point of the facilitated diffusion is that nonspecific binding between the protein and DNA is involved, unlike the searching based only on 3D diffusion. The proteins can first nonspecifically bind to DNA and then search the target sites by 1D diffusion along with the DNA. In this way, the nonspecific binding dramatically increases target size, from recognition sites to a long DNA fragment, and

the search space decreases. The probability and rate of binding then increase. From 1D diffusion along the DNA, protein can disassociate from the DNA randomly. The average length that the protein slides along the DNA before disassociation is the sliding length l_{sl}. If we set the probability of dissociation of a protein at a binding site is p ($p \ll 1$), the sliding length can be estimated:

$$l_{sl} = \frac{h}{\sqrt{p}} \tag{36}$$

Here the h is the size of one step along with DNA, which is a base pair $h = 0.34\ nm$. Before finding the target sites, the protein will go through a series of 3D diffusion 'hopping' and 1D diffusion 'sliding'.

1.3.2 Curvilinear versus linear/hopping diffusion

When the protein is attached to DNA through nonspecific binding and sliding on the DNA to search for the target sites, one important question is whether the protein rotates when sliding. Researchers have discussed the curvilinear movement of the protein (Bagchi, Blainey, and Xie 2008). The normal translational friction is:

$$\zeta = 6\pi\eta r \tag{37}$$

After taking the rotation of the protein into account, the total friction is (Bagchi, Blainey, and Xie 2008):

$$\zeta_{total} = 6\pi\eta r + \left(\frac{2\pi}{10bp}\right)^2 (8\pi\eta r^3 + 6\pi\eta r(r_{OC})^2) \tag{38}$$

Here the r is the radius of protein and r_{OC} is the center-to-center distance between the protein and the DNA. This modification gives a closer result to the single-molecule experimental result

while still containing a large gap. The difference can be caused by the detailed interaction between the protein and DNA (Bagchi, Blainey, and Xie 2008).

CHAPTER 2 INTRODUCTION TO BIOPHYSICAL MEASUREMENT TECHNIQUES

Research relies on the use of different instruments. Here we discuss the instruments we mainly used for our research.

2.1 Nanofluidic device

Around the 1990s, challenges for extensive genome sequencing using gel electrophoresis were identified. Based on the different mobility of DNA in different sizes, gel electrophoresis can separate DNA fragments under action of an electric field (Smisek and Hoagland 1989). However, long DNA will be elongated during gel electrophoresis and becomes hard to separate correctly. Researchers have come up with different methods for solving this problem. One of them is using an artificial matrix to replace agarose gels (Riehn et al. 2008). Although, the ensemble sequencing problem was eventually solved by using capillary electrophoresis (Kemp 1998), the way of using artificial matrix evolved into targeting single long DNA using nanofluidic devices.

2.1.1 Physics behind the nanofluidic device

Compared with DNA expanded in free solution, DNA in a nanofluidic device is the laterally confined by hard walls (Tegenfeldt, Prinz, Huang, et al. 2004). Researchers proposed different confinement regimes based on the size of nanochannels which result in different local DNA conformations (Reisner, Pedersen, and Austin 2012). Here we mainly discuss the case when the diameter of the nanochannels is around several DNA persistence lengths ($l_p \sim 50\ nm$). At this

width, DNA falls between the limiting cases of the Odijk and de Gennes regime, which are characterized by straight and string-of-blobs conformations, respectively.

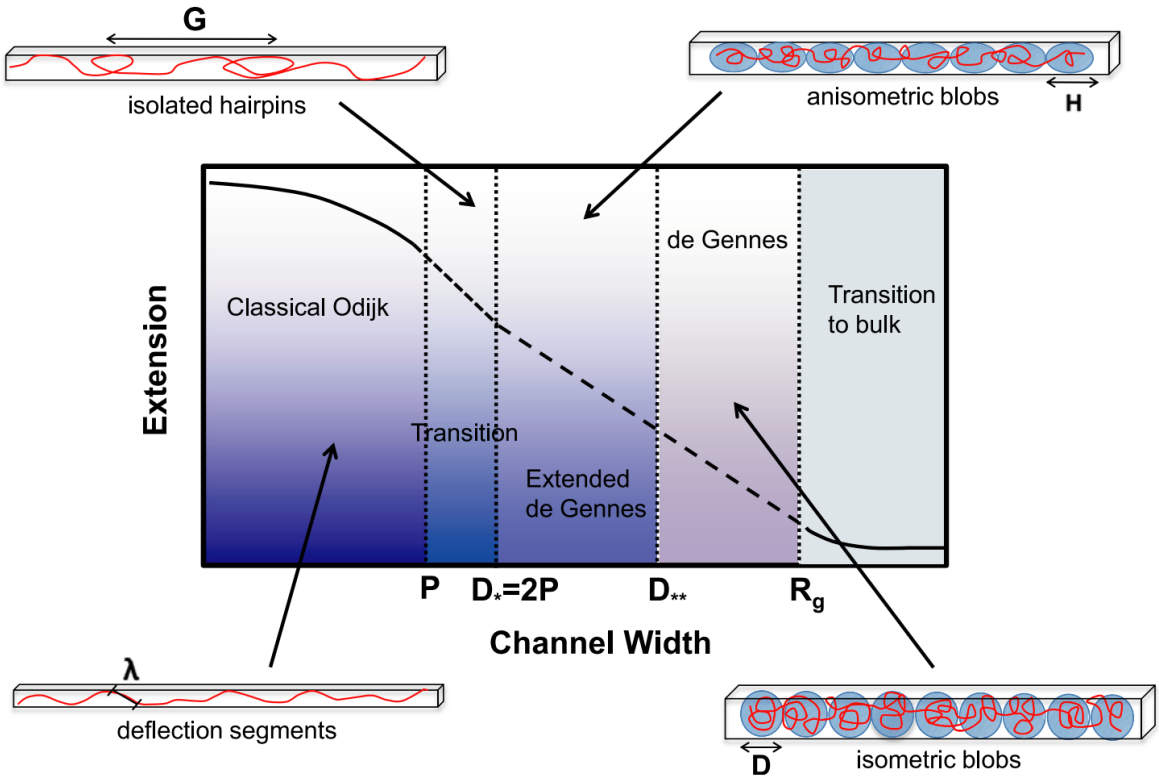

Figure 2.1 Different regimes of nanochannels. The figure was taken from (Reisner, Pedersen, and Austin 2012).

- **De Gennes regime and extended de Gennes regime**

In the previous chapter, we discussed the DNA ideal chain model and the self-exclusion effect. Unlike in free solutions, the self-exclusion effect plays a crucial role in nanochannels. Here we first pull out the free energy we introduced in the last chapter. Using the distribution function (eq.6), we get the free energy of entropy:

$$F_{entropy} = -k_B T ln\big(f(R,N)\big) \cong \frac{3k_B T}{2l_k^2 N} R^2 \qquad (39)$$

Combining the free energy of entropy and the free energy of self-exclusion (eq.23) (excluded volume $\chi = \omega l_k^2$), we get the total free energy for DNA:

$$F_{total} = \frac{3k_BT}{2l_k^2N}R^2 + \frac{\frac{1}{2}k_BT\chi N^2}{R^3} \tag{40}$$

In free solution, we have discussed the result in the previous chapter. The F_{total} minimize at Flory radius (eq.25):

$$\frac{\partial F_{total}}{\partial R} = 3k_BT\left(\frac{R}{l_k^2N} - \frac{\chi N^2}{2R^4}\right) = 0$$

$$R = R_F \cong \left(\omega l_p\right)^{\frac{1}{5}}L^{\frac{3}{5}} \tag{25}$$

In nanoconfinement, DNA will not be able to expand into a sphere. Instead, in the de Gennes regime, which is when the diameter of nanochannel is larger than the persistence length ($D \gg l_p$) DNA forms linked blobs inside the nanochannel (de Gennes and Witten 1980). The diameters of the blobs are the same as the nanochannel's size ($D = R_{blob}$). Considering each blob as a Flory sphere ($R_F = D = R_{blob}$) and combining with eq. 25, we obtain:

$$L_{blob} \cong D^{\frac{5}{3}}\left(\omega l_p\right)^{-\frac{1}{3}} \tag{41}$$

This is the contour length contained in one blob. The extension in the channel direction is the sum of the radius of all blobs at equilibrium:

$$r = r_0 = \frac{L}{L_{blob}}D \cong L\left(\omega l_p\right)^{\frac{1}{3}}D^{-\frac{2}{3}} \tag{42}$$

An interesting note is that this result can be obtained by:

$$F_{total} = \frac{3k_BT}{2l_k^2N}r^2 + \frac{\frac{1}{2}k_BT\chi N^2}{rD^2} \tag{43}$$

Here the minimization comes when

$$\frac{\partial F_{total}}{\partial r} \cong k_B T \left(\frac{3r}{l_k^2 N} - \frac{2 l_p^2 \omega N^2}{r^2 D^2} \right) = 0 \quad (44)$$

We will also get an extension at equilibrium eq.40 (r_0). The free energy equation can be rewritten as:

$$\frac{\partial F_{total}}{\partial r} \cong k_B T \left(\frac{4\omega}{l_k^2 D^2} \right)^{\frac{1}{3}} \left(\frac{r}{r_0} - \left(\frac{r_0}{r} \right)^2 \right) \quad (45)$$

The spring constant at equilibrium ($r = r_0$) was:

$$k = \frac{\partial^2 F_{total}}{\partial^2 r} = k_B T \left(\frac{4\omega}{l_k^2 D^2} \right)^{\frac{1}{3}} \left(\frac{1}{r_0} + \frac{2}{r_0} \right) = \frac{3 k_B T}{r_0} \left(\frac{4\omega}{l_k^2 D^2} \right)^{\frac{1}{3}} \quad (46)$$

Combine with eq. 40, we get:

$$k \cong \frac{6 k_B T}{L l_k} \quad (47)$$

Interestingly it is not related to ω. If the dimension of the cross-section of the nanochannels is $D_1 \times D_2$. Then the total energy provided in Reisner's review is (Reisner, Pedersen, and Austin 2012):

$$F_{total} = \frac{3 k_B T}{2 l_k L} r^2 + \frac{\frac{1}{2} k_B T \chi N^2}{r D_1 D_2} \quad (48)$$

The extended de Gennes regime is when the blobs are no longer spherical. The spherical blobs formed along the DNA were 'squeezed' with a smaller nanochannel cross-section. The longer axis is the direction along the channel. For this anisometric blob problem, the extent of the blob along the channel H and channel width D. The free energy provided in Reisner's review (Reisner, Pedersen, and Austin 2012) is:

$$F_{total} \cong k_B T \left(\frac{R^2}{\left(\frac{L}{L_b} \right) H^2} + \frac{H \left(\frac{L}{L_b} \right)^2}{R} \right) \quad (49)$$

- **Odijk regime**

After the diameter of the nanochannel becomes smaller than the de Gennes regime, DNA molecules will no longer be able to fold back (Odijk 2008). The DNA molecules are extended along the channel direction and deflected by the nanochannel wall. But first, let's discuss the critical size D_{crit} when DNA just be able to fold back.

At critical size D_{crit}, the radius is $R = D_{crit}/2$, and the arc length is $l = \pi D_{crit}/2$ when DNA folds back. So, the bending free energy is:

$$F_{bend} = \frac{kl}{2R^2} = \frac{k\pi}{D_{crit}} \cong k_B T \frac{\pi l_p}{D_{crit}} \tag{50}$$

According to Boltzmann distribution, the probability of DNA folding back is proportional to exp $(-F_{bend}/k_B T)$. It means that $D_{crit} = \pi l_p$ (for DNA, it is around 150nm) when the probability drops. When the size $D \ll D_{crit}$, DNA will not be able to fold back, and the wall inside nanochannels will deflect it. The deflected length is called the Odijk scale given by Odijk:

$$\lambda_0 \cong \left(l_p D^2\right)^{\frac{1}{3}} \tag{51}$$

When the D is small, the angle of deflection is slight. Then, the DNA extension is:

$$r = Lcos(\theta) \cong L\left(1 - \frac{\theta^2}{2}\right) \cong L\left(1 - \frac{D^2}{2\lambda_0^2}\right) = L\left(1 - \frac{1}{2}\left(\frac{D}{l_p}\right)^{\frac{2}{3}}\right) \tag{52}$$

The free energy:

$$F = k_B T N_{deflect} = \frac{k_B T L}{\lambda_0} = \frac{k_B T L}{\left(l_p D^2\right)^{\frac{1}{3}}} \tag{53}$$

2.1.2 Fabrication of nanofluidic devices

The details for the fabrication of our nanofluidic device are mentioned in the next chapter. Here we mainly focus on the general descriptions. The nanofluidic device contains channels at a nanometer scale. For DNA manipulation experiments, those channels are usually around several persistence lengths of DNA. For nanochannel fabrication, the first crucial decision is which material to use. The material used for making nanofluidic devices can be divided into two groups: hard and soft (Riehn et al. 2008).

Soft materials refer to polymeric materials, like PDMS (Peng and Li 2016). The advantage of using soft materials is the fast and low-cost patterning. Unlike the hard materials, which need to be patterned individually, soft materials can be patterned by molds. However, nanochannels made from soft materials can attract DNA molecules onto the surface and have a stronger fluorescence background than fused silica. Especially for single molecules experiments, these properties increase the difficulty of DNA manipulation. Most hard materials refer to crystalline materials, like silica. In our experiment, the material for the nanofluidic device was fused silica (M. Liu et al. 2020). Since the surface charge of fused silica is negative in most physiological buffers, DNA molecules, which have negatively charged backbones, are unlikely to adhere to the nanochannel walls. Also, the fluorescence background of the fused silica is much less than that of soft materials and ordinary glass. These properties enable us to perform biological experiments at the single molecule level.

Unlike soft materials, our fused silica nanofluidic devices are made by e-beam lithography. The nanofluidic device made of hard materials can also be patterned by FIB (focused ion beam) and

imprint lithography. By reactive ion etching (RIE), we could etch the fused silica with nano slits and create nanochannels by sealing it with another piece of fused silica. However, despite the good properties of fused silica for biological experiments, the cost of creating each device is much higher than using soft materials. Also, due to the negative charge along the nanochannel walls, DNA binding proteins, which usually contain positively charged binding domains, are likely to stick onto the channel walls. This can be a big challenge when working with the DNA-protein systems. Researchers also have spent lots of effects on diminishing this nonspecific adhesion (Persson et al. 2012).

After patterning, the next step is sealing the nanofluidic device. We used the thermal annealing method to seal the patterned nanofluidic device. After RCA cleaning and drilling access holes, polished fused silica anneals to another piece of fused silica by heating up to 1000 ℃. This way, we will be able to create nanochannels used for biological experiments.

2.2 Atomic Force Microscopy (AFM)

Nanochannels are an excellent way to analyze the influence of proteins on long, genomic sized DNA molecules. However, when the size of DNA molecules is small, the analysis will be impossible because of the spatial resolution limitation of the light in optics. One of the solutions is using AFM (Atomic Force Microscopy).

The invention of AFM was around 1986, with the first experiment done by Binnig (Binnig, Quate, and Gerber 1986). Since then, AFM has been widely used in different areas. Compared with other imaging methods beyond the optical diffraction limit, like SEM (Scanning electron

microscopy), AFM creates 3D information and does not require extra coating for charge dissipation. Also, unlike SEM, AFM does not use an electron beam and does not create a charge artifact to the sample. The regular operation of AFM can be done in the atmosphere and does not require a high vacuum. These advantages make AFM a suitable candidate for biology experiments since most biological samples are soft and fragile.

2.2.1 Basic principle of AFM

There are mainly three operation modes in AFM: contact mode, tapping mode, and noncontact mode (Jalili and Laxminarayana 2004). Contact mode is when the cantilever tip is dragged along the sample surface when scanning in the x and y directions. Height information of the testing sample is transformed into the deflection of the cantilever, which will, in turn, change the position of the laser spot on the photodetector. The computer records this change of laser with the x and y coordinates information and then outputs a color map showing the height information at each (x,y) position. Usually, a soft cantilever is needed to reduce the force applied to the sample. However, AFM operating in this mode can cause damage to the sample, and the signal is noisy, making it not a good choice for imaging biological samples.

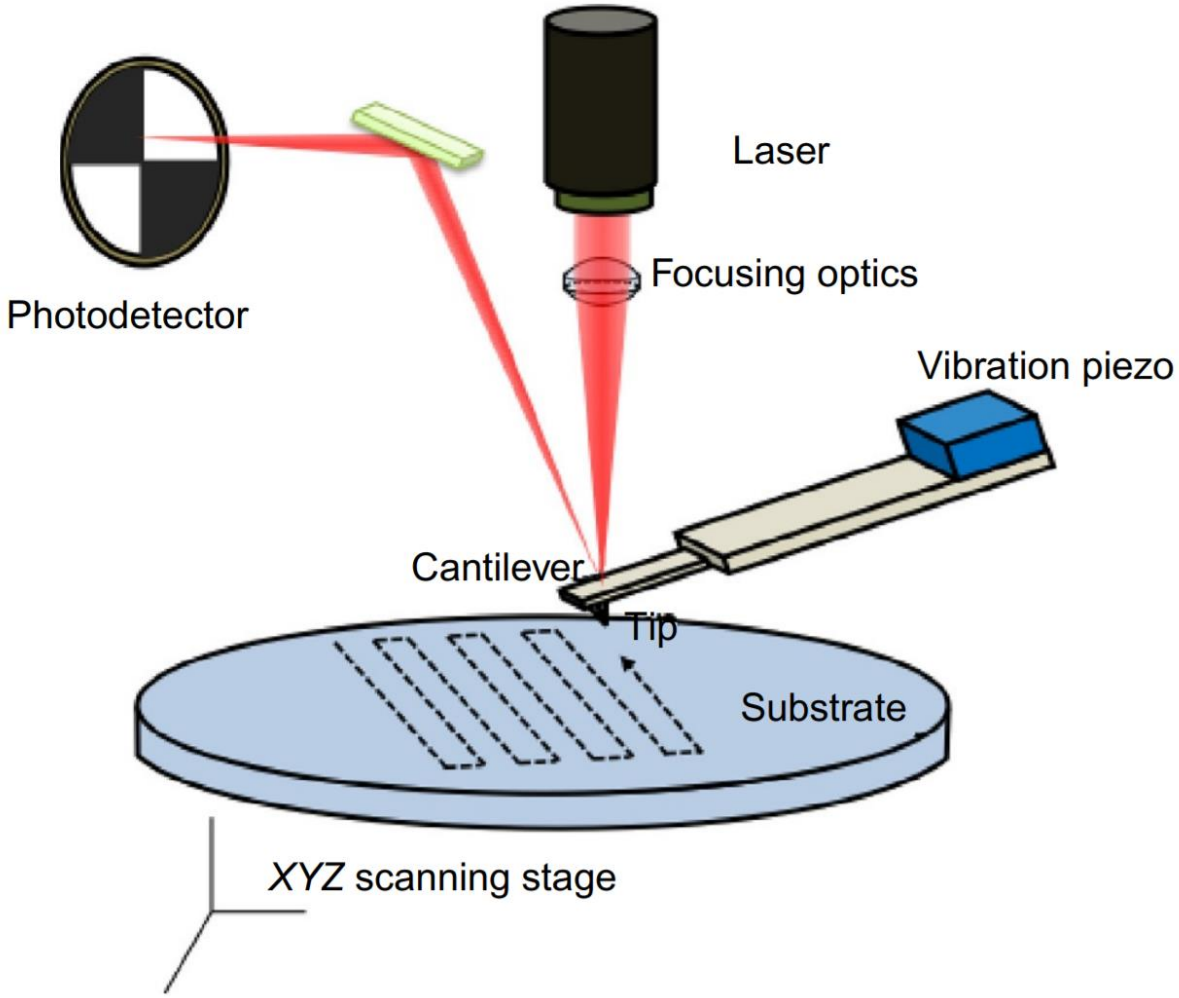

Figure 2.2 A schematic diagram of AFM. The figure was taken from (LeBlanc et al. 2017). The configuration of AFM is shown in Fig. 2.2. The central part is the cantilever with tip, cantilever holder with vibration piezo, laser, and detector systems. The x-y scanning is based on the movement of the cantilever or the stage. In our case, the cantilever is moving in the x-y direction.

The mode we usually use is the tapping mode (M. Liu et al. 2020). In tapping mode, the cantilever tip does not always contact the sample surface. Instead, the cantilever tip is driven to oscillate near the eigen frequency by the vibration piezo elements and tap the sample surface. When the cantilever tip gets close to the sample surface, the oscillation amplitude of the cantilever tip will change because of the van der Waals force or other forces. This amplitude change will be compensated by moving the z-axis of the cantilever through a feedback loop so that the amplitude remains the same. This change will be transformed into the height information being recorded. The phase change is also recorded during the scan, representing different adhesion and stiffness. In our case, it can be used to tell whether only two DNA stack on each other or there is one protein involved. Also, the amplitude is not the only parameter that can be used to modulate. Some AFM uses frequency or phase to modulate the tip holder. Compared to the contact mode, the tapping mode does not create a long-time force on the sample surface, which means it is better for imaging some fragile samples.

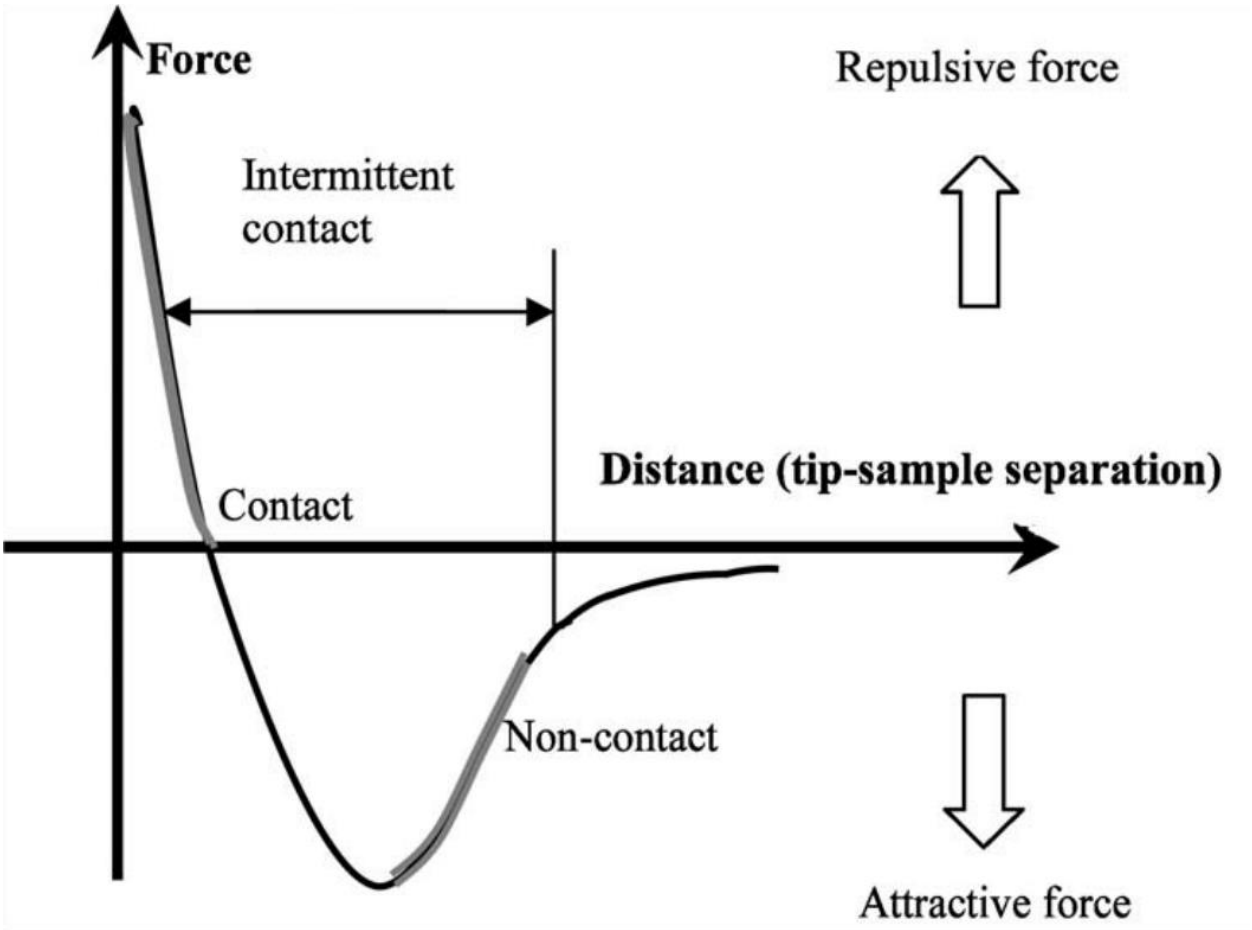

Figure 2.3 Force between tip and sample versus distance. The figure was taken from (Jalili and Laxminarayana 2004). Different modes work at different working distances. The force between tip and sample can be complicated, including van der Waals and electrostatic force.

The non-contact mode does not require contact between the surface and the tip. When the tip comes close to the surface, the force will change the amplitude or phase. Then the feedback loop will automatically modulate to keep the amplitude or phase, and this information can be recorded pixel by pixel. The non-contact mode does not require contact between tip and sample, which causes minor damage and the tip lasts longer. Usually, the force is low at around 10 pN (Jalili

and Laxminarayana 2004). However, one notable difference between the non-contact mode and the tapping mode in our case is a layer of water on mica when imaging in the atmosphere. The non-contact mode will not penetrate the water layer and image the water layer and sample simultaneously. This can influence the quality of imaging.

2.2.2 Applications of AFM

The ability to visualize samples in nanometer scales and measure force at the nano-newtons range makes AFM a good candidate for measurement in molecular metrology and biological sciences (Jalili and Laxminarayana 2004). In molecular metrology, AFM has been used in studies of microstructures (Westra and Thomson 1995) and nanofabrication (Dagata et al. 1998).

In the biological sciences, applications include characterizing conformation changes of DNA (LeBlanc et al. 2017; M. Liu et al. 2020), mechanobiology (Krieg et al. 2019), force spectroscopy (Puchner and Gaub 2009), and cell imaging (Henderson, Haydon, and Sakaguchi 1992).

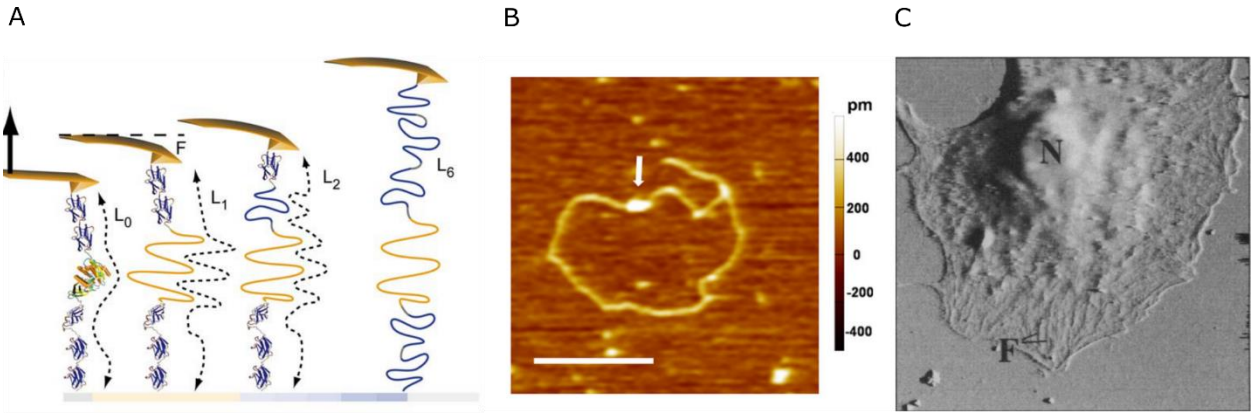

Figure 2.4 Application of AFM on biological science. (A) Force-extension experiment. Image taken from (Puchner and Gaub 2009). (B) The protein binds to DNA. The image is taken from (Countryman et al. 2018). (C) Actin filament imaging in cells. The image is taken from (Henderson, Haydon, and Sakaguchi 1992).

2.3 Fluorescence Anisotropy

There are a number of ways to test the binding between the protein and DNA, like electrophoretic mobility shift assay (EMSA) and surface plasmon resonance. Fluorescence anisotropy can quantify the binding affinity. It does not require enormous amounts of samples, and it is sensitive. These properties make it a suitable candidate for our biological experiments.

2.3.1 Principle of fluorescence anisotropy

The fluorescence anisotropy is measured by comparing the intensity difference between two components of emission light $I_{\parallel}$ and $I_{\perp}$. The $I_{\parallel}$ and $I_{\perp}$ are created by polarizer orient in the vertical and horizontal direction. Here the $I_{\parallel}$ represents the component that is parallel to the polarization direction of excitation light and the $I_{\perp}$ represents the component that is

perpendicular to the polarization direction of excitation light. The definition of fluorescence anisotropy is:

$$F.A. = \frac{I_{\parallel} - I_{\perp}}{I_x + I_y + I_z} = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + 2I_{\perp}} \tag{54}$$

The primary mechanisms for fluorescence anisotropy is rotational Brownian motion. The polarized excitation light can excite fluorescent molecules. Those excited molecules are oriented in a range of directions compared to the excitation light's polarization direction. However, since the fluorescence is inside the liquid, those fluorescence molecules collide with other molecules (Brownian motion) and change their direction. The primary purpose of fluorescence anisotropy is to analyze whether the molecules' size has changed by estimating the direction changes. The change of the fluorescence anisotropy estimates the quantification of the direction changes. The rate of this tumbling behavior can be estimated by the rotational diffusivity:

$$\phi = \frac{k_B T}{\eta V} \tag{55}$$

Here V is the volume of the nanoparticle and η is the viscosity of the liquid. Since we can control temperature and viscosity constant in the experiment, the rotational diffusivity mainly depends on the nanoparticle size. The mean time delay between fluorescence excitation and emission is called the lifetime of the fluorescence dye (τ). The rotation during this time delay follows $\langle \delta\theta^2 \rangle = \phi\tau$. Usually, the rotation during the fluorescence lifetime leads to nearly complete loss of orientation for the operation of fluorescence anisotropy. A slight rotation (short life time) will make it hard to distinguish from the noise, and a significant rotation (large life time) will be averaged because the molecules have rotated several rounds before measurement.

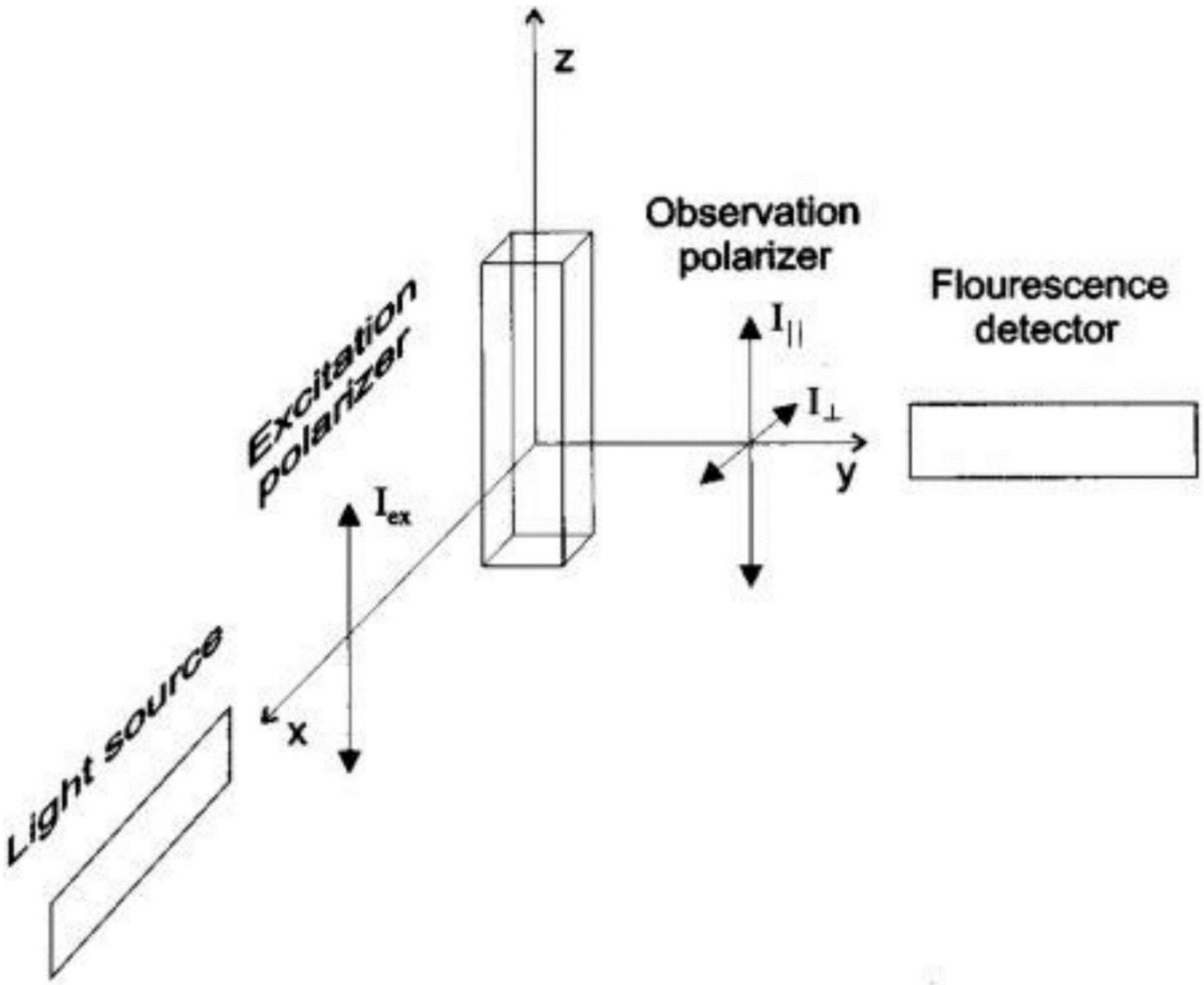

Figure 2.5 Schematic diagram of the fluorescence anisotropy measurement configuration. Taken from (Kuśba and Lakowicz 1999).

Usually, the anisotropy is calculated by a function that contains an instrument-specific G factor that represents the difference between vertically and horizontally polarized light for the machine. The definition of the G factor is:

$$G = \frac{I_{HV}}{I_{HH}} \tag{56}$$

Here, the I_{HH} means both the polarizer for emission and excitation light are horizontal. The I_{HV} means the excitation light is horizontal and emission light is vertical. Using a horizontal

excitation light, the intensity of the excitation in both horizontal and vertical should be the same in theory. So, the difference represents the error caused by the machine. Using the G factor the anisotropy becomes:

$$F.A. = \frac{I_{\parallel} - GI_{\perp}}{I_{\parallel} + 2GI_{\perp}} \tag{57}$$

2.3.2 Application of fluorescence anisotropy

Since the fluorescence anisotropy can detect the size change of the fluorescent molecule, a typical application is measuring the binding affinity of a reaction or a binding event. In our experiments for DNA binding protein, we used fluorescently labeled short DNA as our DNA substrate and did a protein titration. The concentration of initial DNA and protein are $[D]_0$ and $[P]_0$. Then after incubation, the solution contains three parts, free DNA, free protein, and DNA-protein conjugate. Here we simplify by assuming that the molar ratio when binding between DNA and protein is 1:1, which means we do not consider multiple proteins binding. Then, it always follows:

$$[D]_0 = [D] + [DP] \tag{58}$$

$$[P]_0 = [P] + [DP] \tag{59}$$

The definition of the affinity is:

$$K_d = \frac{[D][P]}{[DP]} \tag{60}$$

Combining equations 56, 57, and 59, we can get the expression in terms of $[DP]$, the K_d, and the initial DNA and protein concentration:

$$[P]_0([D]_0 - [DP]) = K_d[DP] + [DP]([D]_0 - [DP]) \tag{61}$$

$$[DP]^2 - ([P]_0 + [D]_0 + K_d)[DP] + [P]_0[D]_0 = 0$$

After simplifying, we will get $[DP]$ expression in terms of the K_d, the initial DNA and protein concentration (Desai et al. 2015):

$$[DP] = -\frac{[P]_0 + [D]_0 + K_d \pm \sqrt{([P]_0 + [D]_0 + K_d)^2 - 4[P]_0[D]_0}}{2} \quad (62)$$

Also, detected fluorescent light is the combination of the light from fluorescently labeled DNA with and without protein. And the intensity depends on concentration. So, it follows:

$$F.A. = ([D]F.A._D + [DP]F.A._{DP}) = [D]_0 F.A._D + [DP](F.A._{DP} - F.A._D) \quad (63)$$

Accordingly, since the fluorescence anisotropy is proportional to the $[DP]$, we can use this equation to fit the experimental result to obtain the dissociation constant K_d for this DNA-protein interaction

CHAPTER 3 METHODS AND MATERIALS

In the previous chapter, we talked about three instruments we usually use and their principle. In this chapter, we will focus more on the details of methods and materials in our experiments.

3.1 Nanochannel experiments

In the previous chapter, we talked about the physics behind nanochannel stretching of DNA. Here we will talk about our experimental setup, fabrication of nanochannels, and the experimental methods.

3.1.1 Experimental setup

Our setup for nanochannels contains three systems: the imaging system, the flow control system, and the nanofluidic system, as shown in Fig. 3.1. The imaging system contains a fluorescence microscope combined with a laser optical setup. We have four laser emitters with different colors, which allow us to imagine different fluorescent dyes. Fluorescent labeling, which we will describe in more detail in a later chapter, includes the DNA dyes YOYO-1 and YOYO-3, the fluorescent dye ATTO 565-maleimide (used for labeling Cysteine), and ATTO 555-NTA (used for labeling His-tag). The inverted fluorescence microscope (Nikon TE-2000) is connected to the laser system. A Nikon 60× or 100× oil immersion objective (Apo TRIF NA=1.40) was used in our experiment, and the data were collected by an emCCD (Andor iXon Life) through a QV2 image splitter (Photometrics) with four bandpass filters at a different wavelength.

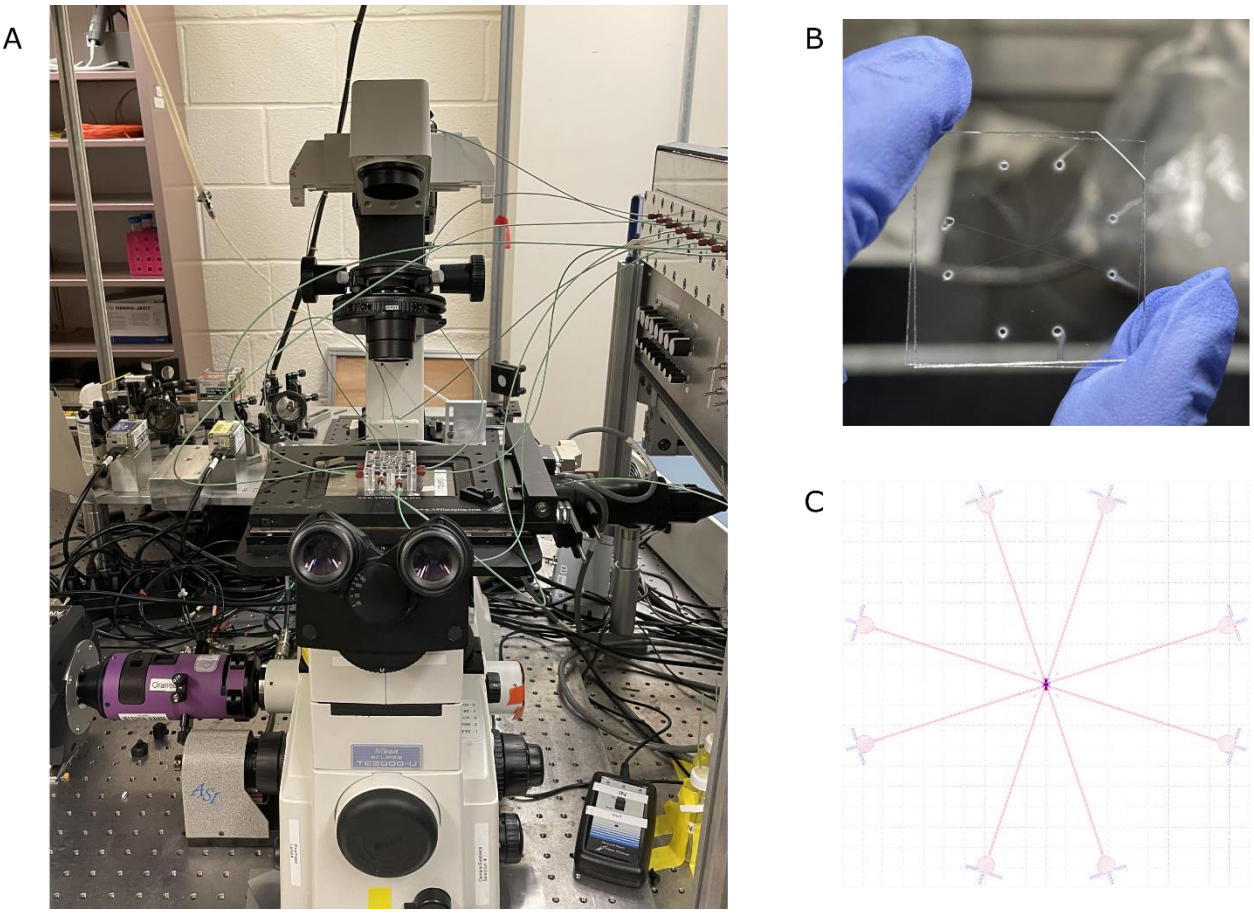

Figure 3.1 Setup for nanofluidic experiments. (A) imaging system setup (B) nanofluidic chip (C) layout of the nanofluidic chip.

The flow control system is to control the nitrogen pressure. Our nanofluidic device is driven by nitrogen flow. So, it is necessary to have a delicate control system. We usually use ~ 2 bars to flush. We flush the sample using nitrogen from a reservoir into the microchannel or nanochannel. For the combinatory channels, we use 100-mbar pressure to prevent the DNA from popping out the nano slots. The detailed methods will be presented in later chapters.

The core of the nanofluidic device is the nanofluidic chip, as shown in Fig.3.1 (B). There are a metal plate and a glass block to hold this nanofluidic chip. Inside the glass block, some channels

are connected to the controlling system for the nitrogen flow. Also, reservoirs inside the glass block hold the sample for the experiment. These samples can be pushed by nitrogen into the center of the chip for a later experiment.

3.1.2 Fabrication of the nanofluidic chip

In the previous chapter, we mentioned using hard material (fused silica) for our chip. Here, we are going to talk more about the details of fabrication.

- **Alignment markers**

Nanofluidic chips are made from fused silica wafers. Those fused silica wafers have a well-polished surface for later thermal bonding. The fabrication of the nanochannels contains several steps of lithography. It is necessary to have alignment markers to make sure we are patterning at the correct position. There are several steps in making those alignment markers:

1. Rinse and dry to clean wafer.
2. Spincoat a layer of LOR-3A at 4000 rpm and prebake at 180 ℃ for 5 min.
3. Spincoat a layer of SPR220 at 4000 rpm and prebake at 115 ℃ for 90 s.
4. Patterning markers using MA16 and post bake at 115 ℃ for 90 s.
5. Development in CD26.
6. Clean the residues with oxygen plasma.
7. Metal (platinum) deposition, thickness is around 50 nm
8. Lift off using MICROPOSIT Remover 1165 (N-Methyl-2-pyrrolidone (NMP))

- **Shunt region**

The Shunt region only shows up in our 8-port device. The purpose of having this region is for buffer exchang or sequential protein exposure. The shunt region's main function is to ensure that the DNA stays in the nano slots rather than popping out into the shunt region. The competition is between the flow and the entropy preference (Mikkelsen et al. 2011). In our case, the typical depth of the shunt region is 30 nm to 50 nm. The depth of nano slots is usually 150 nm to 180 nm. This combination is good enough for the DNA to be held. To make the shunt region, we perform the following steps:

1. Clean the wafer and coat it with a layer of HDMS (use an HDMS oven).

2. Spincoat a layer of photoresist (like SPR220) and prebake at 115 ℃ for 90 s.

3. MA16 aligner for patterning the shutter channel through a mask (40 μm gap, 5.5 s expose time) and post bake at 115 ℃ for 90 s.

4. Use CD-26 developer for 60 s and clean with DI water.

5. Check the pattern and use Oxford NGP80 to etch the wafer. The etching starts with oxygen cleaning (O_2 100 sccm, 35 mTorr, 200 W power) and seasoning for a better etching rate. For our shunt region, which is 30-50 nm deep, we used 2 min and 10 s (O_2 2 sccm, CHF_3 50 sccm, 35 mTorr, 200 W power, usually 22 nm per second but varies on different days).

6. Clean the photoresist with 1165 and oxygen plasma clean.

- **Nanochannels or nano slots**

Nanochannels or nano slots are patterned by e-beam lithography. The steps for making nanochannels or nano slots are similar. So here we combined these two. Usually, the size of

nanochannels in our experiments is around 150 nm. To make the nanochannels or nano slots, here are the steps:

1. Clean the wafer and heat the wafer at 115 ℃ for 90 s to dehydrate.

2. Spincoat a layer of ZEP520A for about 200 nm (usually, we use 1.5 diluted ZEP520A spin at 3000 rpm). The main thing is to ensure the ZEP520A will not be depleted when we do the reactive ion etching. Remember, this step should not have an HDMS layer before the deposition of ZEP520A. Prebake at 180 ℃ for 3 min to dehydrate.

3. Deposit 25 nm of Au through physical vapor deposition. The purpose is to create a conductive layer.

4. Use Raith 150 Two to pattern the target nano patterns onto the wafer (default 20 kV, 30 μm aperture, and 10 mm working distance.

5. Immerge wafer into Au etch (KI /I_2 /DI H_2O) for 30 s to get rid of Au. Clean with DI water and nitrogen. Heat at 115 ℃ for 60 s to dehydrate.

6. Develop ZEP 520A in amyl acetate for 60 s. Stop development by a flush with IPA.

7. Check the pattern under an optical microscope in darkfield mode.

8. Use Oxford NGP80 to etch the wafer. Condition setting for the etching is the same. Etching always starts with oxygen cleaning. Here we used 6 min and 45 s.

9. Clean the ZEP520A with 1165 and oxygen plasma clean.

- **Microchannels**

The microchannels connect the nanochannels and reservoirs with the primary purpose of pass the DNA and protein through. The depth does not matter a lot as long as it is deeper than the shunt region. Usually, we make the microchannels at around 1 μm. The fabrication steps for

microchannels are like the shunt region. So here we are not going to talk about the details about it. The etching time is 55 minutes. Furthermore, we do not strip the photoresist away after etching because we need it for later dicing and bonding.

- **Dicing and bonding**

In our experiments, the chip size is 25 mm × 25 mm. The size of these chips is determined by dicing the treated fused silica wafer. We operate a standard dicing protocol with a few modifications. First, wafers are mounted with the patterned side in contact with the plastic mounting film. Secondly, a resinoid blade is used for dicing. Third, the wafer thickness is 550 μm

After dicing, there are several steps before we get the final channels.

1. Sand blasting the chip at the entry position. This step is to create a connection between the reservoir and the chip.

2. Put the chip into 1165 overnight. This step is to get rid of the plastic mounting tape and photoresist. In 1165, the sands were not easy to stick on the chip. Also, in most cases, we will put our coverslips into 1165 overnight. Those coverslips are used for thermal bonding later.

3. Wash and clean the chip and coverslips with DI water.

4. Do an RCA clean for the coverslips and chips. RCA clean contains two parts. One is the acid ($HCl/H_2O_2/H_2O$ = 1:1:3), and the other is the base ($NH_4OH/ H_2O_2/H_2O$ = 1:1:3). Both solutions need to heat to 75 ℃ during the reaction. The reaction is first using the acid step to clean any metal residues on the chip and then the base step to clean organic residues. The base

also creates a high density of hydroxide groups on thee fused silica surface, which is essential for the bonding.

5. Dry the chips with nitrogen and attach the chip and coverslip. Then put the combined chip into the furnace and heat to larger than 1000 ℃. Make sure there are no trapped bubbles near the channel before the heat.

3.1.3 Surface passivation

Here we focus on one of the major problems during the nanochannel experiment, the need for surface modification of the internal surfaces of the device to prevent excessive absorption of DNA-binding proteins. Nitrogen pushes DNA and protein samples from the reservoir into the microchannel and nanochannel. The fused silica channel wall is negatively charged, and our channel size is larger than the Debye length. Hence, the DNA is efficiently confined since it has a negatively charged backbone. In contrast, DNA-binding protein usually contains positively charged domains that are pulled onto the negatively charged channel walls. Researchers have tried to prevent the sticking by surface passivation using BSA (Sweryda-Krawiec et al. 2004; Riehn et al. 2005), Polyethylene glycol (PEG) (S. Lee and Vörös 2005), and lipid (Persson et al. 2012). However, those methods are only successful for specific protein system. Our experiments focused on using poly-dopamine grafted mPEG-NH_2 as our passivation method (H. Lee et al. 2007). The coating includes a two-hour dopamine coating with 2 mg/ml dopamine solution in 10 mM Tris buffer at pH 8.0 and an overnight mPEG-NH_2 coating with mPEG-NH_2 in 10 mM Tris buffer (pH 8.0). The coating was successful for most proteins attempted in our work.

3.2 AFM experiment

Since the diffraction limit does not allow resolution of features below around 200 nm, we need to use other methods to examine the binding pattern of proteins with a size of 10 nm or less to DNA. An atomic force microscope (AFM) is a good candidate for our purposes. The principle behind the AFM has been discussed in the last chapter. Here we are going to talk more about the details of the experiment.

3.2.1 Biological preparation

Biological samples are attached to the mica before imaging. Since both the mica surface and the DNA surface are negatively charged, we usually use magnesium ions as the linker between them. Here are the steps before imaging:

1. Stick one mica onto the small metal disc.
2. Use plastic tapes to peel the mica and create a flat surface for later deposition.
3. Prepare biological samples, including incubation time.
4. Dilute samples in AFM imaging buffer (HEPES 20 mM, $Mg(OAc)_2$ 10mM, NaCl 100mM, pH 7.5). Since the imaging buffer contains a high concentration of magnesium which may influence DNA-protein interactions, the mixture needs to be deposited onto mica quickly.
5. Deposit mixture on mica and wash with 1 mL of DI water.
6. Dry the mica with nitrogen, and it is ready for imaging.

Although the AFM can image a large area over a long time, we usually have a 1-5 μm frame size for our experiment. In this way, we can obtain a high resolution with few artifacts. For a suitable

density on mica, we usually dilute the biological samples into around 0.7-1 ng/μL of DNA or 10-30 nM of protein.

3.2.2 Imaging steps

After the preparation of the mica, the mica will be put on the imaging stage. The imaging needs some preparation, and here are the steps.

1. Assemble the cantilever with the holder.

2. Attach the holder to the AFM head.

3. Align the laser to the top of the cantilever. This should create a large sum number, meaning laser light was reflected into the detector.

4. Adjust the DP to make the deflection to be zero. This step is to make the reflected laser spot at the center of the detector.

5. Do an auto tune to ensure the cantilever tip is at the correct position. The curve should look like a gaussian curve, and our frequency is usually around 75 kHz. This way, the cantilever can start to vibrate at a low voltage.

6. Set scan size (3 μm), points (512), and scan frequency (1 Hz). Start the engagement of the tip.

7. Lower the tip to make the z voltage smaller. Then change the set point to smaller and lowering down the tip again. This step is the soft approach.

8. Close the cabin to isolate the AFM from the environment noise.

3.3 Biological samples

3.3.1 DNA samples

The nanochannel experiments of the MBD project (chapter 4) were performed using λ-DNA from Sigma-Aldrich (~48 kbp). A subset of these experiments required CpG methylation, which was obtained as described here. Following New England protocol, the lambda DNA was then methylated with M.SssI (New England Biolabs). The enzyme can methylate the CpG sites along with the DNA, which requires S-adenosylmethionine (SAM) for methyl group transfer. The reaction was incubated at 37 ℃ for 2 h for complete methylation of the CpG sites on the lambda DNA. This reaction can be stopped at 65 ℃ for more than 20 minutes. We tested the methylation by a restriction enzyme. We used HpaII (New England Biolabs) for the test. Since HpaII recognizes the CCGG site and is methylation-sensitive, HpaII will not be able to cut this specific site after methylation of CpG. The DNA was labeled by YOYO-1 at a 1:10 dye to base-pair ratio for fluorescence imaging. To break the hybridization between the complementary ends of two different lambda DNA, we heat the DNA to 65 ℃ for 10 min and rapidly cool it down to room temperature before use. For the AFM experiment for the MBD project, we used a linearized DNA plasmid with a 1,697-bp methylated CpG-rich region and two 2,742-bp and 2,724-bp CpG-free regions (Pan et al. 2017).

For the telomeric cohesion project (chapter 5), we need a long telomeric DNA. Because of the diffraction limit, a DNA with a length less than 1 kbp will not be long enough to determine the length change after protein exposure. Also, the DNA inside nanochannels extends to only around one-third of its contour length. We ligated short telomeric DNA into a long one to get long telomeric DNA. We digested the T270 plasmid (5.4 kb) with XbaI, and BglII followed the NEB

protocol. The T270 plasmid, which contains two telomeric sequence regions with 135 (TTAGGG) repeats linked by a 23-bp non-telomeric sequence, was purchased from Addgene (PSXneo(T2AG3)). Two DNA fragments after digestion were separated by gel and purified by a gel purification kit. Since the XbaI and BglII will create complementary ends, we took advantage of this by directly ligating the DNA fragment containing telomeric sequences. To make a telomeric DNA that is long enough, we must do multiple ligations between the DNA. We concentrate the DNA to 500 ng/μl after gel purification to decrease the probability of self-ligation creating looped DNA. The ligation mixture was incubated at room temperature overnight. This is to make sure it was ligated into a long DNA. The resulting DNA was purified by phenol-chloroform extraction. The telomeric DNA was labeled by YOYO1 at a 1:20 dye to base-pair ratio for nanochannel experiments.

3.3.2 Fluorescent labeling of proteins

We investigated proteins that were either prepared by collaborators or purchased from commercial sources. Most proteins were used as provided. However, a subset was labeled using fluorescent dyes using the methods described below.

In our MBD project, MBD2FLsc is a full-length human MBD2 single-chain construct comprising MBD2 (amino acids 150–393) and the scMBD2-GATAD2A expressed in E. coli and purified by nickel affinity chromatography (Desai et al. 2015). We labeled this protein by ATTO-565 maleimide through the covalent interaction between maleimide and the thiol group of the cysteine amino acids. The protein was first treated by TCEP for 20 min at room temperature to decrease the disulfide bond. Then mix with 20× fold excess dye and incubate at 4 ℃

overnight. To separate the protein and dye, we used gel filtration for purification. Those gel beads contain holes that can store dyes while it is small for proteins, making proteins come out faster than dyes. The labeling efficiency was measured by measuring the absorbance spectrum. It follows:

$$[P] = \frac{A_{280} - A_{max} * CF}{\epsilon_{protein}} * dilution\ factor \quad (64)$$

$$[dye] = \frac{A_{max}}{\epsilon_{dye}} * dilution\ factor \quad (65)$$

Here, CF is the correction factor representing the component at 280 nm from the dye and ϵ is the extinction factor. It represents the light absorbance per molar concentration of the dye or protein. The labeling efficiency is the ratio. The activity of the protein is measured by fluorescence anisotropy, described in the last chapter.

CHAPTER 4 COMPACTION BY PROTEINS FROM THE MBD FAMILY

The following chapter is chiefly based on (M. Liu et al. 2020). All data presented here are my work. Collaborators contributed by preparing proteins and establishing baseline conditions for DNA preparation and protein-DNA binding experiments. I performed all data analysis and formed conclusions. The manuscript was written collaboratively.

4.1 Introduction

Epigenetic regulation, the inheritable regulation of gene expression without changing the DNA sequence, often involves modifications of DNA bases and the specific binding of proteins to those modifications. One of the most widely studied epigenetic mechanisms is cytosine methylation, which is usually referred to as DNA methylation. It is an inheritable chemical modification that regulates gene transcription (Dor and Cedar 2018). This regulation is either by dedicated binding proteins that serve as a reader of the epigenetic state (Du et al. 2015) or by modulating the assembly of the transcription machinery directly (Jin et al. 2016). Understanding the mechanisms through which DNA methylation influences gene expression is an active research field due to the growing evidence of its relation to human development and disease, and in particular, aging (Horvath and Raj 2018), environmental response (Barouki et al. 2018), and cancer (Gokul and Khosla 2013).

Our focus is on probing how members of the methyl-CpG-binding domain (MBD) protein family, which share a common methyl-binding domain (Mahmood and Rabbani 2019), alter the conformation of DNA molecules. Our interest arises from two lines of thought. Firstly, demonstrating that an MBD protein compacts DNA as part of binding leads to the suggestion

that the mechanism of modulation of transcriptional activity can be an effect of steric compaction. Such mechanisms have been explicitly proposed for some MBD (Georgel et al. 2003). Secondly, MBD binding has been established as a label in epigenetic mapping technologies, and the lack of compaction would be a necessary condition for a specific epigenetic label. Such mapping technologies are emerging in a niche where bisulfite-based sequencing has remained very difficult, such as for single cells (Miura et al. 2012; Farlik et al. 2015; Zhu et al. 2018), or where the variability over large scales (>100 kbp) is required without single-base resolution but with high throughput.

Mapping of methylation patterns on extended DNA molecules intrinsically yields single-molecule information, and there are three main pathways to obtaining CpG methylation patterns using fluorescent probes (Levy-Sakin and Ebenstein 2013). The first is using antibody probes, which, however, work best in the context of fixed and denatured DNA (Mathieu, Picard, and Tourmente 2002; Koo et al. 2011). Thus, this approach is incompatible with high-throughput DNA stretching in our nanochannels. The second approach is the covalent labeling of short motifs using methylation-specific proteins or methylases to create a barcode-like pattern on DNA molecules (Deen et al. 2017). Usually, the flanking sequence of the CG site will strongly influence the specificity. The third method is the use of DNA-binding proteins to investigate the epigenetic information (Cerf et al. 2011; S. F. Lim et al. 2011). In this paper, we expand on that last method by investigating two full-length MBD proteins that could be used to provide methylation-specific labeling of DNA molecules.

All MBD proteins have a methyl-CpG-binding domain that enables them to bind to methylated CpG sites specifically. We focus on the family of proteins related to MeCP2. Members of this family with methylation-specific binding are MeCP2, MBD1, MBD2, and MBD4 (Hendrich and Bird 1998). Based on their specificity in binding to methylated CpG substrates, MBD proteins are good candidates for mapping methylation patterns on stretched DNA (Cerf et al. 2011; S. F. Lim et al. 2011). Previous publications demonstrating mapping using MBDs have used either capillary force DNA stretching combined with a peptide from MBD1 as a probe (Cerf et al. 2011), or nanofluidic channel stretching combined with an MBD peptide from MeCP2 (S. F. Lim et al. 2011). While both studies demonstrated specific labeling, and both utilized only the methyl-binding domain from the respective protein, Lim et al. (S. F. Lim et al. 2011) reported shortening of the DNA, while Cerf et al. (Cerf et al. 2011) did not report this phenomenon.

The stretching technique employed by Lim et al. is confinement to nanofluidic channels with a cross-section of 100nm×100nm to 200nm×200nm, which extends DNA through the interplay of excluded volume (self-avoidance), DNA stiffness, and confinement (Tegenfeldt, Prinz, Cao, et al. 2004). DNA is in an equilibrium state and thus free to fluctuate and change its conformation as proteins bind, buffer conditions change, or proteins perform catalytic reactions (Frykholm, Nyberg, and Westerlund 2017). Studies of DNA configurations in nanochannels under protein exposure have yielded insights into the modification of the mechanical parameters of DNA. Of specific interest for this paper is the probing of DNA–protein–DNA bridges, which can stabilize DNA loops or manifest themselves in an apparent modified effective DNA width. The formation of such DNA–protein–DNA bridges scales with the rate of DNA–DNA contacts and thus is strongly dependent on DNA concentration. Typical concentrations of in vitro assays (10s of

μg/ml) are far too low to mimic the environment of a cell nucleus. The high local concentration of nanochannel-stretched DNA (>100 μg/mL), while still not quite at a physiological level, leads to a much higher rate of DNA–DNA contacts (Reisner et al. 2007). In addition, nanochannel stretching requires no tethering and lacks external stresses, thus making nanochannel stretching an excellent platform to test protein-mediated DNA–DNA interactions and DNA–protein–DNA bridge formation.

In this study, we focus on two proteins from the MBD family, MBD2 and MeCP2. MBD2 interacts with the nucleosome remodeling and histone deacetylation (NuRD) complex (Y. Zhang et al. 1999). It is expressed throughout the body (Wood et al. 2016). Studies have shown that MBD2 helps NuRD target methylated CpG sites to modify the chromatin and regulate gene transcription (Desai et al. 2015; Pan et al. 2017). We are not aware of a co-operative mechanism in MBD2 binding, including in vivo correlations with the density of target sites (Baubec et al. 2013). MeCP2 is highly expressed in the brain and well-studied because of its impact on brain function and disease (Fasolino and Zhou 2017). MeCP2 is capable of co-operative binding to DNA oligomers that contain a single specific binding site (Khrapunov et al. 2016), with salt-dependent binding constants and cooperativity. However, this cooperativity in general decreases the specificity of binding. Interestingly, an effect of methyl CpG density-dependent binding, with higher affinity for higher densities, is apparent in vivo data (Baubec et al. 2013). That points to a second possible mechanism of cooperativity that is regulated through the DNA density itself. MeCP2 has been shown to be able to bridge and loop DNA molecules (Georgel et al. 2003; Nikitina et al. 2007). While the apparent compaction of DNA by an MBD peptide from MeCP2 (S. F. Lim et al. 2011) suggests the presence of MBD dimers that are able to form protein bridges

between DNA molecules, surprisingly no cooperativity was observed in the formation of dimers on short substrates (Khrapunov et al. 2014). That may point to the fact that the compaction observed in (S. F. Lim et al. 2011) is driven by the fluorescent label.

Here we show that while both proteins bind with similar affinity to their target sequences, the extension of DNA is very different in both cases. In particular, MBD2 did not change the conformation of DNA molecules, and the extension is insensitive to fluorescent labeling of the MBD2. In contrast, MeCP2 compacted the DNA molecules to less than 50% of their native extension. By using atomic force microscopy (AFM), we show that the probable mechanism for compaction by MeCP2 is the formation of DNA–protein–DNA bridges that create local loop configurations. We thus believe that MBD2 is a superior epigenetic probe to MeCP2, and that full-length proteins can not only be used for labeling reagents but may be performing better than the isolated MBD peptides that have been used in the past.

4.2 Materials and Methods

4.2.1 Nanochannel fabrication

Devices with nanochannels and microchannels were fabricated on fused silica wafers. Nanochannels (180 nm×180 nm) were patterned by electron beam lithography, and microchannels were patterned by optical lithography. We obtained both nanochannels and microchannels by reactive ion etching and sealed the device with a second fused silica substrate by thermal bonding (Riehn et al. 2008). To prevent proteins from sticking to the surface, we used dopamine-mPEG to passivate the interior surfaces of the device (H. Lee et al. 2007). The device was first exposed to 2 mg/ml of dopamine in 10 mM Tris buffer (pH 8.0) for 2 h and then coated

using 100 mg/ml of methoxy polyethylene glycol amine (mPEG-NH2, MW=5000 kDa) in 10 mM Tris buffer (pH 8.0) overnight.

4.2.2 Biological materials

Methylcytosine-free linearized λ-DNA (48.5 kbp) was purchased from Sigma-Aldrich. We used M.SssI (New England Biolabs) to methylate λ-DNA following the supplier protocol. CpG methylation was confirmed by showing the suppression of digestion by HpaII (New England Biolabs) after methylation, as presented in Fig. 4.1A. The methylase was finally removed by purification using a QIAEX II Gel Extraction Kit (Qiagen). In figures throughout this manuscript, DNA without 5-methylcytosine is labeled as "C-DNA", and DNA carrying the maximum density of 5-methylcytosine is labeled as "5mC-DNA".

MBD2FLsc is a full-length human MBD2 single-chain construct comprising MBD2 (amino acids 150–393) and the scMBD2-GATAD2A expressed in E. coli and purified by nickel affinity chromatography (Desai et al. 2015). The as-purified MBD2FLsc is labeled as "MBD2" in the figures. For part of the experiments, MBD2FLsc was labeled using ATTO 565-maleimide (ATTO-TEC) and purified by gel filtration chromatography (Sephadex G-50, GE Healthcare). The labeled MBD2FLsc is labeled "ATTO-MBD2" in the figures. The recombinant human MeCP2 protein produced by the mammalian expression system (Accession P51608) with a 6His tag at the C-terminus was purchased from Novoprotein. The purity of both proteins was confirmed using SDS-PAGE (Fig. 4.1B)

To test the binding affinity of proteins before and after labeling, we performed fluorescence anisotropy measurements using a FAM-labeled dsDNA with one methylated CpG binding site as the substrate (Desai et al. 2015; Pollard 2010). We performed protein titrations with 24 nM of the DNA substrate in a working buffer (HEPES 20mM, NaCl 50 mM, $MgCl_2$ 3 mM,

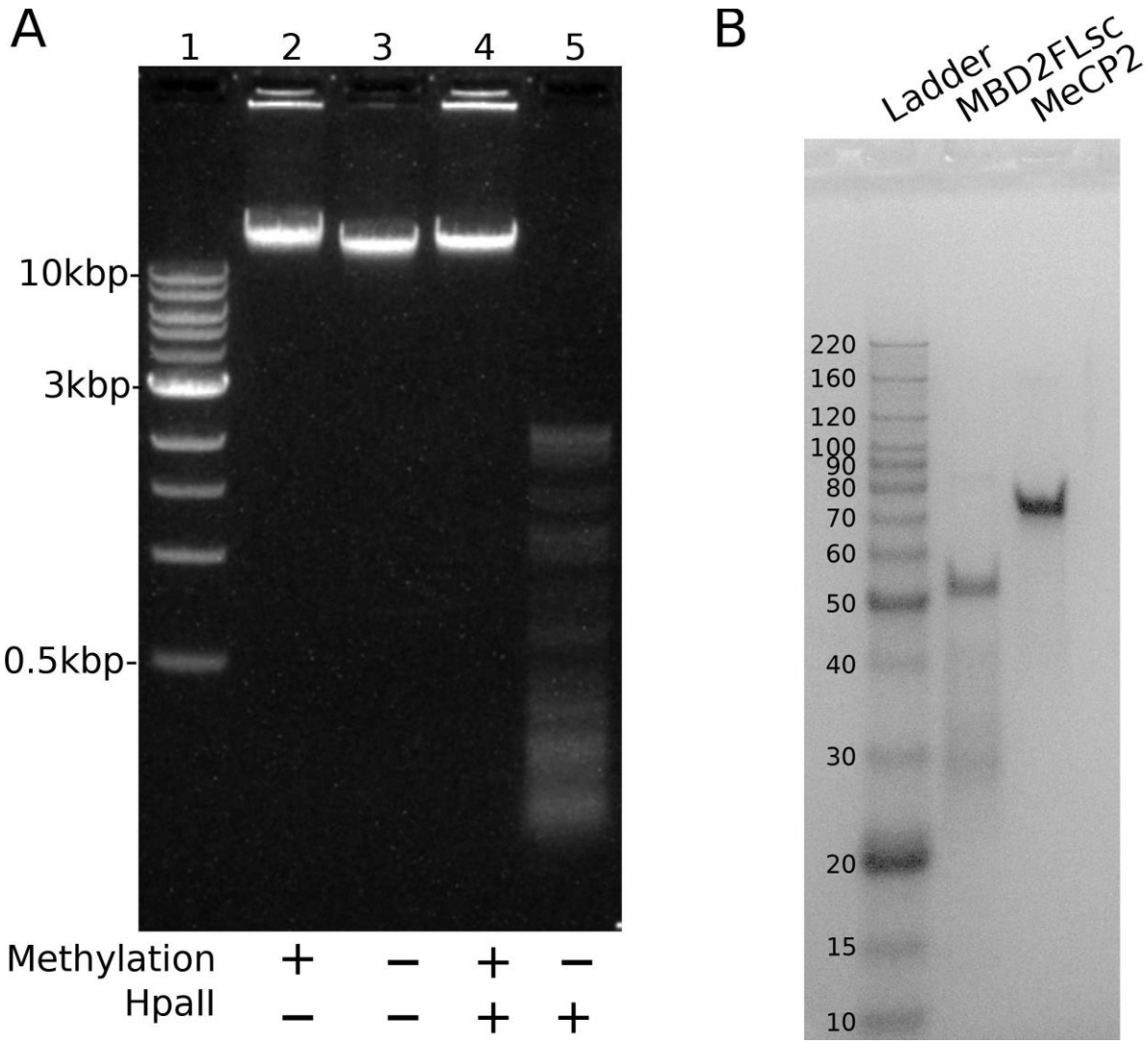

Figure 4.1 Gel electrophoresis of biological samples. Taken from (M. Liu et al. 2020). (a) Agarose gel electrophoresis for verification of cytosine methylation. Lane 1 is a DNA ladder, lane 2 is methylated λ-DNA, and lane 3 is unmethylated λ-DNA. Methylated λ-DNA was not

digested by HpaII (lane 4), while the unmethylated λ-DNA was digested by HpaII (lane 5). (b) SDS-PAGE demonstrating purity of proteins used in experiments.

EDTA 3 mM, pH 7.5). The fluorescence anisotropy was measured using a PTI QuantaMaster 40 (Horiba), and the results were fitted using the Langmuir (single-binding site) model.

4.2.3 Fluorescence microscopy

Both methylated and unmethylated λ-DNA were stained using YOYO-1 (Life Sciences) at a 10:1 molar ratio of base pairs to dye before imaging. To prevent the non-specific binding of the protein, we added mPEG-NH2 to the working buffer with a final concentration of 30 mg/ml. For the MeCP2 experiment, we further added 1% of polyvinylpyrrolidone (PVP).

We first incubated the λ-DNA with the respective protein in our working buffer for 30 min, if applicable. The solution was then introduced into the device using pressurized nitrogen, and DNA was driven into nanochannels using hydrostatic pressure. When a molecule entered, the pressure was removed, which caused molecules to stop in the nanochannel. Imaging was performed on an inverted fluorescence microscope (Nikon TE-2000) with a Nikon 60× oil immersion objective (NA=1.40) in near TIRF mode, and the data were collected by an emCCD (Andor iXon Life) and a QV2 image splitter (Photometrics). For observation of YOYO-1-stained DNA, we used a 488 nm laser for illumination and a 525/40 bandpass filter for collection. For the protein signal (ATTO 565), we used a 561 nm laser and a 600/35 bandpass filter. The data were analyzed using ImageJ and Matlab (Mathworks).

Bare nanochannel-stretched DNA shows a constant ratio between contour length and extension and thus a constant brightness along the molecule (Tegenfeldt, Prinz, Cao, et al. 2004). We have found that this assumption does not hold for DNA strongly compacted by a protein and thus needed to find a different way to characterize the size of a DNA configuration. For each data frame, we calculate the autocorrelation function of the signal intensity $I(x)$ along a DNA molecule

$$C(x) = \frac{\sum_i I(x_i + \delta x) I(x_i)}{l - \delta x} \tag{66}$$

where I is the background-corrected intensity, x_i is the positions of the individual pixels, l is the length of the analysis window, and δx are displacements that are integer multiples of the pixel size. The autocorrelation function can be averaged over all frames to form $\langle C \rangle(\delta x)$. For a molecule with constant intensity along the channel, the autocorrelation function would be a triangle with a full width at half maximum that matches the extension of the molecule. However, we explicitly do not want to make that assumption and instead obtain the length information by calculating the radius of gyration of the autocorrelation function

$$R_g = \sqrt{\frac{\sum_{\delta x} \langle C \rangle(\delta x) \cdot (\delta x)^2}{\sum_{\delta x} \langle C \rangle(\delta x)}} \tag{67}$$

where sums run such that δx sweeps $-l/2 \ldots l/2$, which is chosen about 3 times the average extension of observed molecules. For a uniformly stretched molecule that has an intensity profile that is a Gaussian-widened boxcar function of length L (Riehn et al. 2008), we obtain $R_g = \sqrt{\frac{L^2}{6} + \sigma^2}$. Here σ^2 is the sum of the variance of the point spread function of the microscope objective, and the variance due to the blurring by thermal fluctuations of the system. For large

extended molecules, R_g scales linearly with the extension along the channel axis, while for compacted molecules the σ-term begins to dominate.

The mean extensions for a given condition were found by first determining the R_g for each molecule of the set separately, and then finding the numerical mean and variance of the distribution. From these, the error of the mean was determined, and a Gaussian distribution function was determined that is used in the histogram plots of R_g. The Gaussian assumption is expected to hold well for large molecules (Ödman et al. 2018).

The advantage of this approach is that no assumption about the underlying fluorescence profile along the DNA molecule has to be made, and that averaging of correlations functions is robust because no fitting of any kind is required prior to averaging. Importantly, the more compacted the molecule is, the smaller the R_g value will be. By comparing the R_g result with and without incubation with a protein, we can determine whether a protein can compact the DNA molecule without requiring a detailed understanding of the stretching process.

4.2.4 Atomic force microscopy

For atomic force microscopy (AFM), we used a 7,163-bp linear DNA substrate which contains a 1,697-bp methylated CpG-rich region that is flanked by 2,742-bp and 2,724-bp CpG-free regions (Pan et al. 2017). For MeCP2, the DNA substrate and the protein were diluted in AFM imaging buffer (HEPES 20 mM, $Mg(OAc)_2$ 10 mM, NaCl 100 mM, pH 7.5), mixed together and deposited on freshly peeled mica. For MBD2FLsc, we first mixed the protein and DNA and then diluted the sample in AFM buffer before deposition. The final MeCP2 concentration deposited on mica was 7.5 nM, and the MBD2FLsc concentration was 14 nM. The mica samples were then

washed with filtered deionized water and dried with nitrogen. We used an MFP-3D-Bio AFM from Asylum Research with Pointprobe®PPP-FMR probes (Nanosensors, ≈2.8 N/m) to image the sample at a scan resolution of 5.9 nm and a scan rate of 3 μm/s. The data were analyzed using Asylum Research software and Matlab.

We introduce two quantitative measures for DNA configuration. First, we want to determine whether an object in the AFM image is indeed a single DNA molecule, or whether it is a complex of two or more molecules. To this end, we count the number of DNA ends in an image. A single molecule should have two ends at most. Note that one or two ends are possible if the molecule is randomly deposited in a configuration that includes a loop. We can further introduce a quantitative measure for the "loop count" within the single molecules by counting the number of unoccupied areas that are fully enclosed by DNA contour within a molecule. This numerical measure is a 2-d projection of the 3-d configuration of the molecule prior to deposition, and thus an over-count of the physical number of loops that exist in free solution. However, it is operationally very robust, while determining the number of physical loops that are stabilized by bound protein is considerably slower and carries a higher uncertainty. Both quantitative measures are illustrated in Fig. 4.6, and the definition of the "loop count" is specifically illustrated in Fig. 4.6c.

4.3 Results

Protein modification can lead to a modulation of binding affinities, both between the protein and its substrate as well as between proteins. This is particularly important since we are interested in finding an MBD protein for potential use as an epigenetic labeling agent. The maleimide

chemistry underlying the labeling of MBD2FLsc is prone to disrupt protein configuration. In particular, an attack of protein disulfide bonds may cause the structure of MBD2FLsc to change, thus lowering its binding affinity to DNA. Fluorescence anisotropy titration yielded a K_d of 25±5 nM and 12±4 nM for labeled and unlabeled MBD2FLsc, respectively (Fig. 4.2a). The value for unlabeled MBD2FLsc is consistent with (Desai et al. 2015), and we conclude that a large fraction of the labeled MBD2FLsc population retained its ability to bind to methylated CG sequences.

We observed a significant difference between the extension of methylated and unmethylated DNA, with a larger extension for unmethylated DNA ($p < 0.015$). That is contrary to the finding by Sun et al., who found that methylated DNA is more extended (Sun et al. 2016). However, it is consistent with the decreased persistence length of DNA sequences containing 5-methylcytosine reported by Ngo et al. (Ngo et al. 2016). Another possibility for a decreased extension of methylated DNA compared to unmethylated DNA is a drop in DNA persistence length due to damage during the methylation and purification protocol, which is exacerbated by the relatively large length of our substrate. Note that R_g values for DNA in Fig. 4.3 are consistent with the observed extension of about 6μm according to the argument in the "Materials and methods" section (4.2.3).

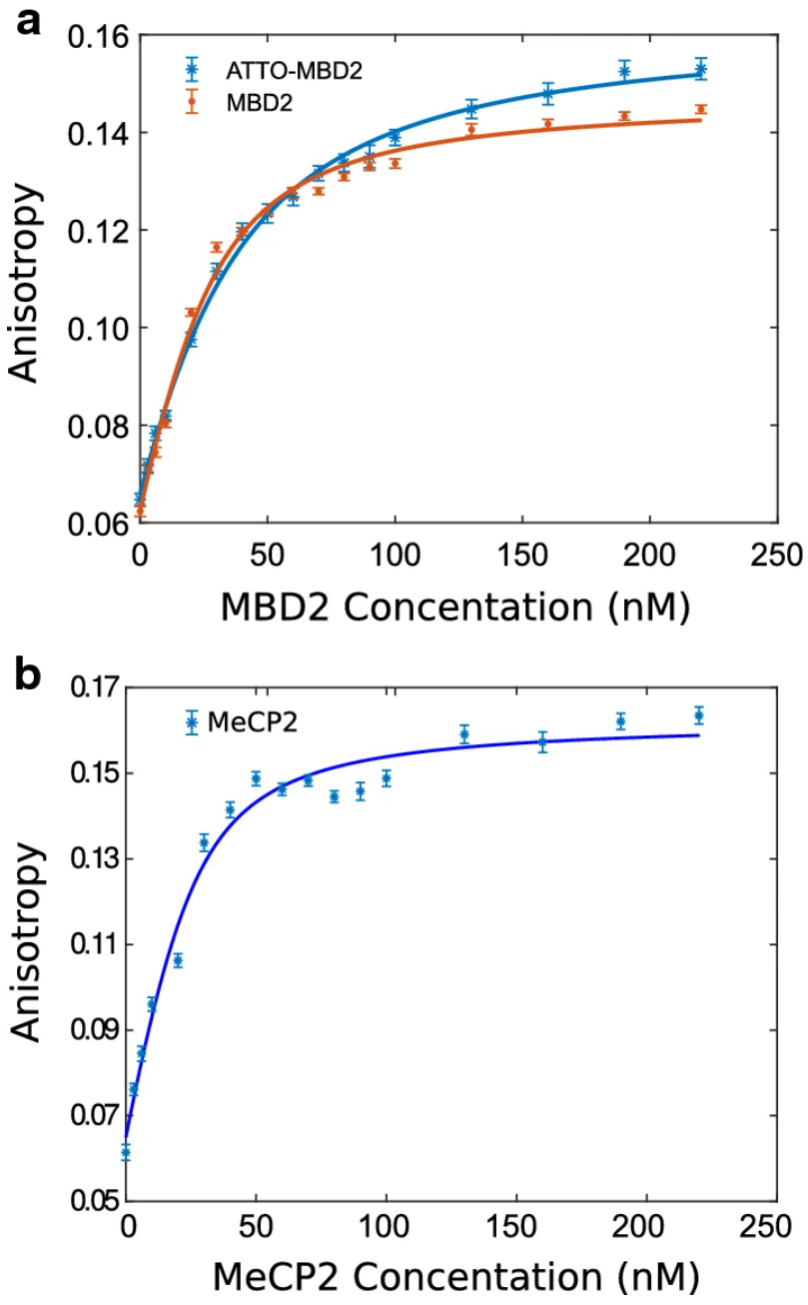

Figure 4.2 Fluorescence anisotropy titration. Taken from (M. Liu et al. 2020). Fluorescence anisotropy titration of protein binding to 24 nM methylated dsDNA substrate. (a) Labeled (blue *) and unlabeled MBD2FLsc (red •) yielded affinities of 24.5±4.6 nM and 11.9±3.8 nM, respectively. (b) MeCP2 yielded an affinity of 7.5±4.3 nM. Error bars are standard deviations over 3 measurements.

The R_g variations between different protein conditions for unmethylated DNA were not statistically relevant (all likelihoods of null-hypothesis p >0.24). Some variations are expected in light of the strong dependence of the extension with nanochannel width (Reisner et al. 2007), our manufacturing precision over a 4-in. wafer, and the variability of the thickness of the channel coating that we apply. For methylated DNA substrates, the R_g for bare DNA and unlabeled MBD2FLsc were not significantly different ($p < 0.43$). However, fluorescently labeled MBD2FLsc lead to an approximately 4% larger R_g ($p< 0.05$), possibly due to a larger effective volume of the labeled protein. Importantly, the effects of smaller extension of methylated DNA substrates (compared to unmethylated) and larger extension of DNA with bound ATTO 565-labeled MBD2FLsc (compared to bare methylated substrates) cancel each other out, so that the difference in extension between bare unmethylated DNA and methylated DNA with bound labeled MBD2FLsc is statistically insignificant ($p < 0.52$).

Table 4.1 Summary of $\boldsymbol{R_g}$ for nanochannel-stretched DNA with MBD2FLsc as determined in Fig. 4.3

Substrate	Bare DNA (μm)	ATTO-MBD2 (μm)	MBD2 (μm)
Methylated DNA	2.68±0.5	2.85±0.07	2.63±0.05
Unmethylated DNA	2.91±0.04	3.03±0.07	2.94±0.05

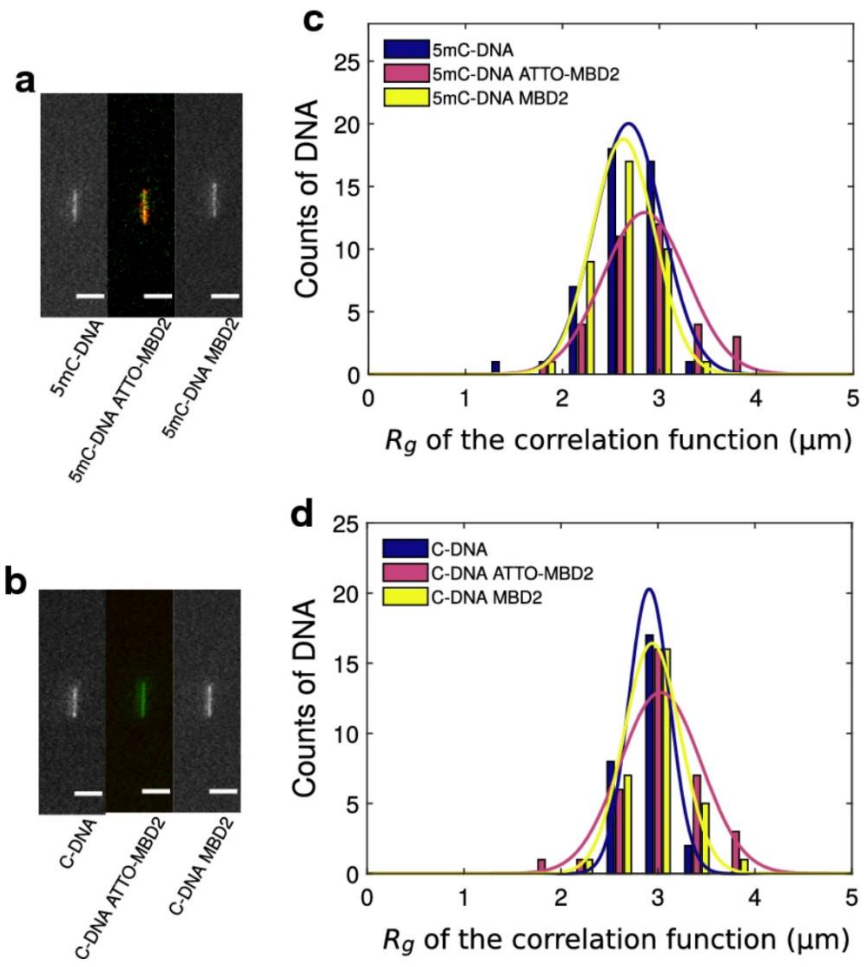

Figure 4.3 Nanochannel stretching of YOYO-1-stained DNA under MBD2FLsc binding. (a) Fluorescence images of methylated DNA (left), methylated DNA with labeled MBD2FLsc (center, ATTO 565-labeled MBD2FLsc in red and methylated λ-DNA in green), and methylated DNA with unlabeled MBD2FLsc (right). (b) Fluorescence images of unmethylated DNA (left), unmethylated DNA with labeled MBD2FLsc (center with ATTO 565 labeled MBD2FLsc in red and methylated λ-DNA in green), and unmethylated DNA with unlabeled MBD2FLsc (right). The scale bar represents 5μm. (c) Histograms of $\boldsymbol{R_g}$ for methylated DNA (blue: bare DNA, N=44, magenta: ATTO 565-MBD2FLsc, N=35, yellow: MBDFLsc, N=38). (d) Histograms of $\boldsymbol{R_g}$ for unmethylated DNA (blue: bare DNA, N=27, magenta: ATTO 565-MBD2FLsc, N=34, yellow: MBDFLsc, N=30).

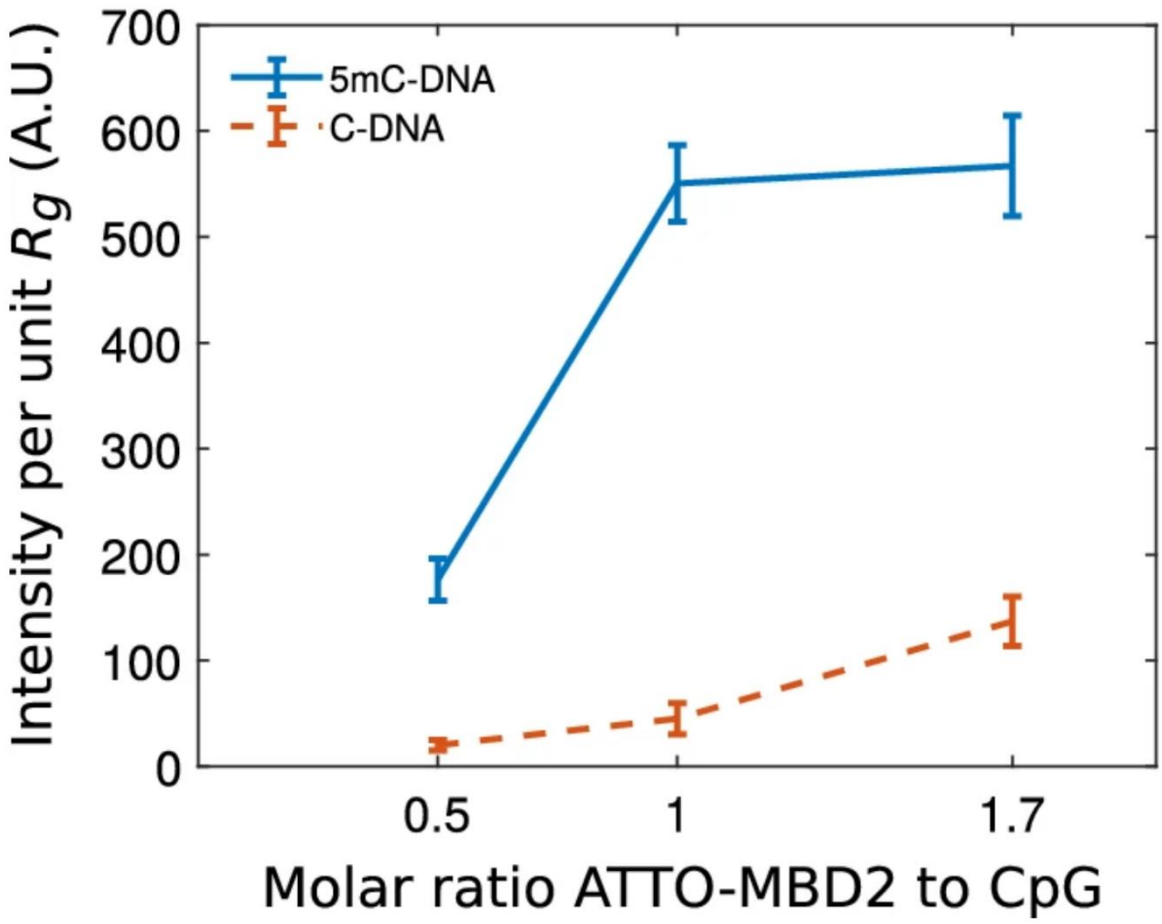

Figure 4.4 Protein fluorescence signal intensity per unit $\boldsymbol{R_g}$ of nanochannel-stretched DNA as function ATTO 565-labeled MBD2FLsc concentration. The Blue (upper) curve is for methylated substrates, and the red (lower) curve is for unmethylated substrates. Error bars represent standard deviations.

Table 4.2 Summary of $\boldsymbol{R_g}$ for nanochannel-stretched DNA with ATTO-565-labeled MBD2FLsc as a function of concentration of MBD2FLsc determined from the dataset underlying Fig. 4.4

Substrate	1:0.5 (μm)	1:1 (μm)	1:1.7 (μm)
Methylated DNA	2.73±0.10	2.85±0.07	2.72±0.11
Unmethylated DNA	3.03±0.06	3.03±0.07	2.85±0.08

The characterization up to this point was performed at a ratio of one MBD2FLsc per target site along DNA. We next characterized whether this result is sensitive to the concentration of ATTO 565-MBD2FLsc. We summarize these results in Table 4.2. We find only a weak dependence of R_g for methylated substrates as a function protein concentration when the ratio [CpG]:[MBD2] is varied from 1:0.5 to 1:1.7, with the lowest likelihood of the zero-hypothesis being p=0.38. For unmethylated substrates, the dependence on protein concentration is stronger, with a mild contraction at [CpG]:[MBD2]=1:1.7. However, the level of significance within the statistics of our experiments is low, with $p < 0.26$ for the hypothesis that the R_g is independent of protein concentration.

The same dataset lets us determine the protein concentration that leads to the highest specificity, which we define as the ratio of bound fluorescent MBD2FLsc bound to a methylated and an unmethylated λ-DNA substrate, respectively. For densely bound proteins in fluorescence imaging, the number of proteins is proportional to the integrated intensity of the fluorescence signal. To account for a possible variation in DNA lengths due to handling, the fluorescence signal was normalized to R_g, which is proportional to the DNA extension in a nanochannel. As anticipated, the specificity is a function of the concentration of the labeling agent since the number of proteins is limited to the number of recognition sites (Fig. 4.4). We observe that the number of bound proteins to the methylated substrate plateaus at parity between protein concentration and concentration of CpG sites. The non-specific binding to the unmethylated substrate does not show a plateau. At the point of parity between MBD2FLsc and binding site concentrations, the specificity reaches approximately a factor of 10.

Turning to MeCP2, we first confirmed that the MeCP2 binding affinity is appropriate for the protein using fluorescence anisotropy (Fig. 4.2b). We find a K_d of 7.5±4.4 nM, which is in agreement with (Khrapunov et al. 2016). For the imaging of MeCP2 on DNA confined to nanochannels, we added 1% of polyvinylpyrrolidone (PVP) inside the working buffer to improve surface passivation (S. F. Lim et al. 2011). Fluorescence images of methylated and unmethylated λ-DNA under nanochannel stretching with MeCP2 show strong compaction of methylated substrates (Fig. 4.5). Quantitative analysis shows that the R_g of unmethylated DNA contracted only very slightly upon exposure with MeCP2 (Table 4.3), with $p < 0.09$. That is in contrast to methylated DNA, which contracted strongly upon MeCP2 exposure ($p<6\times10^{-7}$). We did not attempt to label MeCP2 since the compaction of DNA both is evidence of DNA binding, as well as an undesired feature for any epigenetic profiling application. A similar difference between methylated and unmethylated DNA was observed as in the first data set, with a shorter extension for methylated DNA ($p< 0.0015$).

Table 4.3 Summary of $\boldsymbol{R_g}$ for nanochannel-stretched DNA as determined in Fig. 4.5 for MeCP2

Substrate	Bare DNA (μm)	MeCP2 (μm)
Methylated DNA	2.68±0.03	1.75±0.04
Unmethylated DNA	2.82±0.02	2.74±0.03

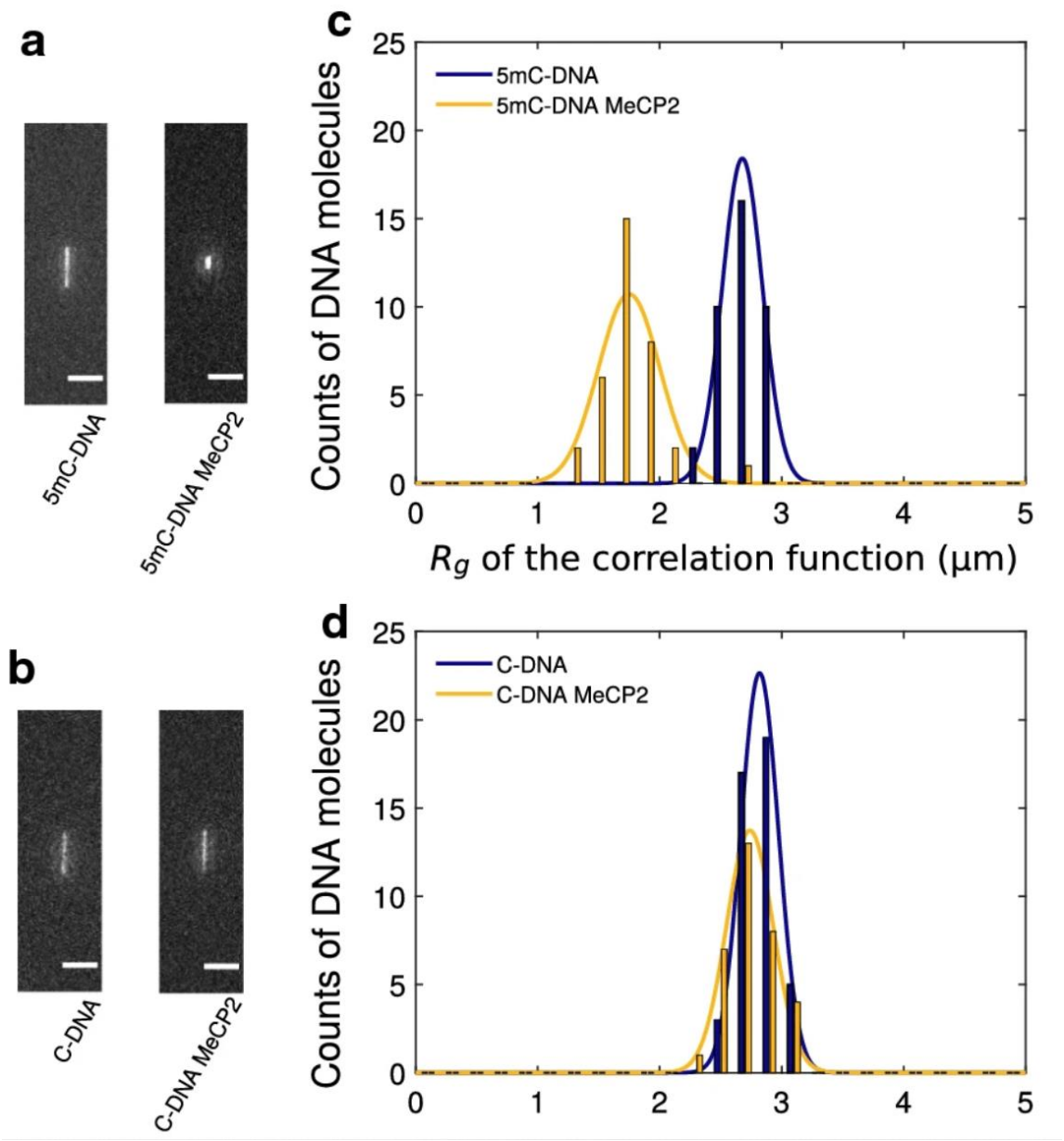

Figure 4.5 Nanochannel stretching of YOYO-1-stained DNA under MeCP2 binding. (a) Fluorescence images of methylated λ-DNA (left) and methylated DNA with MeCP2 (right). (b) Fluorescence images of unmethylated λ-DNA (left) and unmethylated DNA with MeCP2 (right). The scale bar represents 5μm. (c) Impact of MeCP2 binding on $\boldsymbol{R_g}$ for methylated DNA (blue bare DNA N=38, yellow with MeCP2 N=34). (d) Impact of MeCP2 binding on $\boldsymbol{R_g}$ for unmethylated DNA (blue bare DNA N=44, yellow with MeCP2 N=33).

We consider four mechanisms for the compaction of DNA by MeCP2 (Frykholm, Nyberg, and Westerlund 2017). First, the compaction could be due to molecular crowding as reported by Zhang et al. (C. Zhang et al. 2009), but note that this is unlikely in light of our protein concentration. Furthermore, if crowding were the cause, then a similar compaction would have been expected for methylated and unmethylated substrates. The second possibility is a condensation that effectively lowers the contour length through packing into a chromatin-like filament (Streng et al. 2009; Frykholm et al. 2016). The other two mechanisms could be a shortening through a local modulation of physical parameters such as the persistence length and the effective width (the inverse to (Frykholm et al. 2014)), or the formation of localized loops that are stabilized by MeCP2 locking the looped configurations that form as part of thermal fluctuations. To distinguish between the latter three mechanisms, we used atomic force microscopy (AFM) to image the conformations of DNA molecules. In order to obtain configurations that are easy to analyze, we chose a linear DNA as the substrate that carries methylated CpG-rich and CpG-poor regions (Pan et al. 2017). Example images with bare DNA, MBD2FLsc, and MeCP2 are shown in Fig. 4.6a–c. In general, DNA appeared more compact in presence of MeCP2 than MBD2FLsc, with an apparent excess of looped configurations for MeCP2, suggesting the compaction by that protein in nanochannels is driven by the formation of loops. We also observed sporadic large DNA–protein clusters formed in the presence of MeCP2 at high concentrations.

In the "Materials and methods" section (4.2.4), we describe a pathway for quantifying DNA configuration based on ends and open loops. For bare DNA, in excess of 90% of molecules were non-overlapping as indicated by the fraction of DNA with zero, one, and two ends (Fig. 4.6d).

That number decreased marginally for MBD2 ($p < 0.53$) and somewhat for MeCP2 ($p < 0.04$), but in all cases more than 80 % of molecules were isolated. Moving to the count of visible DNA loops for molecules with two or fewer ends (Fig. 4.6e), we find that MBD2FLsc does not alter the conformation of DNA molecules ($p < 0.53$). 98.3±1.2% of bare DNA and 98.3±1.2% of DNA with MBD2FLsc had between 0 and 3 loops. That value dropped to 57.4±4.2% when MeCP2 was added to DNA, which is a significant effect ($p < 7 \times 10^{-14}$). At the same time, 35.2±4.3% of molecules showed 4 to 7 loops, and 7.4±2.4% showed 8 or more loops.

To determine whether the loops are linked to the presence of protein, we determined the apparent height of DNA–DNA crossovers for all three conditions (Fig. 4.6f). The height of both bare DNA and DNA with MBD2FLsc follow an approximately monomodal distribution with mean height and standard deviation of 0.76±0.08 nm and 0.80±0.12 nm, respectively. The difference is at the border of statistical significance ($p < 0.1$). We cannot resolve the exact location of proteins along the DNA, and thus cannot determine whether the small increase is due to protein that is bound in the vicinity of the cross-over on one segment, or to both DNA at the cross-over point.

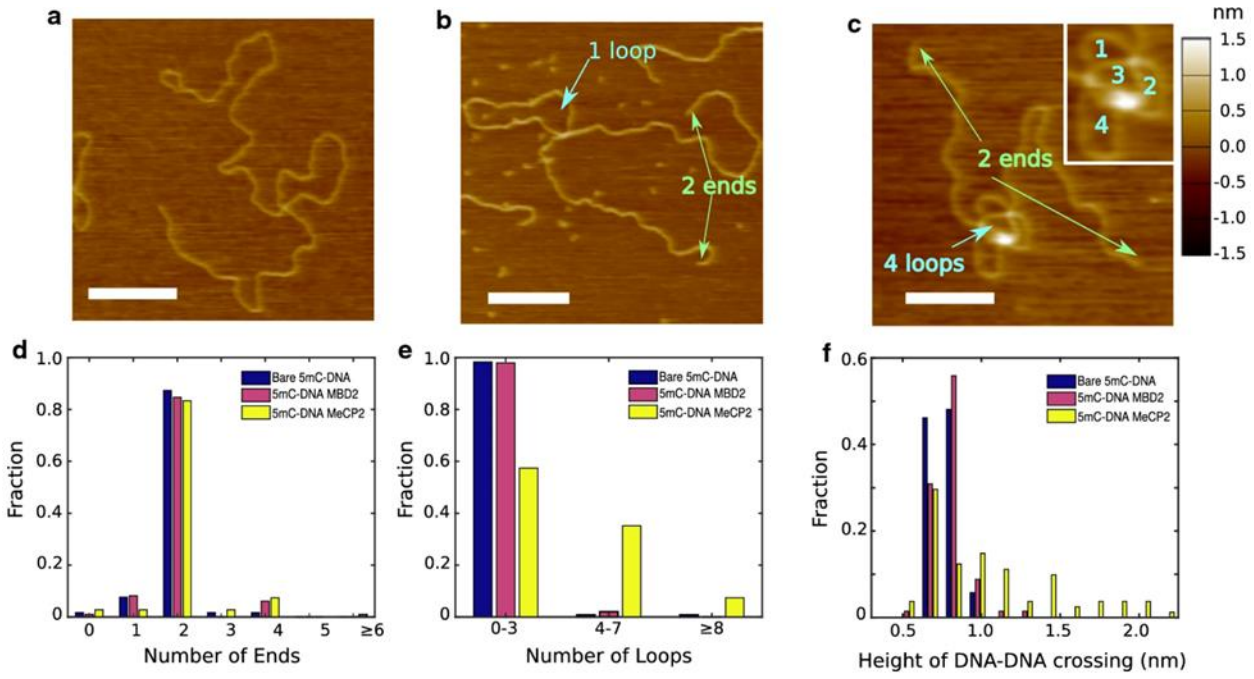

Figure 4.6 Atomic force microscopy (AFM) of methylated substrates under various conditions. AFM of bare methylated dsDNA oligomer a, the same oligomer with MBD2Flsc b, and with MeCP2 c. Scale bars are 200nm. The green arrows point at ends, and cyan arrows point at loops. The inset in d illustrates the counting method for loops. The distribution of number of free ends (d) and the distribution of number of loops (e) for DNA or DNA–protein complexes was determined from such images (bare DNA N=118, MBD2FLsc N=98, MeCP2 N=108). f Height of isolated DNA–DNA crossings (bare DNA N=52, MBD2FLsc N=68, MeCP2 N=83).

The distribution of heights at DNA crossovers in presence of MeCP2 does not follow a monomodal distribution and is rather characterized by the location of apparent protein clusters of varying size at DNA–DNA contact points with a mean cross-over height of 1.12±0.61 nm. Difference to both DNA and MBD2FLsc is statistically significant with $p<8\times10^{-4}$. Interestingly, a subset of cross-overs with MeCP2 displays the same height as for bare DNA, likely due to the

fact that these were not true loop anchoring points before deposition, but rather are random cross-overs caused by deposition onto a 2-dimensional substrate. We further note that in dense configurations, the apparent height could point to a collection of more than two co-localized DNA segments in the probed volume of the AFM tip.

4.4 Discussion and conclusion

We have demonstrated that ATTO 565-labeled MBD2FLsc is a promising candidate for epigenetic mapping applications, since unmethylated DNA without labels and methylated DNA with labels exhibit the same extension. We have further demonstrated specificity of about tenfold on our substrate. Note that the apparent specificity is likely a function of the substrate. Specifically, λ-DNA carries only about 300 CpG site in 48 kbp overall. That means that non-specific sites are considerably more abundant than specific sites. Furthermore, footprint-limited binding is well documented for DNA-binding proteins with multiple binding sites, such that the specific binding sites are less than fully occupied (Wang et al. 2005). This can be reconciled with our finding of maximum occupancy close to equal concentrations of MBD2 and substrate (Fig. 4.4) by noting that not all proteins may be active.

The compaction effect for MeCP2 agrees with prior studies (Georgel et al. 2003; Nikitina et al. 2007), but a number of concluding remarks are warranted. Importantly, the compaction described here does not require interaction with any additional proteins, such as nucleosomes, ribosomes, or similar. We also note the compaction reported by us is the result of DNA methylation. Within the scope of this paper, we are not able to resolve whether the DNA–protein–DNA bridges contain single proteins or are in fact DNA–protein–protein–DNA bridges.

However, we can make the statement that both scenarios lead to a co-operative effect that will enhance the effective binding of MeCP2 to highly methylated sequences.

This can be seen from a thermodynamic argument: Since the closing of a loop requires two helices of DNA, the binding probability of a protein must scale with the square of the DNA concentration. If a randomly formed DNA–DNA contact is stabilized by one or two copies of MeCP2, the recruitment of further MeCP2 becomes thermodynamically more favorable. As more and more proteins are recruited, the DNA configuration will become increasingly denser, thus further enabling the binding of MeCP2. While the interaction of MeCP2 with other nucleoproteins certainly is required for full function, we suggest that the repression of expression of methylated substrates by MeCP2 can be partially explained by this compaction. This is in direct contrast to MBD2FLsc, which does not impact the DNA configuration. Instead, MBD2 requires other proteins of the NuRD complex for the regulation of gene expression (Desai et al. 2015).

In conclusion, ATTO 565-labeled MBD2FLsc appears to be a promising labeling agent specificity targeting methylated CpG sites without changing DNA length under nanochannel stretching. We have found no evidence that MBD2FLsc binding is modulated by binding of proximal MBD2FLsc. On the other hand, MeCP2 shows a strong impact on the organization of DNA and is unlikely to be a good labeling agent.

CHAPTER 5 IMPACT OF SA-1 CONTAINING COMPLEXES ON DNA CONFIGURATION

The following chapter is chiefly based on unpublished manuscript. All data presented here are my work. Collaborators contributed by preparing proteins and establishing baseline conditions for DNA preparation and protein-DNA binding experiments. I performed all data analysis and formed conclusions. The manuscript was written collaboratively.

5.1 Introduction

During eukaryotic cell division, the cohesion of sister chromatids is essential for the correct segregation of chromosomes into two daughter cells. The cohesion of sister chromatids is controlled by a protein complex named cohesin (Peters, Tedeschi, and Schmitz 2008; Davidson and Peters 2021). In addition to its critical role in cell division, cohesin is also involved in 3D chromatin organization and gene regulation (Merkenschlager and Nora 2016). The minimal form of canonical human cohesin consists of four subunits, specifically the cohesin ring subunits Smc1, Smc3, and Scc1 (also known as Rad21), as well as the fourth, subunit Scc3 that either takes the form of SA1 or SA2 (Haering et al. 2002; Gruber, Haering, and Nasmyth 2003). Cohesins containing SA1 or SA2 show different distributions along chromosomes and fulfill different roles in chromatin organization (Kojic et al. 2018). In particular, SA1 is required for telomere cohesion, while SA2 is required for cohesion at centromeres (Canudas and Smith 2009). Researchers recently showed that this functional difference is connected to the different DNA-binding properties, whereas SA1 binds to telomeres while SA2 recognizes intermediate DNA structures during DNA repair and replication (Lin et al. 2016; Countryman et al. 2018). Importantly, cohesion at the telomeres is not mediated by the canonical form of the cohesin ring,

but rather only one of the proteins commonly considered part of cohesin that closely interacts with parts of the shelterin complex (Bisht, Daniloski, and Smith 2013).

The shelterin complex contains six subunits (TRF1, TRF2, TIN2, Rap1, TPP1, and POT1) that assemble along the repeating TTAGGG sequence at the end of chromosomes, and prevent the telomeres from being sensed by the DNA damage response (Verdun and Karlseder 2007; de Lange 2018). Dysfunction of telomeres results in genome instability and can lead to cancer (Maciejowski and de Lange 2017). TRF1 and TRF2 are telomeric DNA binding proteins that can specifically recruit other shelterin proteins, including TIN2, Rap1, and TPP1 to telomeric sequences (Erdel et al. 2017). Deletion of TRF1 makes telomeres fragile and results in replication stalling at telomeres (Sfeir et al. 2009). A large number of proteins and their complex interaction has led to a number of hypothesized models of the complex (Veverka, Janovič, and Hofr 2019; C. J. Lim and Cech 2021), where it is possible that multiple structures exist at different times of the cell cycle, are physiologically significant, and may present future drug-able targets.

TRF1 contains an N-terminal acidic domain, a TRFH domain, and a C-terminal Myb domain that recognizes the telomeric DNA sequence (Erdel et al. 2017). The TRFH domain is one of two recurring binding domains that enable a complex network of possible interactions within the shelterin complex. Different versions of the TRFH domain are found in TRF1, TRF2, and TIN2. TRF1 self-dimerizes with its TRFH domain mediating its TRF1-TRF1 interaction (Griffith, Bianchi, and de Lange 1998; Bianchi 1997; Fairall et al. 2001) but the significance of tetramers observed in single-molecule experiments and molecular dynamics simulations warrants further

exploration (Griffith, Bianchi, and de Lange 1998; Lin et al. 2016; Kalathiya, Padariya, and Baginski 2018). In vitro, TRF1 promotes parallel DNA-DNA pairing and DNA tracts (Griffith, Bianchi, and de Lange 1998; Lin et al. 2016), which is enhanced by TIN2 (S. H. Kim et al. 2003; Pan et al. 2021). While TIN2 does not bind directly to DNA, TIN2 stabilizes TRF1 on telomere DNA (de Lange 2018), and deletion of TIN2 destabilizes the shelterin complex and activates the DNA damage response (Takai et al. 2011). The binding of TIN2 to TRF1 is based on the binding of $TIN2_{TBM}$ to $TRF1_{TRFH}$ (S. Kim, Kaminker, and Campisi 1999). The TBM domain is a second recurring binding domain among shelterin proteins that is also present in TPP1 and TRF2. The stoichiometry and significance of $TRF1_{TRFH}$ to $TIN2_{TBM}$ poses some difficulty, in that crystal structures indicated a 1:1 stoichiometry (Chen et al. 2008) that would suggest a destabilization of $TRF1_{TRFH}$ dimerization. On the other hand, DNA-DNA binding assays indicate that only a single TIN2 is involved in binding to a TRF1 dimer (S. H. Kim et al. 2003). Computational studies showed that $TIN2_{TBM}$ binding to $TRF1_{TRFH}$ leads to an increase in the tetramerization (by enabling the coordination of two $TRF1_{TRFH}$ dimers) (Kalathiya, Padariya, and Baginski 2018). However, binding experiments demonstrated that full-length TIN2 is required for efficient TRF1-based coordination of independent DNA molecules, apparently indicating that the TRFH domain at the N-terminus of TIN2 plays some role as well (S. H. Kim et al. 2003).

The addition of SA1 to TRF1 and TIN2 leads to an even more complex picture since SA1 binds to telomeric DNA as well as both TRF1 and TIN2 (Canudas et al. 2007). The DNA-binding AT-hook domain at the N-terminus of SA1 provides specificity for the telomeric sequence (Bisht, Daniloski, and Smith 2013). The N-terminus is also the point where TRF1 and SA1 bind, while the binding of SA1 to TIN2 is towards its C-terminus and the N-terminus of TIN2. When

compared to SA2, the specificity of SA1 to telomeres arises both from the binding affinity of SA1 to the telomeric repeat sequence, as well as its interaction with TIN2 and TRF1 that are not shared with SA2 (Canudas et al. 2007). siRNA depletion and chromatin immunoprecipitation *in vivo* indicate that telomere cohesion in the presence of SA1 arises from interaction with the shelterin proteins TRF1 and TIN2 (Bisht, Daniloski, and Smith 2013). The structure of neither the TRF1-SA1 nor the TIN2-SA1 complexes has been resolved.

Since sister telomere cohesion is largely independent of the cohesin ring (Bisht, Daniloski, and Smith 2013), we hypothesized that SA1, TRF1, and TIN2 are sufficient to induce DNA-DNA pairing through the establishment of protein bridges. *In vitro* measurements to quantify such protein-induced DNA-DNA pairing can either observe conformation changes in a single DNA molecule (cis-bridging), i.e., compaction, or pairing of independent molecules (trans-bridging), i.e., DNA aggregation. Possible bridging mechanisms by SA1, TRF1, and TIN2 are illustrated in Fig. 5.1. Researchers have found that SA1 interacts with TRF1 and enhances TRF1-mediated telomeric DNA-DNA pairing *in vitro*, as illustrated in Fig. 5.1 **B** and **C** (Lin et al. 2016). Recently we showed that TIN2 enhances the bridging and compaction mediated by TRF (Pan et al. 2021) (Fig. 5.1 **D**). The present paper explores the ternary system combining all three proteins (Fig. 5.1 **E** and **F**), and further investigates whether the final DNA configuration is dependent on the sequence of protein assembly on DNA or not.

The present study extensively utilizes DNA extension under nanofluidic stretching (Tegenfeldt, Prinz, Cao, et al. 2004) to find a quantitative, time-resolved measure of cis-bridging under time-dependent action of a combination of proteins. The technique avoids the shortfalls of other

possible techniques. For instance, it can utilize long substrates (10s of kb) while avoiding the need to tether the molecule and apply axial stress that is common to other time-resolved single-molecule methods, like magnetic and optical tweezers (Heller et al. 2014) or DNA curtains (Collins et al. 2014). At the same time, it yields time-dependent information which is not obtained in typical atomic force microscopy (AFM) imaging in air. In nanofluidic devices, DNA molecules stretch due to a balance of self-avoidance and entropy when confined by channels with a diameter of a few DNA persistence lengths (Reisner, Pedersen, and Austin 2012). DNA extension is changed under protein action when the contour length changes (Streng et al. 2009), the entropy is changed through a crowding environment (Guttula et al. 2018; C. Zhang et al. 2009), the persistence length of DNA is changed (Frykholm et al. 2014), or DNA bridges are established (Öz et al. 2020). The study of protein-based conformation changes has benefited from design innovations, importantly the introduction of grooves recessed into nanoslits (Mikkelsen et al. 2011) that then enable real-time observation of the response to buffer exchange (Sharma et al. 2020). We have further refined that concept (Saroj Dangi unpublished result) and use it here to investigate the interactions between DNA and three proteins, namely SA1, TRF1, and TIN2. The device enables us to introduce proteins into nanochannels sequentially, and thus identify not only aggregate effects but also binding sequences. We test the relevance of the cis-bridging results obtained from nanofluidics with trans-bridging results obtained from AFM to verify their findings in the context of bridging multiple molecules as is the case in sister chromatid cohesion.

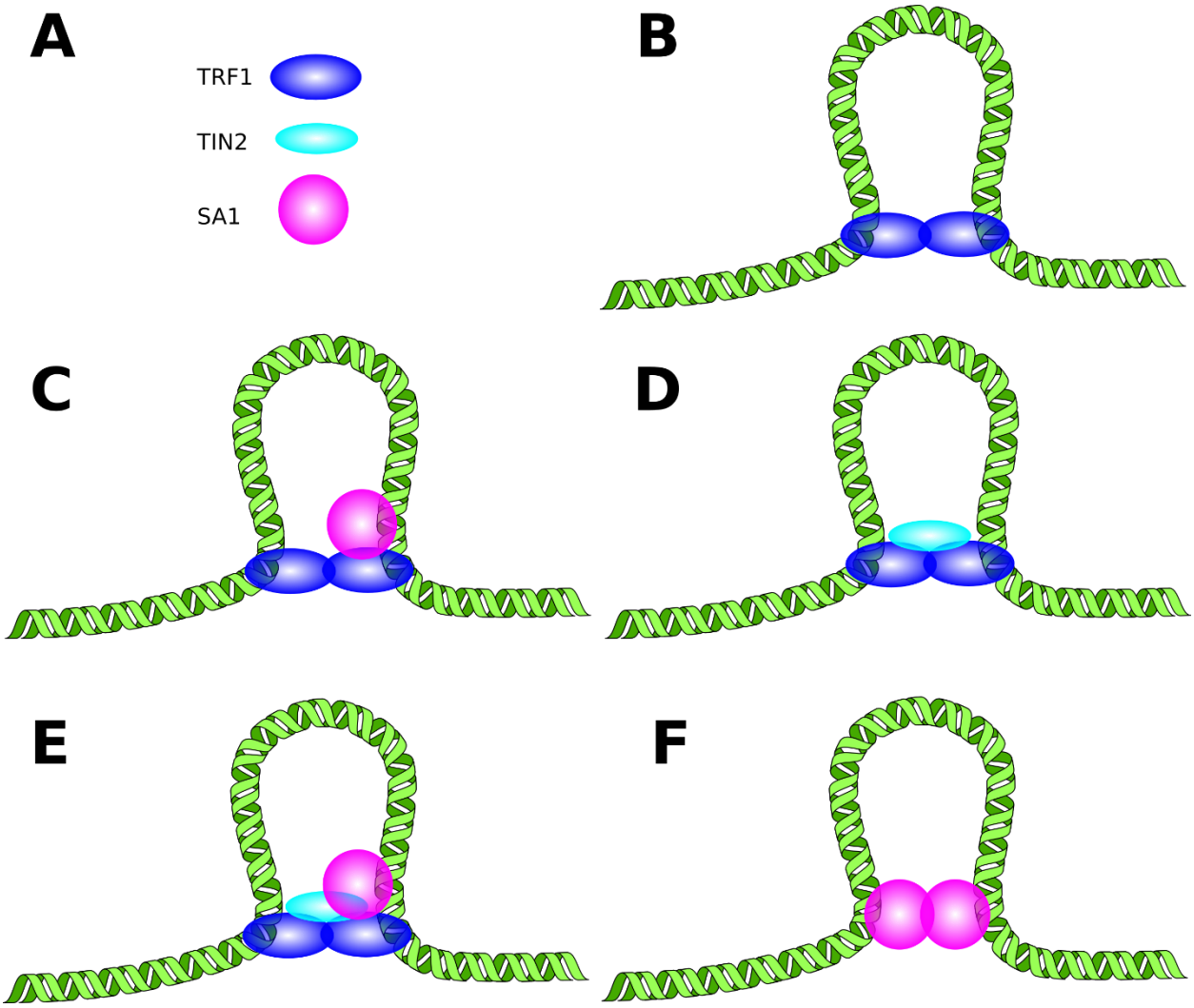

Figure 5.1 Models of protein-stabilized loops in telomeres. (A) TRF1 is indicated blue, TIN2 is cyan, and SA1 is red. (B) TRF1 can stabilize DNA loops and stabilize tracks of parallel DNA. (C) SA1 enhances the stabilization of loops and track formation due to TRF1 while binding directly to DNA. (D) TIN2 enhances the stabilization of loops due to TRF1 without binding directly to DNA. (E) Suggested structure of a complex incorporating TRF1, TIN2, and SA1. (F) Hypothetical structure of SA1 forming dimers while bound to distant loci of the same DNA.

5.2 Methods

5.2.1 Combinatory nanofluidic device fabrication

An in-depth discussion of the device and its operation are currently under review in another publication (Saroj Dangi et al. unpublished results) (Fig. 5.2 **A**). Our combinatory devices were fabricated on fused silica wafers. Nanogrooves (150 nm × 150 nm) were patterned by electron beam lithography into the floor of a nanoslit region (40 nm depth) that was patterned by optical lithography, both followed by reactive ion etching to transfer the patterns. This reaction region is linked to 1 µm-deep microchannels that allow interfacing to pressurized nitrogen infrastructure. We obtained a closed fluidic system by sealing the device with a second fused silica substrate (Riehn et al. 2008).

For surface passivation, we used dopamine-mPEG to passivate the interior surfaces of the device (H. Lee et al. 2007; M. Liu et al. 2020). The device was coated by first exposing to 2 mg/mL of dopamine in 10 mM HEPES buffer (pH 8.0) for 2 hr and then coated using 100 mg/mL of methoxy polyethylene glycol amine (mPEG-NH_2, MW=5,000 kDa) in 10 mM HEPES buffer (pH 8.0) overnight.

5.2.2 Biological materials

The λ-DNA substrate (48.5 kbp) was purchased from Promega. The T270 plasmid (5.4 kb), which contains two telomeric sequence regions $(TTAGGG)_{135}$ linked by 23-bp non-telomeric sequence, was purchased from Addgene (PSXneo(T2AG3), #12403) (Hanish, Yanowitz, and de Lange 1994). The T270 plasmid was first digested by BglII and XbaI for 4 hr at 37 ℃. The telomeric region (pT270) was separated by agarose gel electrophoresis and purified by a gel

purification kit (Qiagen). Then the pT270 was ligated (LpT270) using the Quick Ligation Kit (New England Biolabs) and purified using phenol-chloroform extraction.

N-terminal His_6-tagged SA1 and TRF1, and N-terminal HA-tagged TIN2 (1-354 aa) were expressed in insect cells using the pFastBac system, as described previously (Lin et al. 2014, 2016; Pan et al. 2021). The protein concentrations were determined by the Bradford assay.

5.2.3 Nanofluidic Experiments

Both the telomeric DNA and λ-DNA substrates were stained using YOYO-1 at a 1:20 (dye:base pairs) molar ratio. All nanofluidic experiments were conducted under identical buffer conditions (20 mM HEPES, 100 mM KCl, 0.1 mM $MgCl_2$, pH 7.5). YOYO-1-stained DNA was excited by a 488 nm laser. The fluorescence imaging was performed on an inverted fluorescence microscope (Nikon TE-2000) with a Nikon 100× Apo TIRF oil immersion objective (1.49 N.A.) in total internal reflection mode. The data were obtained by using an emCCD camera (Andor iXon Life) and a QV2 image splitter (Photometrics).

The experimental procedure is illustrated in Fig. 5.2 **B**. During the experiment, ligated telomeric DNA molecules were brought to the reaction region by the application of ≃2 bar of nitrogen gas to the ports of the DNA-containing microchannel. Upon removing the driving pressure, molecules became trapped inside nano grooves due to local minima of the free energy (Mikkelsen et al. 2011). The configuration of the molecules was observed for 10 s under a no-flow condition. We then drove protein-containing buffer into the shunt region at 100 mbar under constant flow to incubate the DNA for 5 min. In experiments investigating the action of multiple

proteins as a function, this was followed by 5 min of flow from the second protein-containing buffer under equal conditions. Note that the molecules were transiently deformed by the applied pressure since the flow field is not fully perpendicular to the axis of all nano grooves. We then removed the flow, let the molecules equilibrate for 100 s, and finally recorded molecules for 10 s under a no-flow condition.

By comparing the DNA length before and after protein incubation, we obtain the relative length change of a telomeric DNA substrate. YOYO-1-stained bare DNA stretched in nano grooves usually shows a constant brightness along the DNA. In experiments that expose DNA to DNA-binding proteins, the DNA molecules are generally compacted by proteins inhomogeneously along the molecule backbone, which makes it hard to fit a Gaussian-widened boxcar function as is customary in the field (Riehn et al. 2008). Our analysis of DNA molecules presented here is based on the radius of gyration (R_g) of the correlation function. In reference (M. Liu et al. 2020) we have detailed the computation of the radius of gyration (R_g) of the correlation function. For long DNA molecules with homogenous stretching, R_g is proportional to the extension of the DNA molecule. Here we use the relative length change of DNA with and without proteins as our indicator. The relative length change of DNA was defined as

$$\frac{\delta L}{L} = \frac{L_0 - L_{proteins}}{L_0} \tag{68}$$

For statistical analysis, the data was plotted in box-whisker plots with indications for the median and the 25-75 % range. Pair-wise p-values were determined by Wilcoxon signed-rank test unless indicated otherwise. In particular, we used the Kruskal-Wallis and Tukey's range test if multiple similar conditions were compared as a group. Errors in the text are given by twice the

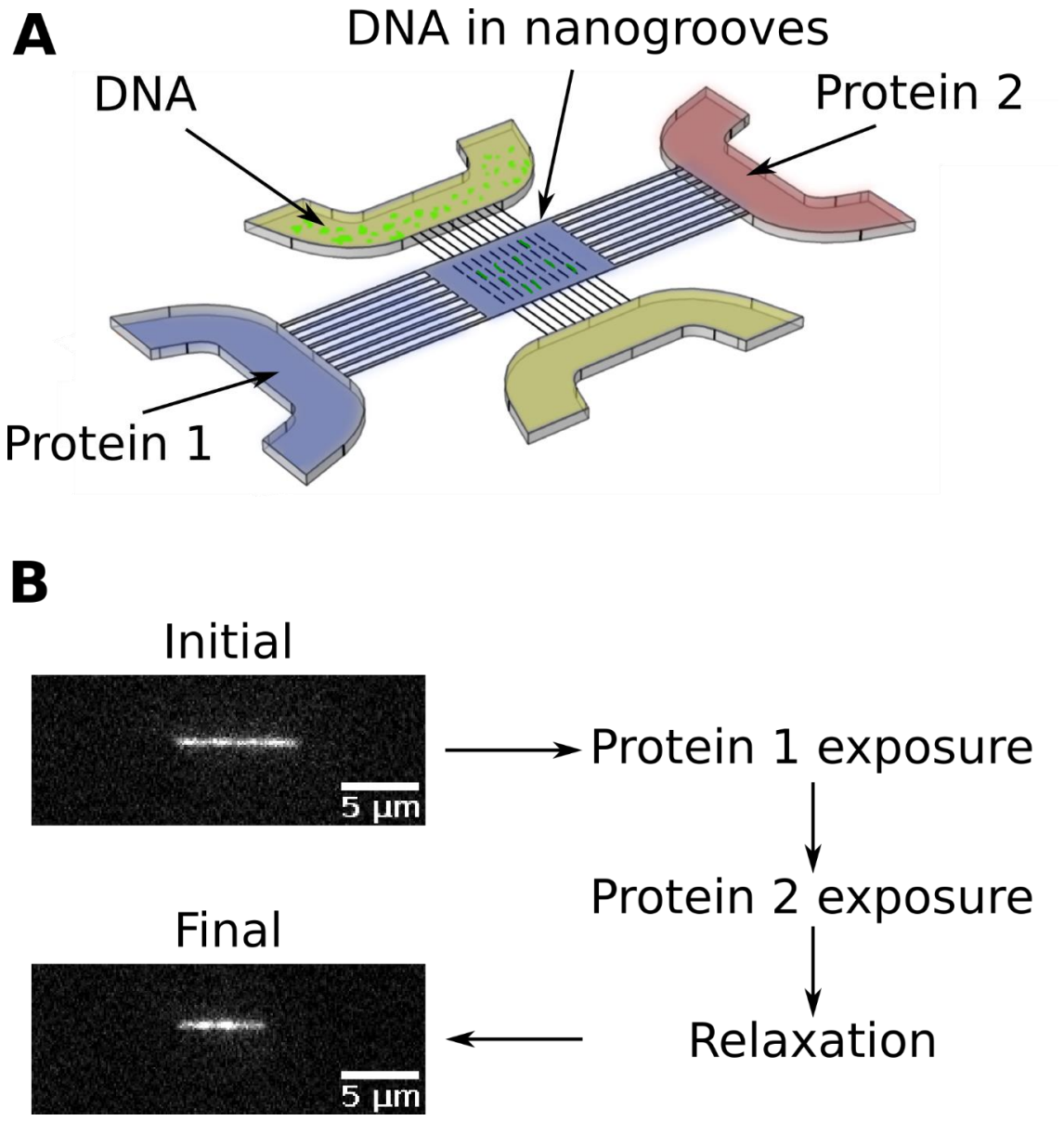

Figure 5.2 Schematic of device and principle of use. (A) Shows the device which holds stretched DNA inside the nano grooves of a central reaction region. Four double-ended microchannels supply DNA, two alternate protein solutions, or a blank buffer for rinsing, respectively. (B) illustrates the experimental procedure starting with the observation of the molecule before protein exposure (labeled "Initial"), followed by exposure to proteins (here TRF1+TIN2 first, followed by SA1) for 5 minutes each, followed by relaxation, and post-exposure observation (labeled "Final").

median absolute deviation, which, depending on the assumed underlying distribution, is on the order of the 95 % confidence interval.

5.2.4 Atomic force microscopy

We used the pRST5 DNA (3513 bp, a gift from Jack Griffth at UNC-Chapel Hill), which contains 96 TTAGGG repeats (Kar, Willcox, and Griffith 2016). We linearized pRST5 plasmid using SacI. We set three comparison groups with 25 nM proteins and linear pRST5 DNA (6 nM): (1) DNA incubated with SA1, TRF1, and TIN2; (2) DNA incubated with TIN2 and TRF1; (3) DNA incubated with TRF1. All incubations were done in a working buffer (20 mM HEPES, 100 mM KCl, 0.1 mM $MgCl_2$, pH 7.5) at room temperature for 5 min. The mixture of DNA and proteins were diluted in AFM buffer (20 mM HEPES, 10 mM $Mg(OAc)_2$, 100 mM NaCl, pH 7.5) and then deposited onto freshly peeled mica. The mica samples were washed with filtered deionized water and then dried with nitrogen gas. We used an MFP-3D-Bio AFM from Asylum Research and Pointprobe® PPP-FMR probes (Nanosensors) to image the sample at a scan resolution of 512 × 512 and a scan rate of ~1 Hz. Data were then analyzed using Asylum Research software and Matlab.

Since different depositions may cause different DNA densities on the mica surface, we choose to use the ratio between free DNA and DNA in protein-DNA clusters as our indicator for the efficiency of DNA-DNA bridging. With large clusters, it is hard to distinguish individual DNA inside. Instead, we use the AFM volumes of clusters divided by the number of free DNA. The data were then plotted using the Box-Whisker plots with the median and the 25-75 %. The p-valves were determined by Wilcoxon signed-rank test.

5.3 Results

5.3.1 TIN2 facilitates TRF1-mediated telomeric DNA compaction through cooperative interactions

The interaction between TRF1 and TIN2 is well-documented (S. Kim, Kaminker, and Campisi 1999), with an affinity of 310 nM (Chen et al. 2008). We have shown in prior publications that TIN2 enhances TRF1-mediated telomeric DNA compaction (Pan et al. 2021). In particular, we found that TRF1 and TIN2 together efficiently compact individual telomeric DNA molecules, as well as induce large clusters of multiple DNA strands with coordinated proteins. We attributed this to the bridging of segments of DNA by the protein complex. While it is well-established that TIN2 stabilizes TRF1 dimers, it is currently not known whether TIN2 can invade previously formed TRF1 dimers that are bound to unbridged DNA, and whether this possible invasion then leads to an increase in DNA-DNA bridging.

Before investigating the impact of the order of protein exposure on DNA compaction inside nanochannels, we first characterized how TRF1 and TIN2 alone change DNA extension under nanofluidic stretching (Fig. 5.3 **A**). At a protein concentration of 450 nM, we did not observe a statistically significant change in DNA extension after exposure to either TRF1 or TIN2 individually. In contrast, the median relative DNA length changed by -12 $\pm$ 2% when telomeric DNA was incubated with 450~nM of the TRF1-TIN2 mixture. This reduction is significant in the Tukey test spanning the experiments in Fig. 5.3 **A** ($p<8 \times 10^{-4}$).

Since the affinity of TIN2-TRF1 binding (310 nM (Chen et al. 2008)) as well as TRF1-DNA binding (200 nM (Hanaoka, Nagadoi, and Nishimura 2009) are both roughly half the above-

chosen protein concentration, we investigated whether a doubling of the protein concentration to 900 nM, and thus raising the occupancy of protein on the DNA, could lead to a stronger compaction result (Fig. 5.3 **A** last column). While we did observe a median relative length change of -17 ± 3% at 900 nM, versus -12 ± 2% for 450 nM, the difference between the experiments with 450 nM and 900 nM of both TRF1 and TIN2 only carried a significance of $p<0.1$. Thus, we cannot exclude the null hypothesis that the difference in protein concentration was irrelevant. The small additional compaction upon doubling of the protein concentration suggests that a large fraction of TRF1-TIN2 binding sites on DNA is occupied at 450 nM.

To determine whether DNA compaction is driven by the binding of TRF1-TIN2-TRF1 complexes that are present in the free pool, or through the capture of TIN2 by DNA-bound TRF1-TRF1, we quantify the DNA extension under sequential exposure to TRF1 and TIN2. In particular, we test whether the two possible exposure sequences (TRF1 first/TIN2 second or TIN2 first/TRF1 second) are equivalent, and whether either of those sequences yields a compaction that is equal to the compaction obtained using premixed TRF1 and TIN2 (Fig. 5.3 **B**). We used a protein concentration of 900 nM to utilize the small increase in DNA compaction mentioned above. The median telomeric DNA compaction after incubation with TIN2 followed by TRF1, with TRF1 followed by TIN2, and with the TRF1-TIN2 mixture were -6 ± 3%, -16 ± 4%, and -17 ± 3%, respectively. A Tukey test shows that the difference between the sequential addition of TRF1/TIN2 and the TRF1-TIN2 mixture is not significant ($p<0.65$), while the TIN2/TRF1 addition order gave significantly lower compaction ($p<0.03$) than the TRF1-TIN2 mixture. Note that this mild DNA compaction under the sequential addition of TIN2/TRF1 does not necessarily imply a compaction by TRF1-TIN2 complexes under this condition but may

occur simply due to DNA compaction by TRF1 alone that is higher in Fig. 5.3 **B** than in Fig. 5.3 **A** due to the increased protein concentration. The sequential exposure experiments suggest a DNA compaction mechanism through TIN2 interactions with TRF1 dimers that are already bound to DNA (C. J. Lim et al. 2017).

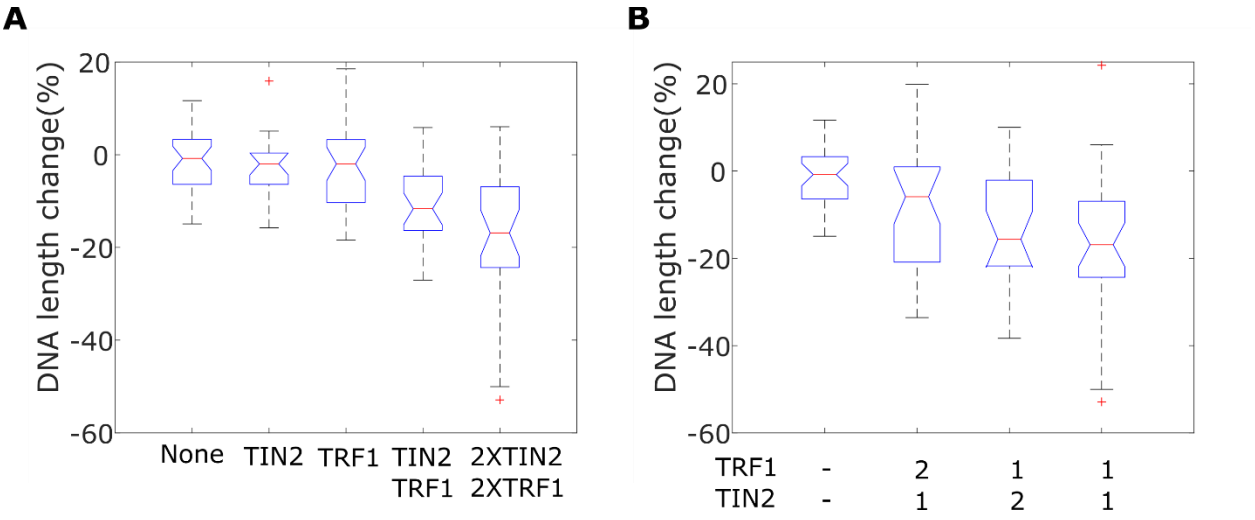

Figure 5.3 Testing the relative length change of ligated pT270 DNA under exposure to TRF1 and TIN2 proteins. (A)Shows single-step 5-minute exposure experiments with (1) control with protein-free buffer (N=35), (2) 450 nM TIN2 (N=20), (3) 450 nM TRF1 (N=35), (4) mixture of 450 nM TIN2 and 450 nM TRF1 (N=29), and (5) mixture of 900 nM TIN2 and 900 nM TRF1 (N=30). (B) Investigates dependence of the compaction on the exposure sequence with (1) control with buffer (N=35), (2) first 900 nM TIN2 and then 900 nM TRF1 (N=30), (3) first 900 nM TRF1, and then 900 nM TIN2 (N=23), and (4) pre-mixed 900 nM TIN2 and TRF1 each (N=30). The Box-Whisker plots show the median, the 25-75 percentiles, and the range.

5.3.2 SA1 drives DNA compaction independent of TIN2 presence

In the section below, we will consider combinations of TRF1-TIN2 on the one hand and SA1 on the other hand and their activity in the compaction of DNA. The interacting system is complex, with six possible pairwise interactions, four three-way interactions, and the potential for more complicated allosteric pathways. While some of the interactions were explored above (DNA-TRF1, DNA-TIN2, DNA-TRF1-TIN2 (Pan et al. 2021)), and others have been considered in the literature (TRF1-TIN2 (Chen et al. 2008), TRF1-SA1, SA1-TIN2, SA1-TRF1-TIN2 (Canudas et al. 2007), DNA-TRF1-SA1 (Lin et al. 2016)), the DNA-SA1 and DNA-SA1-TIN2 interactions are less explored. In this section, we directly compare the DNA compaction by SA1 and TIN2 versus SA1 alone.

In Fig. 5.4 **A**, we present a comparison of compaction of the ligated pT270 DNA substrate in the presence of 450 nM SA1, 900 nM SA1, or 450 nM SA1 combined with 450 nM TIN2. All SA1-containing conditions drive compaction with $p<0.004$ shown in Fig. 5.4 **A** compared to the control, with a median compaction of $-5 \pm 1\%$. The individual pair-wise significance tests of SA1-containing conditions relative to the control also yielded $p<0.05$. However, no significant differences among the three conditions containing SA1 were found, indicating that SA1 mediates DNA compaction independent of TIN2.

The finding of DNA compaction by SA1 independent of TRF1 is novel in as far as it was not found in a prior study at lower SA1 concentration (30 nM) (Lin et al. 2016). The lack of difference between the 450 nM and 900 nM conditions in Fig. 5.4 **A** indicated the saturation of

SA1 binding sites along the telomeric DNA, and thus we cannot determine with the current data what the minimal required concentration of SA1 for DNA compaction is.

5.3.3 Compaction by TRF1, TIN2, and SA1 are additive, and facilitated through separate DNA-binding events

In a prior exploration of the impact of TRF1 and SA1 on telomeric DNA, we found an enhancement of the DNA-DNA pairing when compared to TRF1 alone (Lin et al. 2016). There we hypothesized that the interaction stems from the formation of SA1-TRF1 complexes (of unknown stoichiometry) from initially independent DNA-bound SA1 and TRF1. We were not able to determine, however, whether this model of independent binding is consistent with the experiment since premixed proteins were applied to the DNA. Here we answer that question while expanding the system to include TRF1, TIN2, and SA1 because the three-protein system has been shown necessary for the cohesion of the sister telomeres (Canudas et al. 2007). For binding sequence experiments, we consider TRF1-TIN2 as one compacting agent and SA1 as two the second due to the lack of impact by TIN2 on SA1-mediated compaction above.

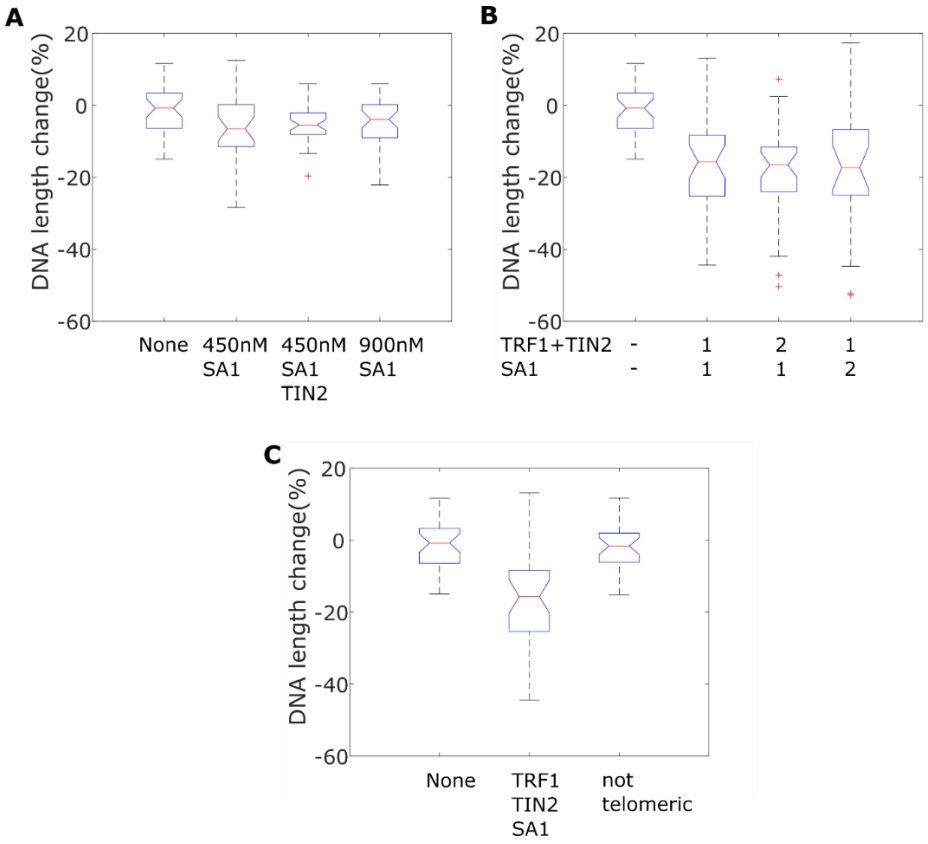

Figure 5.4 Testing the relative DNA length change under exposure to TRF1, TIN2, and SA1. (A) Demonstration that SA1 compacts ligated pT270 DNA, and that that compaction is independent of TIN2. Conditions are (1) control with buffer (N=35), (2) 450 nM SA1 (N=29), (3) 450 nM SA1 and TIN2 premixed (N=38), and (4) 900 nM SA1 (N=36). (B) Tests whether the exposure sequence of TRF1-TIN2 and SA1 impact the compaction of ligated pT270 DNA. (1) control with buffer (N=35), (2) mixture of 450 nM SA1, TIN2 and TRF1 (N=33), (3) first 450 nM SA1 and then 450 nM TRF1-TIN2 mixture (N=33) (3) and first 450 nM TRF1-TIN2 mixture and then 450 nM SA1 (N=23). (C) Demonstration that compaction by the combination of SA1, TRF1, and TIN2 is telomere specific. (1) Ligated pT270 exposed to protein-free buffer (N=35), (2) Ligated pT270 exposed to premixed 450 nM SA1, TIN2, and TRF1 each (N=33), and (3) λ-DNA exposed to premixed 450 nM SA1, TIN2 and TRF1 each (N=29). The Box-Whisker plots show the median, the 25-75 percentiles, and the range.

In Fig. 5.4 **B**, we present the compaction of telomeric DNA under exposure of SA1 first/ TRF1-TIN2 second, TRF1-TIN2 first/ SA1 second, and a premixed combination of SA1 and TRF1-TIN2. The concentration for each protein was 450 nM, which was chosen to prevent runaway compaction and clogging of the device. We observe a DNA length change of -16 $\pm$ 3% when all three proteins are added simultaneously, which is significantly greater than both the 450 nM TIN2-TRF1 mixture from Fig. 5.4 **A** ($p=0.05$), as well as the 450 nM SA1 dataset from Fig. 5.4 **A** ($p<2 \times 10^{-4}$). Meanwhile, we did not observe a significant difference among the three conditions with all three proteins (Kruskal-Wallis $p=0.84$). In particular, we find that the median of the DNA length change for telomeric DNA exposed to SA1 followed by TRF1-TIN2, TRF1-TIN2 followed by SA1, and simultaneous exposure to all three proteins were -17 $\pm$ 2%, -17 $\pm$ 4%, -16 $\pm$ 3%, respectively. The insensitivity to the exposure sequence directly implies that TRF1-TIN2 and SA1 undergo independent DNA binding events.

To strengthen the validity of our assumption that loop formation through bridging of distant DNA segments is the cause of the compaction, and establish the relevance for telomere cohesion, we quantified the formation of clusters of short DNA substrates (pRST5) using AFM (Fig. 5.5). We observed a significant change in the number of DNA in clusters after incubating with SA1, TIN2, and TRF1 for 5 min, comparing with incubating with TRF1 and TIN2 ($p<2 \times 10^{-4}$). Clustering of DNA without confinement strongly implies the formation of protein-mediated bridges between fully independent DNA segments.

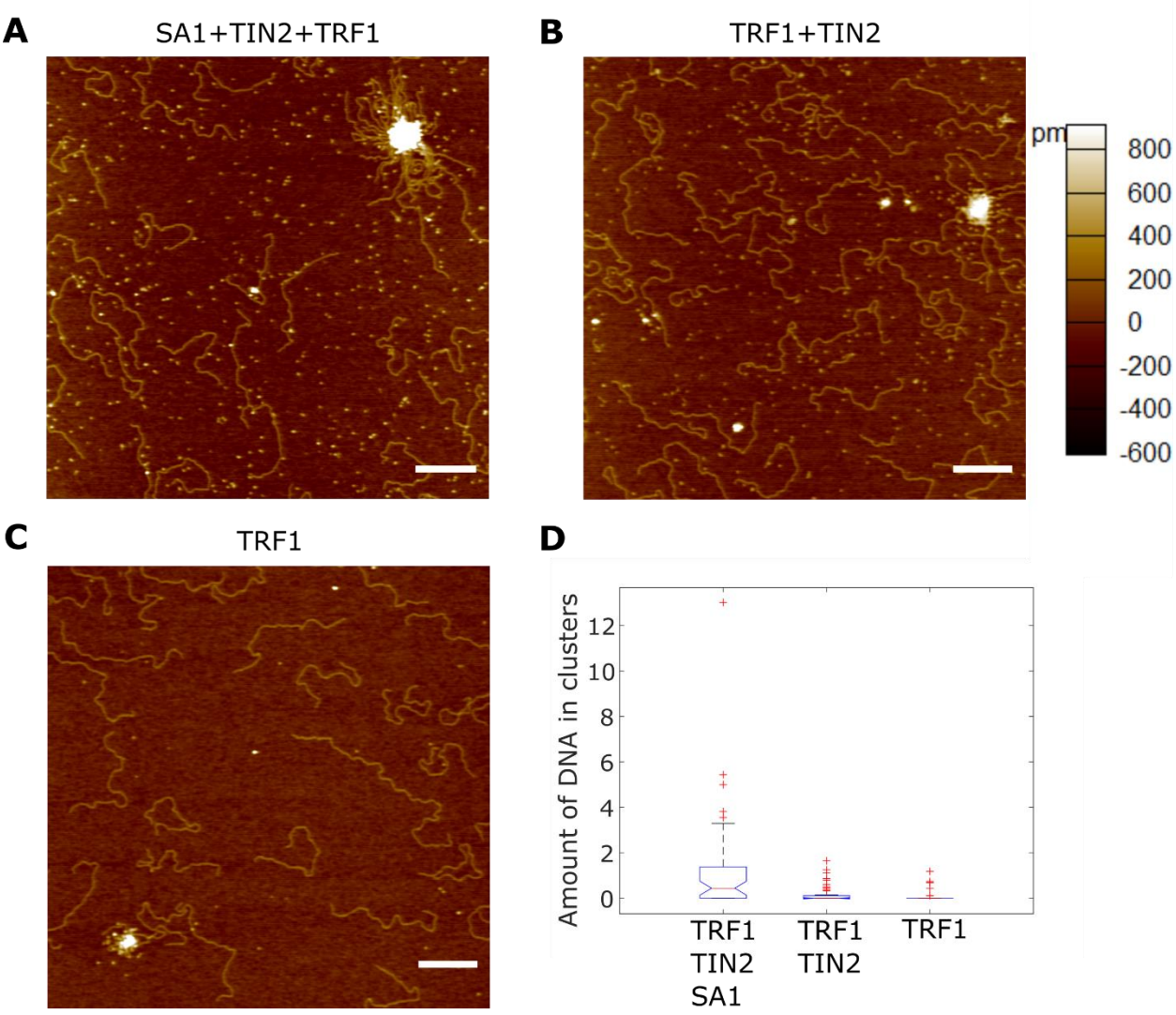

Figure 5.5 AFM result of DNA in clusters under different protein conditions. AFM image of pRST5 incubating with 25 nM SA1, TIN2, and TRF1 (A), pRST5 with 25 nM TIN2 and TRF1(B), and pRST5 with 25 nM TRF1 (C) for 5mins. White scale bars represent 400 nm. The Box-Whisker plots of the amount of DNA under different protein conditions for each frame (D). The red line shows the median. The box shows 25-75% of the dataset.

Above, we motivated the TRF1-TIN2-SA1 system by its significance to telomere cohesion (Canudas et al. 2007; Lin et al. 2016). In order to strengthen that argument, we verified that the observed compaction is specific to telomeric substrates. In particular, we compared the length decrease of λ-DNA and telomeric DNA under exposure to the mixture of 450 nM each of SA1, TRF1, and TIN2 (Fig. 5.4 **C**). We found that λ-DNA length changed by -2 $\pm$ 1% under the exposure to SA1, TRF1, and TIN2 mixture, which is significantly different from the result on the telomeric substrate presented in Fig. 5.4 **B** ($p<3 \times 10^{-4}$). The sequence specificity is attributable to the specificity of TRF1 and SA1 to telomeric DNA (Erdel et al. 2017; Lin et al. 2016).

5.4 Discussion

We started this study with the question of how a cohesion of DNA arises from an interaction network containing three proteins and two DNA segments. In such a complex system, co-location, or even binding, in an *in vitro* experiment does not imply a functional role since there is a large number of contact points. Usually, the question of which interactions are necessary for the function is answered through a time-consuming process of *in-vivo* observation of deletion mutants. Here we outline a procedure that allows some of this information to be found through a less labor-intensive pathway of performing sequential exposure experiments in nanofluidic devices.

We first consider the system of DNA, TRF1, and TIN2, but without SA1, for which prior publications indicate that the addition of full-length TIN2 stabilizes both compaction of telomeric DNA as well as mutual binding of independent DNA segments. We confirmed the compaction in nanochannel measurements and the mutual binding using AFM. We interpret the

compaction as a capture of two genetically distant (say 100s of bp) segments of the same molecule by a protein bridge. Two overall possibilities for the stoichiometry of a single bridge exist, which are that the bridge is formed by a TRF1 dimer with one $TRF1_{Myb}$ bound to each distant DNA segment, or that the bridge is formed by two TRF1 dimers that each are bound to one DNA segment. In the following, we will use an argument of elimination to that the latter possibility is favored by our data by eliminating all pathways that lead to a single dimer with two $TRF1_{Myb}$ bound to two distant DNA segments.

The near equality of the compaction of TRF1+TIN2 and TRF1 first/TIN2 second cases leads to start the consideration with a single TRF1 dimer that is initially bound to a single DNA segment. This must be the initial configuration since, in a nanochannel, DNA is initially stretched out and TRF1 is flushed over it. The stability of the dimer can be, in principle, computed by knowing the affinity of $TRF1_{Myb}$ to telomeric DNA and the affinity for dimerization of $TRF1_{TRFH}$. Let us assume first that the lifetime of $TRF1_{TRFH}$ dimerization is smaller than that of the $TRF1_{Myb}$-DNA binding. Due to the mobility of the $TRF1_{Myb}$ on DNA, we would expect that these dimers randomly dissociate. TIN2 binding would reduce the TRF1 dimer dissociation rate. However, we have not observed such TRF1 dimer splitting events into prior studies (Lin et al. 2014), TRF1 is typically reported to bind in a dimeric state, and the stabilization of $TRF1_{TRFH}$ dimer by $TIN2_{TBM}$ without the N-terminal TRFH domain was not sufficient to induce coordination of independent DNA segments (S. H. Kim et al. 2003). We further speculate about how bridges would be formed in the TRF1/TIN2 experiment: a bridge would require that a TRF1 dimer without TIN2 opens up, that $TIN2_{TBM}$ binds to the $TRF1_{TRFH}$, and that in exactly that moment the newly formed $TRF1_{TRFH}$-$TIN2_{TBM}$ finds a $TRF1_{TRFH}$ on a genetically distant DNA segment,

rather than the other half of the TRF1 dimer of which it initially was part of. It would appear that this process strongly favors a redimerization rather than the bridge formation. So we exclude the hypothesis of single trans-segment dimers with time-limited $TRF1_{TRFH}$, and rather assume that the binding of $TRF1_{Myb}$ is the shorter-lived binding location along the DNA-TRF1-TRF1-DNA bridge. However, we also have no possible mechanism to suggest that TIN2 binding at $TRF1_{TRFH}$ increases the stability of the $TRF1_{Myb}$-DNA contact, especially in light of the flexible linker between $TRF1_{TRFH}$ and $TRF1_{Myb}$. We can gauge the probability of the single bridging TRF1-dimer hypothesis by looking at the quantitative results.

We approximate the statistical mechanics of the nanoconfined DNA as a harmonic spring, i.e., we use a mean-field approximation. In this approximation, the protein-based stabilizing energy is proportional to the square of the change of length relative to the protein-free equilibrium energy. We first consider the impact of TIN2 on the bridges formed by TRF1 alone at a concentration of 450 nM of each protein. We conclude that the free energy of TRF1-based loop stabilization increases at least 10-fold with the addition of TIN2. This drastic increase appears unlikely in the case when the $TRF1_{Myb}$ is limiting the stability of the bridge since TIN2 does not act on it. Thus, we have ruled out both hypotheses that see a compaction because the two TRF1 making up one TRF1 dimer have bound to genetically distant DNA segments. That leads us to consider the alternative hypothesis, which is that of two DNA dimers that have both $TRF1_{Myb}$ bound to the same DNA segment bind to form a tetramer. This tetramerization has been discussed in the literature, but it is currently unclear how the tetramerization depends on TIN2 domains that do not directly bind to $TRF1_{TRFH}$ (S. H. Kim et al. 2003). Possible pathways would be the interaction of two TIN2 that are each bound to one TRF1 dimer, or a second, weak binding locus

between TIN2 and TRF1 that is independent of $TIN2_{TBM}$ and $TRF1_{TRFH}$. We conclude this discussion by comparing experiments with TRF1+TIN2 at 450 nM to those with TRF1+TIN2 at 900 nM. We infer that the latter has about twice the free energy due to the formation of protein bridges (data in Fig. 5.1 **A**). Following established polymer physics, we would anticipate the number of bridges (and thus the energy) to scale with the square of the bound TRF1+TIN2 concentration, which is apparently not the case. However, it is conceivable that the protein-DNA binding at 900 nM is becoming footprint-limited such that the concentration of DNA-bound TRF1+TIN2 is not twice that at 450 nM. The doubling will also be consistent if the rate of loop formation is not limited by the protein concentration, but rather by the frequency of DNA-DNA contacts within a DNA molecule stretched by nanoconfinement. Once a spontaneous contact is made, short-range diffusion of protein along DNA would allow the efficient formation of bridges.

We now turn the discussion to combinations of TRF1+TIN2 and SA1 (all proteins at 450 nM). First, we note that the exposure sequence does not impact the DNA compaction, and thus we conclude that SA1 binding and TRF1+TIN2 binding occur independently of each other. The open question is whether the compaction by the 3-protein system requires the interaction of SA1 with TRF1+TIN2, or whether these proteins work independently. In the quantitative analysis, we find that the free energy of compaction by TRF1+TIN2 is about 5 to 6 times the free energy in the presence of SA1, but that the free energy of compaction under the combination of all three proteins is twice the free energy under TRF1+TIN2. Thus, the free energy of the three-protein system is not simply the sum of free energies of the TRF1+TIN2 and SA1 conditions but exceeds it considerably. We attribute this to an interaction of SA1 with both TRF1 and TIN2. Such interactions have been reported in the literature (Canudas et al. 2007). Without a detailed

model of the complex, we are not able to determine whether this stabilization is principally based on the additive effects of binding energies within the cluster, or whether additional allosteric effects play a role.

While the nanochannel system is an interesting tool for quantification under sequential protein exposure, some important differences to prior studies and their interpretation are warranted. In general, we used protein concentrations that exceed those used in the references (Chen et al. 2008; Hanaoka, Nagadoi, and Nishimura 2009; Lin et al. 2016). The concentrations were chosen such that statistically significant differences between conditions could be observed. This is illustrated in the compaction by TRF1 alone, which at 450 nM was barely observable, while AFM observations gave an unambiguous signal (Lin et al. 2014, 2016; Pan et al. 2021). The first reason for this suppression of compaction in nanochannels is that the extension within a nanochannel imposes a higher energy penalty than in a free solution. The second reason is that the non-specific binding of protein to channel walls may render the protein concentration in devices lower than at first intended. A third reason is linked to the specific structure of TRF1-based DNA bridging. In particular, DNA-DNA tracks have been reported, which in a nanochannel carry a considerable energy penalty because a z-fold in which a channel segment is occupied by three DNA will have to be formed.

We note that the question of DNA configuration may be an underappreciated detail that complicates the assessment of relative contributions of individual proteins and binding sites, even in DNA that is not under nanochannel confinement. In particular, the high persistence length of DNA leads to different effective affinities for the binding of the first and second

$TRF1_{Myb}$ to DNA, and in a similar way, leads to different affinities of binding TRF1 dimers to two nearby DNA segments that are either parallel or perpendicular. Once a parallel orientation is achieved, the free energy of binding increases. Nanochannels clearly modify the probabilities of specific mutual DNA configuration, but so do many nuclear proteins *in vivo* experiments. We also note that recently developed descriptions of nucleoprotein-DNA complexes in terms of phase transitions would predict a susceptibility of DNA compaction to the effects of confinement.

Finally, we can comment on the hypothesis that TRF1, TIN2, and SA1 form the minimal system required for telomere cohesion. Through quantification of compaction in nanochannels as well as AFM imaging of DNA aggregation, we have determined that the three-protein system has an activity that considerably exceeds that of its constituents. We further showed that this activity is specific to telomeric DNA substrates.

CHAPTER 6 CONCLUSIONS

My research focuses on the interaction between the DNA and protein inside nanochannels, mainly focusing on the DNA conformation changes caused by the proteins. These protein-DNA interactions are required for DNA to fulfill its functional role. Investigations on DNA-protein interactions, including binding, looping, and contour length change, help understand the protein function. Nanochannels provide an excellent platform to investigate the interaction between DNA and protein at the molecule level because they do not require DNA modification to tether it onto a surface (although they do require fluorescence labeling for the observation). Common problems for this system include controlling the flow, the design and fabrication of nanofluidic chips, and surface passivation. My efforts focus on the latter two parts since we already have a good flow controlling system. We used dopamine-PEG passivation for our nanofluidic device to suppress the nonspecific binding of proteins, which enabled later experiments with different proteins inside nanochannels. A second focus is that the combination of AFM results with our nanofluidic device results, which gives us a better understanding of the mechanism of the DNA-protein interaction. Compared to fluorescence microscopy, AFM has a better spatial resolution and enables us to observe at a nanometer scale.

In our first project, we investigated the two proteins MeCP2 and MBD2 from the MBD family. We find a significant difference between the MeCP2 and MBD2 in their ability to change DNA configuration. Based on the AFM result, we propose that these configuration changes may result from the DNA looping ability of MeCP2. However, we were not able to resolve how the MeCP2 loops DNA. We determined that the MeCP2 binding involves co-operativity. We also suggest that the compaction may partially explain the repression of expression in cytosine methylation by

MeCP2. Also, the different function of MBD2 makes ATTO 565-labeled MBD2FLsc a promising labeling agent specifically targeting methylated CpG sites without changing DNA length under nanochannel stretching. This work can be further expanded in different aspects, including searching for the functional domains responsible for this different behavior of MBD2 and MeCP2 and the interactions of MBD2 or MeCP2 with other nucleoproteins.

We extended our investigations to a three-protein system in our second project and demonstrated the feasibility of sequentially exposing DNA molecules to proteins using our combinatory device. We suggested a possible scenario for explaining SA1, TRF1, and TIN2 work together as a minimum system for the sister telomere cohesion *in vivo*. We also found that for TRF1 and TIN2 systems, exposure sequence does impact results, while for SA1 and TRF1-TIN2 systems, it does not matter. The AFM result also suggests that these three components work together to compact DNA rather than working independently. Although we cannot solve the atomic structure of these three proteins on the DNA, we proposed a possible model for the interaction in between.

In summary, my work contributes to biological studies of the interaction between proteins and DNA within a nanofluidic device. We demonstrate the feasibility of using the nanofluidic device with different designs and combine it with AFM in studying different DNA-protein systems. We discovered a novel difference in behavior in the MBD family which may help explain gene expression repression. We proposed possible models for SA1, TRF1, and TIN2-induced sister telomere cohesion *in vivo*. Further work can expand in two ways: new biological discoveries and new nanofluidic device development.

CHAPTER 7 FUTURE WORK

7.1 Cohesin loop extrusion

The nature of interactions between protein and DNA that can be studied inside nanofluidic devices can be further expanded by focusing on new nanofluidic designs for ATPase-dependent active protein-DNA systems. My main interest lies in using the nanofluidic device for measuring DNA loop extrusion by cohesin. Canonical human cohesin is a protein complex that contains four subunits, Smc1, Smc3, Scc1 (Rad21), and Scc3 (SA1 or SA2) (Gruber, Haering, and Nasmyth 2003). Cohesin was initially known for its crucial role in controlling sister chromatids cohesion during cell division (Davidson and Peters 2021). In addition to its role in sister chromatids cohesion, cohesin is also involved in chromatin organization and gene regulation (Merkenschlager and Nora 2016). This implies that the cohesin can connect DNA both *in trans* and *cis*. Researchers have proposed a loop extrusion model for the mechanism of how cohesin creates loops. This loop extrusion of cohesin has been observed in vitro, and it has been found that NIPBL is necessary for DNA binding (Y. Kim et al. 2019). Cryo-EM solved the structure of cohesion and NIPBL on DNA (Shi et al. 2020). However, the fluorescence imaging methods applied in almost all prior publications require the tethering of the DNA molecule, which can induce tension on DNA during loop extrusion. This tension creates extra force on the DNA and influences the extrusion speed. A tension-free method is needed to solve this problem. Researchers have observed cohesin enrichment near the site where the CCCTC binding factor (CTCF) is (Wendt et al. 2008). Based on this observation, one of the models of the loop extrusion is that the cohesion can extrude DNA into a loop, and CTCF can stop the loop extrusion by the cohesin (Wendt et al. 2008).

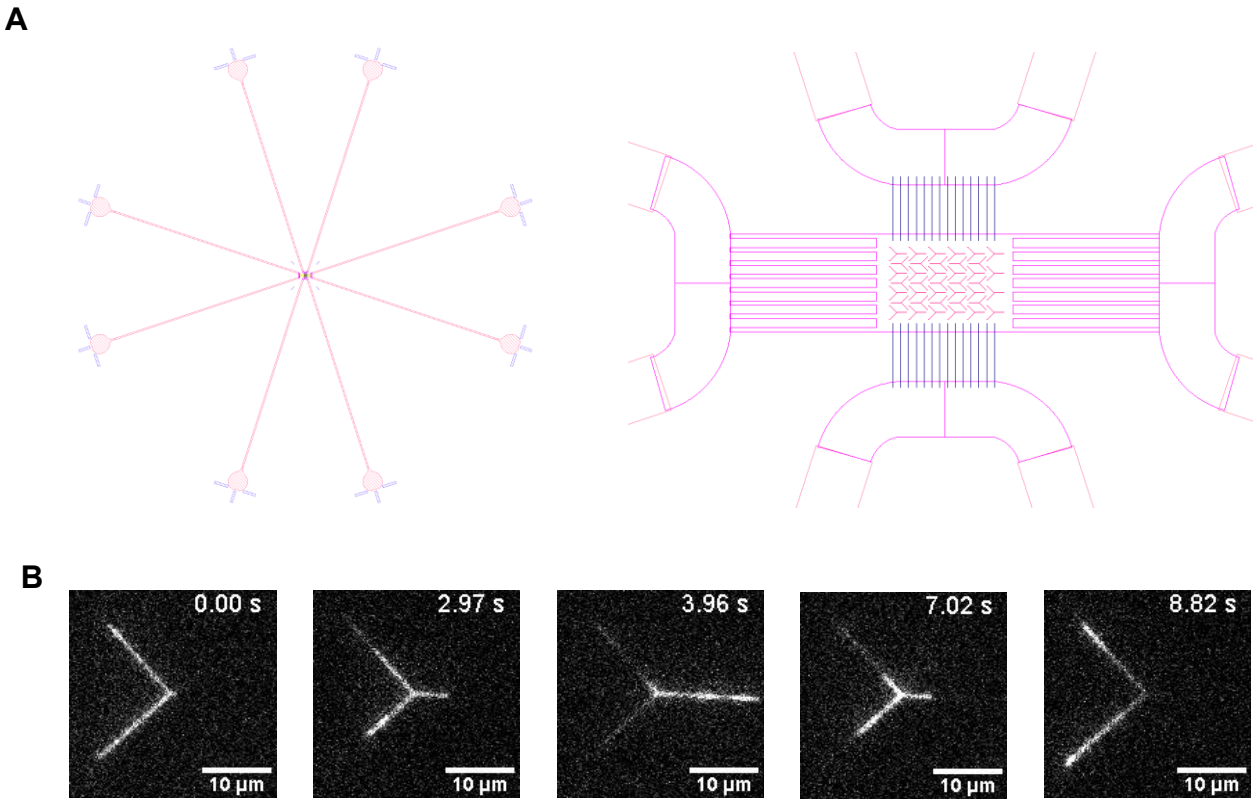

Figure 7.1 The layout of the new Y-channel and an example of manipulating DNA inside Y-channels. (A) The layout of our new combinatory nanofluidic device with a Y-channel. (B) Fluorescence image of manipulating a T4 phage DNA inside the Y-channel.

Nanofluidic devices are a good candidate for solving this problem. DNA is stretched inside the nanochannel due to the nanoconfinement through a competition between the entropy and the self-avoidant interaction. We designed a new nanochannel pattern based on the combinatory channels. Instead of the straight nano slots, those devices contain nano slots in a 'Y' shape. Manipulation of DNA inside Y channels has been achieved by Zubair (Azad, Roushan, and Riehn 2015), and they are all available now in the combinatory device. The Y-channels allow the DNA to extend in two of the three channels, and the loop will extend in the third. Using this new

nanochannel design, we can observe the DNA loop extrusion in real-time and measure the extrusion speed without tension.

7.2 Preliminary results and discussion

The introduction of the new channel was discussed in the previous chapter. The main change here is the Y channels. Since the primary step of fabrication did not change, we will not discuss the details of the fabrication. Our Y-channel is a combination of three nano slots. The size of each nano slot is 150 nm × 150 nm × 10 μm. An ideal DNA substrate for loop extrusion would be longer than 10μm so that the DNA can occupy two nano slots. Here we used T4 phage DNA (169 kbps) as our DNA substrate for observation. The T4 DNA was stained using YOYO-1 at a 1:20 dye to base-pair ratio. The cohesin complex, containing NIPBL as the DNA loader, was a generous gift from Hongtao Yu's group. During the nanochannel experiment, the cohesion-NIPBL was incubated with the T4 phage DNA in a working buffer (20 mM Tris-HCl pH 7.5, 1 mM MgCl2, 50 mM NaCl). The mixture was then pushed into the shunt region by nitrogen, and the DNA was trapped in Y channels due to entropy. For control, we also used ATPase-deficient mutant cohesin for the nanofluidic experiments. A summary of preliminary data is shown below.

We observed loop lengths increasing when incubating with cohesin-NIPBL, while we observed stalled loops when incubating with ATPase-deficient mutant cohesin. This agrees with the publications reporting that loop extrusion requires ATP hydrolysis. We attempted to label cohesion fluorescently to provide future proof that a single cohesin ring drives loop extrusion. We used Tris-NTA to label the cohesion and observed a fluorescent colocalization between the

DNA and the Tris -NTA inside a nanochannel. However, we did not yet find buffer conditions that allowed loop extrusion with fluorescently labeled protein.

The long-term aim is to study the cohesion activity through cofactors of the DNA-binding protein CTCF. T4 phage DNA contains about 24 CCCTC sites for the CTCF. Future work will use fluorescently labeled CTCF bound to T4 DNA. We will then induce cohesin to start DNA loop extrusion. We can then determine whether the CTCF is stopping the cohesin loop extrusion in real-time imaging. However, due to the high surface-to-volume ratio inside the nanofluidic device, initial challenges in the surface passivation to prevent CTCF absorption proved difficult to solve. Future work on this will need better surface passivation. Also, since there are only around 24 specific binding sites on T4 DNA for CTCF, the nonspecific binding will need to be lowered until it is not significant compared to the specific binding. Future work on this will need a well-designed experimental procedure for solving this problem.

REFERENCES

Adair, G.S., A.V. Bock, and H. Field. 1925. “THE HEMOGLOBIN SYSTEM.” *Journal of Biological Chemistry* 63 (2): 529–45. https://doi.org/10.1016/S0021-9258(18)85018-9.

Anderson, W. F., D. H. Ohlendorf, Y. Takeda, and B. W. Matthews. 1981. “Structure of the Cro Repressor from Bacteriophage λ and Its Interaction with DNA.” *Nature* 290 (5809): 754–58. https://doi.org/10.1038/290754a0.

Azad, Zubair, Maedeh Roushan, and Robert Riehn. 2015. “DNA Brushing Shoulders: Targeted Looping and Scanning of Large DNA Strands.” *Nano Letters* 15 (8): 5641–46. https://doi.org/10.1021/acs.nanolett.5b02476.

Bagchi, Biman, Paul C. Blainey, and X. Sunney Xie. 2008. “Diffusion Constant of a Nonspecifically Bound Protein Undergoing Curvilinear Motion along DNA †.” *The Journal of Physical Chemistry B* 112 (19): 6282–84. https://doi.org/10.1021/jp077568f.

Barouki, R., E. Melén, Z. Herceg, J. Beckers, J. Chen, M. Karagas, A. Puga, et al. 2018. “Epigenetics as a Mechanism Linking Developmental Exposures to Long-Term Toxicity.” *Environment International* 114 (May): 77–86. https://doi.org/10.1016/j.envint.2018.02.014.

Baubec, Tuncay, Robert Ivánek, Florian Lienert, and Dirk Schübeler. 2013. “Methylation-Dependent and -Independent Genomic Targeting Principles of the MBD Protein Family.” *Cell* 153 (2): 480–92. https://doi.org/10.1016/j.cell.2013.03.011.

Bianchi, Alessandro. 1997. “TRF1 Is a Dimer and Bends Telomeric DNA.” *The EMBO Journal* 16 (7): 1785–94. https://doi.org/10.1093/emboj/16.7.1785.

Binnig, G., C. F. Quate, and Ch Gerber. 1986. “Atomic Force Microscope.” *Physical Review Letters* 56 (9): 930–33. https://doi.org/10.1103/PhysRevLett.56.930.

Bisht, Kamlesh K., Zharko Daniloski, and Susan Smith. 2013. “SA1 Binds Directly to DNA via

Its Unique AT-Hook to Promote Sister Chromatid Cohesion at Telomeres." *Journal of Cell Science* 126 (15): 3493–3503. https://doi.org/10.1242/jcs.130872.

Bohr, Chr., K. Hasselbalch, and August Krogh. 1904. "Ueber Einen in Biologischer Beziehung Wichtigen Einfluss, Den Die Kohlensäurespannung Des Blutes Auf Dessen Sauerstoffbindung Übt1." *Skandinavisches Archiv Für Physiologie* 16 (2): 402–12. https://doi.org/10.1111/j.1748-1716.1904.tb01382.x.

Canudas, Silvia, Benjamin R. Houghtaling, Ju Youn Kim, Jasmin N. Dynek, William G. Chang, and Susan Smith. 2007. "Protein Requirements for Sister Telomere Association in Human Cells." *The EMBO Journal* 26 (23): 4867–78. https://doi.org/10.1038/sj.emboj.7601903.

Canudas, Silvia, and Susan Smith. 2009. "Differential Regulation of Telomere and Centromere Cohesion by the Scc3 Homologues SA1 and SA2, Respectively, in Human Cells." *Journal of Cell Biology* 187 (2): 165–73. https://doi.org/10.1083/jcb.200903096.

Cerf, Aline, Benjamin R. Cipriany, Jaime J. Benítez, and Harold G. Craighead. 2011. "Single DNA Molecule Patterning for High-Throughput Epigenetic Mapping." *Analytical Chemistry* 83 (21): 8073–77. https://doi.org/10.1021/ac202506j.

Chen, Yong, Yuting Yang, Megan Van Overbeek, Jill R. Donigian, Paul Baciu, Titia De Lange, and Ming Lei. 2008. "A Shared Docking Motif in TRF1 and TRF2 Used for Differential Recruitment of Telomeric Proteins." *Science* 319 (5866): 1092–96. https://doi.org/10.1126/science.1151804.

Collins, Bridget E., Ling F. Ye, Daniel Duzdevich, and Eric C. Greene. 2014. "DNA Curtains." In *Methods in Cell Biology,* 123:217–34. https://doi.org/10.1016/B978-0-12-420138-5.00012-4.

Countryman, Preston, Yanlin Fan, Aparna Gorthi, Hai Pan, Jack Strickland, Parminder Kaur,

Xuechun Wang, et al. 2018. “Cohesin SA2 Is a Sequence-Independent DNA-Binding Protein That Recognizes DNA Replication and Repair Intermediates.” *Journal of Biological Chemistry* 293 (3): 1054–69. https://doi.org/10.1074/jbc.M117.806406.

Crick, Francis. 1970. “Central Dogma of Molecular Biology.” *Nature* 227 (5258): 561–63. https://doi.org/10.1038/227561a0.

Dagata, J. A., J. Schneir, H. H. Harary, C. J. Evans, M. T. Postek, and J. Bennett. 1998. “Modification of Hydrogen-passivated Silicon by a Scanning Tunneling Microscope Operating in Air.” *Applied Physics Letters* 56 (20): 2001. https://doi.org/10.1063/1.102999.

Davidson, Iain F., and Jan-Michael Peters. 2021. “Genome Folding through Loop Extrusion by SMC Complexes.” *Nature Reviews Molecular Cell Biology* 22 (7): 445–64. https://doi.org/10.1038/s41580-021-00349-7.

Deen, Jochem, Charlotte Vranken, Volker Leen, Robert K. Neely, Kris P. F. Janssen, and Johan Hofkens. 2017. “Methyltransferase-Directed Labeling of Biomolecules and Its Applications.” *Angewandte Chemie International Edition* 56 (19): 5182–5200. https://doi.org/10.1002/anie.201608625.

Desai, Megha A., Heather D. Webb, Leander M. Sinanan, J. Neel Scarsdale, Ninad M. Walavalkar, Gordon D. Ginder, and David C. Williams. 2015. “An Intrinsically Disordered Region of Methyl-CpG Binding Domain Protein 2 (MBD2) Recruits the Histone Deacetylase Core of the NuRD Complex.” *Nucleic Acids Research* 43 (6): 3100–3113. https://doi.org/10.1093/nar/gkv168.

Dor, Yuval, and Howard Cedar. 2018. “Principles of DNA Methylation and Their Implications for Biology and Medicine.” *The Lancet* 392 (10149): 777–86. https://doi.org/10.1016/S0140-6736(18)31268-6.

Du, Qian, Phuc-Loi Luu, Clare Stirzaker, and Susan J. Clark. 2015. “Methyl-CpG-Binding Domain Proteins: Readers of the Epigenome.” *Epigenomics* 7 (6): 1051–73. https://doi.org/10.2217/epi.15.39.

Erdel, Fabian, Katja Kratz, Smaranda Willcox, Jack D. Griffith, Eric C. Greene, and Titia de Lange. 2017. “Telomere Recognition and Assembly Mechanism of Mammalian Shelterin.” *Cell Reports* 18 (1): 41–53. https://doi.org/10.1016/j.celrep.2016.12.005.

Fairall, Louise, Lynda Chapman, Heidi Moss, Titia De Lange, and Daniela Rhodes. 2001. “Structure of the TRFH Dimerization Domain of the Human Telomeric Proteins TRF1 and TRF2.” *Molecular Cell* 8 (2): 351–61. https://doi.org/10.1016/S1097-2765(01)00321-5.

Farlik, Matthias, Nathan C. Sheffield, Angelo Nuzzo, Paul Datlinger, Andreas Schönegger, Johanna Klughammer, and Christoph Bock. 2015. “Single-Cell DNA Methylome Sequencing and Bioinformatic Inference of Epigenomic Cell-State Dynamics.” *Cell Reports* 10 (8): 1386–97. https://doi.org/10.1016/j.celrep.2015.02.001.

Fasolino, Maria, and Zhaolan Zhou. 2017. “The Crucial Role of DNA Methylation and MeCP2 in Neuronal Function.” *Genes* 8 (5): 141. https://doi.org/10.3390/genes8050141.

Frykholm, Karolin, Mohammadreza Alizadehheidari, Joachim Fritzsche, Jens Wigenius, Mauro Modesti, Fredrik Persson, and Fredrik Westerlund. 2014. “Probing Physical Properties of a DNA-Protein Complex Using Nanofluidic Channels.” *Small* 10 (5): 884–87. https://doi.org/10.1002/smll.201302028.

Frykholm, Karolin, Ronnie Per-Arne Berntsson, Magnus Claesson, Laura de Battice, Richard Odegrip, Pål Stenmark, and Fredrik Westerlund. 2016. “DNA Compaction by the Bacteriophage Protein Cox Studied on the Single DNA Molecule Level Using Nanofluidic Channels.” *Nucleic Acids Research* 44 (15): gkw352. https://doi.org/10.1093/nar/gkw352.

Frykholm, Karolin, Lena K. Nyberg, and Fredrik Westerlund. 2017. “Exploring DNA–Protein Interactions on the Single DNA Molecule Level Using Nanofluidic Tools.” *Integrative Biology* 9 (8): 650–61. https://doi.org/10.1039/C7IB00085E.

Gajiwala, Ketan S., Hua Chen, Fabrice Cornille, Bernard P. Roques, Walter Reith, Bernard Mach, and Stephen K. Burley. 2000. “Structure of the Winged-Helix Protein HRFX1 Reveals a New Mode of DNA Binding.” *Nature* 403 (6772): 916–21. https://doi.org/10.1038/35002634.

Gennes, P. G. de, and Thomas A. Witten. 1980. “Scaling Concepts in Polymer Physics.” *Physics Today* 33 (6): 51–54. https://doi.org/10.1063/1.2914118.

Georgel, Philippe T., Rachel A. Horowitz-Scherer, Nick Adkins, Christopher L. Woodcock, Paul A. Wade, and Jeffrey C. Hansen. 2003. “Chromatin Compaction by Human MeCP2.” *Journal of Biological Chemistry* 278 (34): 32181–88. https://doi.org/10.1074/jbc.M305308200.

Gokul, Gopinathan, and Sanjeev Khosla. 2013. “DNA Methylation and Cancer.” In *Subcellular Biochemistry*, 61:597–625. Academic Press. https://doi.org/10.1007/978-94-007-4525-4_26.

Griffith, Jack, Alessandro Bianchi, and Titia de Lange. 1998. “TRF1 Promotes Parallel Pairing of Telomeric Tracts in Vitro.” *Journal of Molecular Biology* 278 (1): 79–88. https://doi.org/10.1006/jmbi.1998.1686.

Gruber, Stephan, Christian H. Haering, and Kim Nasmyth. 2003. “Chromosomal Cohesin Forms a Ring.” *Cell* 112 (6): 765–77. https://doi.org/10.1016/S0092-8674(03)00162-4.

Grutsch, Sarina, Sven Brüschweiler, and Martin Tollinger. 2016. “NMR Methods to Study Dynamic Allostery.” Edited by Jin Liu. *PLOS Computational Biology* 12 (3): e1004620. https://doi.org/10.1371/journal.pcbi.1004620.

Guttula, Durgarao, Fan Liu, Jeroen A. van Kan, Véronique Arluison, and Johan R. C. van der Maarel. 2018. “Effect of HU Protein on the Conformation and Compaction of DNA in a Nanochannel.” *Soft Matter* 14 (12): 2322–28. https://doi.org/10.1039/C7SM02118F.

Haering, Christian H., Jan Löwe, Andreas Hochwagen, and Kim Nasmyth. 2002. “Molecular Architecture of SMC Proteins and the Yeast Cohesin Complex.” *Molecular Cell* 9 (4): 773–88. https://doi.org/10.1016/S1097-2765(02)00515-4.

Halford, Stephen E. 2004. “How Do Site-Specific DNA-Binding Proteins Find Their Targets?” *Nucleic Acids Research* 32 (10): 3040–52. https://doi.org/10.1093/nar/gkh624.

Hanaoka, Shingo, Aritaka Nagadoi, and Yoshifumi Nishimura. 2009. “Comparison between TRF2 and TRF1 of Their Telomeric DNA-Bound Structures and DNA-Binding Activities.” *Protein Science* 14 (1): 119–30. https://doi.org/10.1110/ps.04983705.

Hanish, J P, J L Yanowitz, and T. de Lange. 1994. “Stringent Sequence Requirements for the Formation of Human Telomeres.” *Proceedings of the National Academy of Sciences* 91 (19): 8861–65. https://doi.org/10.1073/pnas.91.19.8861.

Heller, Iddo, Tjalle P. Hoekstra, Graeme A. King, Erwin J. G. Peterman, and Gijs J. L. Wuite. 2014. “Optical Tweezers Analysis of DNA–Protein Complexes.” *Chemical Reviews* 114 (6): 3087–3119. https://doi.org/10.1021/cr4003006.

Henderson, E., P. Haydon, and D. Sakaguchi. 1992. “Actin Filament Dynamics in Living Glial Cells Imaged by Atomic Force Microscopy.” *Science* 257 (5078): 1944–46. https://doi.org/10.1126/science.1411511.

Hendrich, Brian, and Adrian Bird. 1998. “Identification and Characterization of a Family of Mammalian Methyl-CpG Binding Proteins.” *Molecular and Cellular Biology* 18 (11): 6538–47. https://doi.org/10.1128/MCB.18.11.6538.

Horvath, Steve, and Kenneth Raj. 2018. “DNA Methylation-Based Biomarkers and the Epigenetic Clock Theory of Ageing.” *Nature Reviews Genetics* 19 (6): 371–84. https://doi.org/10.1038/s41576-018-0004-3.

Jalili, Nader, and Karthik Laxminarayana. 2004. “A Review of Atomic Force Microscopy Imaging Systems: Application to Molecular Metrology and Biological Sciences.” *Mechatronics* 14 (8): 907–45. https://doi.org/10.1016/j.mechatronics.2004.04.005.

Jin, Jianshi, Tengfei Lian, Chan Gu, Kai Yu, Yi Qin Gao, and Xiao-Dong Su. 2016. “The Effects of Cytosine Methylation on General Transcription Factors.” *Scientific Reports* 6 (1): 29119. https://doi.org/10.1038/srep29119.

Kalathiya, Umesh, Monikaben Padariya, and Maciej Baginski. 2018. “The Structurally Similar TRFH Domain of TRF1 and TRF2 Dimers Shows Distinct Behaviour towards TIN2.” *Archives of Biochemistry and Biophysics* 642 (March): 52–62. https://doi.org/10.1016/j.abb.2018.02.005.

Kar, Anirban, Smaranda Willcox, and Jack D. Griffith. 2016. “Transcription of Telomeric DNA Leads to High Levels of Homologous Recombination and T-Loops.” *Nucleic Acids Research* 44 (19): gkw779. https://doi.org/10.1093/nar/gkw779.

Kemp, Graham. 1998. “Capillary Electrophoresis.” *Biotechnology and Applied Biochemistry* 27 (1): 9–17. https://doi.org/10.1111/j.1470-8744.1998.tb01369.x.

Khrapunov, Sergei, Yisong Tao, Huiyong Cheng, Camille Padlan, Richard Harris, Aristea S. Galanopoulou, John M. Greally, Mark E. Girvin, and Michael Brenowitz. 2016. “MeCP2 Binding Cooperativity Inhibits DNA Modification-Specific Recognition.” *Biochemistry* 55 (31): 4275–85. https://doi.org/10.1021/acs.biochem.6b00451.

Khrapunov, Sergei, Christopher Warren, Huiyong Cheng, Esther R. Berko, John M. Greally, and

Michael Brenowitz. 2014. “Unusual Characteristics of the DNA Binding Domain of Epigenetic Regulatory Protein MeCP2 Determine Its Binding Specificity.” *Biochemistry* 53 (21): 3379–91. https://doi.org/10.1021/bi500424z.

Kim, Joseph L., Dimitar B. Nikolov, and Stephen K. Burley. 1993. “Co-Crystal Structure of TBP Recognizing the Minor Groove of a TATA Element.” *Nature* 365 (6446): 520–27. https://doi.org/10.1038/365520a0.

Kim, Sahn-ho, Patrick Kaminker, and Judith Campisi. 1999. “TIN2, a New Regulator of Telomere Length in Human Cells.” *Nature Genetics* 23 (4): 405–12. https://doi.org/10.1038/70508.

Kim, Sahn-Ho Ho, Seungil Han, Young-Hyun Hyun You, David J. Chen, and Judith Campisi. 2003. “The Human Telomere-Associated Protein TIN2 Stimulates Interactions between Telomeric DNA Tracts in Vitro.” *EMBO Reports* 4 (7): 685–91. https://doi.org/10.1038/sj.embor.embor872.

Kim, Yoori, Zhubing Shi, Hongshan Zhang, Ilya J. Finkelstein, and Hongtao Yu. 2019. “Human Cohesin Compacts DNA by Loop Extrusion.” *Science* 366 (6471): 1345–49. https://doi.org/10.1126/science.aaz4475.

Klug, Aaron. 2010. “The Discovery of Zinc Fingers and Their Applications in Gene Regulation and Genome Manipulation.” *Annual Review of Biochemistry* 79 (1): 213–31. https://doi.org/10.1146/annurev-biochem-010909-095056.

Kojic, Aleksandar, Ana Cuadrado, Magali De Koninck, Daniel Giménez-Llorente, Miriam Rodríguez-Corsino, Gonzalo Gómez-López, François Le Dily, Marc A. Marti-Renom, and Ana Losada. 2018. “Distinct Roles of Cohesin-SA1 and Cohesin-SA2 in 3D Chromosome Organization.” *Nature Structural & Molecular Biology* 25 (6): 496–504.

https://doi.org/10.1038/s41594-018-0070-4.

Koo, D.-H., Fangpu Han, James A. Birchler, and Jiming Jiang. 2011. “Distinct DNA Methylation Patterns Associated with Active and Inactive Centromeres of the Maize B Chromosome.” *Genome Research* 21 (6): 908–14. https://doi.org/10.1101/gr.116202.110.

Krieg, Michael, Gotthold Fläschner, David Alsteens, Benjamin M. Gaub, Wouter H. Roos, Gijs J. L. Wuite, Hermann E. Gaub, Christoph Gerber, Yves F. Dufrêne, and Daniel J. Müller. 2019. “Atomic Force Microscopy-Based Mechanobiology.” *Nature Reviews Physics* 1 (1): 41–57. https://doi.org/10.1038/s42254-018-0001-7.

Kuśba, Józef, and Joseph R. Lakowicz. 1999. “Definition and Properties of the Emission Anisotropy in the Absence of Cylindrical Symmetry of the Emission Field: Application to the Light Quenching Experiments.” *The Journal of Chemical Physics* 111 (1): 89–99. https://doi.org/10.1063/1.479256.

Lange, Titia de. 2018. “Shelterin-Mediated Telomere Protection.” *Annual Review of Genetics* 52 (1): 223–47. https://doi.org/10.1146/annurev-genet-032918-021921.

LeBlanc, Sharonda, Hunter Wilkins, Zimeng Li, Parminder Kaur, Hong Wang, and Dorothy A. Erie. 2017. “Using Atomic Force Microscopy to Characterize the Conformational Properties of Proteins and Protein–DNA Complexes That Carry Out DNA Repair.” In *Methods in Enzymology*, 592:187–212. https://doi.org/10.1016/bs.mie.2017.04.004.

Lee, Haeshin, Shara M. Dellatore, William M. Miller, and Phillip B. Messersmith. 2007. “Mussel-Inspired Surface Chemistry for Multifunctional Coatings.” *Science* 318 (5849): 426–30. https://doi.org/10.1126/science.1147241.

Lee, Seunghwan, and Janos Vörös. 2005. “An Aqueous-Based Surface Modification of Poly(Dimethylsiloxane) with Poly(Ethylene Glycol) to Prevent Biofouling.” *Langmuir* 21

(25): 11957–62. https://doi.org/10.1021/la051932p.

Levy-Sakin, Michal, and Yuval Ebenstein. 2013. “Beyond Sequencing: Optical Mapping of DNA in the Age of Nanotechnology and Nanoscopy.” *Current Opinion in Biotechnology* 24 (4): 690–98. https://doi.org/10.1016/j.copbio.2013.01.009.

Lim, Ci Ji, and Thomas R. Cech. 2021. “Shaping Human Telomeres: From Shelterin and CST Complexes to Telomeric Chromatin Organization.” *Nature Reviews Molecular Cell Biology* 22 (4): 283–98. https://doi.org/10.1038/s41580-021-00328-y.

Lim, Ci Ji, Arthur J. Zaug, Hee Jin Kim, and Thomas R. Cech. 2017. “Reconstitution of Human Shelterin Complexes Reveals Unexpected Stoichiometry and Dual Pathways to Enhance Telomerase Processivity.” *Nature Communications* 8 (1): 1075. https://doi.org/10.1038/s41467-017-01313-w.

Lim, Shuang Fang, Alena Karpusenko, John J. Sakon, Joseph A. Hook, Tyra A. Lamar, and Robert Riehn. 2011. “DNA Methylation Profiling in Nanochannels.” *Biomicrofluidics* 5 (3): 034106. https://doi.org/10.1063/1.3613671.

Lin, Jiangguo, Preston Countryman, Noah Buncher, Parminder Kaur, Longjiang E, Yiyun Zhang, Greg Gibson, et al. 2014. “TRF1 and TRF2 Use Different Mechanisms to Find Telomeric DNA but Share a Novel Mechanism to Search for Protein Partners at Telomeres.” *Nucleic Acids Research* 42 (4): 2493–2504. https://doi.org/10.1093/nar/gkt1132.

Lin, Jiangguo, Preston Countryman, Haijiang Chen, Hai Pan, Yanlin Fan, Yunyun Jiang, Parminder Kaur, et al. 2016. “Functional Interplay between SA1 and TRF1 in Telomeric DNA Binding and DNA–DNA Pairing.” *Nucleic Acids Research* 44 (13): 6363–76. https://doi.org/10.1093/nar/gkw518.

Liu, Jin, and Ruth Nussinov. 2016. “Allostery: An Overview of Its History, Concepts, Methods,

and Applications." *PLoS Computational Biology* 12 (6). https://doi.org/10.1371/journal.pcbi.1004966.

Liu, Ming, Saeid Movahed, Saroj Dangi, Hai Pan, Parminder Kaur, Stephanie M. Bilinovich, Edgar M. Faison, et al. 2020. "DNA Looping by Two 5-Methylcytosine-Binding Proteins Quantified Using Nanofluidic Devices." *Epigenetics & Chromatin* 13 (1): 18. https://doi.org/10.1186/s13072-020-00339-7.

Luger, Karolin, Armin W. Mäder, Robin K. Richmond, David F. Sargent, and Timothy J. Richmond. 1997. "Crystal Structure of the Nucleosome Core Particle at 2.8 Å Resolution." *Nature* 389 (6648): 251–60. https://doi.org/10.1038/38444.

Maciejowski, John, and Titia de Lange. 2017. "Telomeres in Cancer: Tumour Suppression and Genome Instability." *Nature Reviews Molecular Cell Biology* 18 (3): 175–86. https://doi.org/10.1038/nrm.2016.171.

Mahmood, Niaz, and Shafaat A. Rabbani. 2019. "DNA Methylation Readers and Cancer: Mechanistic and Therapeutic Applications." *Frontiers in Oncology* 9 (JUN). https://doi.org/10.3389/fonc.2019.00489.

Mathieu, Olivier, Georges Picard, and Sylvette Tourmente. 2002. "Methylation of a Euchromatin-Heterochromatin Transition Region in Arabidopsis Thaliana Chromosome 5 Left Arm." *Chromosome Research* 10 (6): 455–66. https://doi.org/10.1023/A:1020936229771.

Merkenschlager, Matthias, and Elphège P. Nora. 2016. "CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation." *Annual Review of Genomics and Human Genetics* 17 (1): 17–43. https://doi.org/10.1146/annurev-genom-083115-022339.

Mikkelsen, Morten Bo, Walter Reisner, Henrik Flyvbjerg, and Anders Kristensen. 2011.

"Pressure-Driven DNA in Nanogroove Arrays: Complex Dynamics Leads to Length- and Topology-Dependent Separation." *Nano Letters* 11 (4): 1598–1602. https://doi.org/10.1021/nl1044764.

Miura, Fumihito, Yusuke Enomoto, Ryo Dairiki, and Takashi Ito. 2012. "Amplification-Free Whole-Genome Bisulfite Sequencing by Post-Bisulfite Adaptor Tagging." *Nucleic Acids Research* 40 (17): e136–e136. https://doi.org/10.1093/nar/gks454.

Monod, Jacque, Jeffries Wyman, and Jean-Pierre Changeux. 1965. "On the Nature of Allosteric Transitions: A Plausible Model." *Journal of Molecular Biology* 12 (1): 88–118. https://doi.org/10.1016/S0022-2836(65)80285-6.

Ngo, Thuy T. M., Jejoong Yoo, Qing Dai, Qiucen Zhang, Chuan He, Aleksei Aksimentiev, and Taekjip Ha. 2016. "Effects of Cytosine Modifications on DNA Flexibility and Nucleosome Mechanical Stability." *Nature Communications* 7 (1): 10813. https://doi.org/10.1038/ncomms10813.

Nikitina, Tatiana, Xi Shi, Rajarshi P. Ghosh, Rachel A. Horowitz-Scherer, Jeffrey C. Hansen, and Christopher L. Woodcock. 2007. "Multiple Modes of Interaction between the Methylated DNA Binding Protein MeCP2 and Chromatin." *Molecular and Cellular Biology* 27 (3): 864–77. https://doi.org/10.1128/MCB.01593-06.

Odijk, Theo. 2008. "Scaling Theory of DNA Confined in Nanochannels and Nanoslits." *Physical Review E* 77 (6): 060901. https://doi.org/10.1103/PhysRevE.77.060901.

Ödman, D., E. Werner, K. D. Dorfman, C. R. Doering, and B. Mehlig. 2018. "Distribution of Label Spacings for Genome Mapping in Nanochannels." *Biomicrofluidics* 12 (3): 034115. https://doi.org/10.1063/1.5038417.

Öz, Robin, Sean M. Howard, Rajhans Sharma, Hanna Törnkvist, Ilaria Ceppi, Sriram KK, Erik

Kristiansson, Petr Cejka, and Fredrik Westerlund. 2020. “Phosphorylated CtIP Bridges DNA to Promote Annealing of Broken Ends.” *Proceedings of the National Academy of Sciences* 117 (35): 21403–12. https://doi.org/10.1073/pnas.2008645117.

Pan, Hai, Stephanie M. Bilinovich, Parminder Kaur, Robert Riehn, Hong Wang, and David C. Williams. 2017. “CpG and Methylation-Dependent DNA Binding and Dynamics of the Methylcytosine Binding Domain 2 Protein at the Single-Molecule Level.” *Nucleic Acids Research* 45 (15): 9164–77. https://doi.org/10.1093/nar/gkx548.

Pan, Hai, Parminder Kaur, Ryan Barnes, Ariana C. Detwiler, Samantha Lynn Sanford, Ming Liu, Pengning Xu, et al. 2021. “Structure, Dynamics, and Regulation of TRF1-TIN2-Mediated Trans- and Cis-Interactions on Telomeric DNA.” *Journal of Biological Chemistry* 297 (3): 101080. https://doi.org/10.1016/j.jbc.2021.101080.

Pauling, Linus. 1935. “The Oxygen Equilibrium of Hemoglobin and Its Structural Interpretation.” *Proceedings of the National Academy of Sciences* 21 (4): 186–91. https://doi.org/10.1073/pnas.21.4.186.

Peng, Ran, and Dongqing Li. 2016. “Fabrication of Polydimethylsiloxane (PDMS) Nanofluidic Chips with Controllable Channel Size and Spacing.” *Lab on a Chip* 16 (19): 3767–76. https://doi.org/10.1039/c6lc00867d.

Persson, Fredrik, Joachim Fritzsche, Kalim U. Mir, Mauro Modesti, Fredrik Westerlund, and Jonas O. Tegenfeldt. 2012. “Lipid-Based Passivation in Nanofluidics.” *Nano Letters* 12 (5): 2260–65. https://doi.org/10.1021/nl204535h.

Peters, J.-M., Antonio Tedeschi, and Julia Schmitz. 2008. “The Cohesin Complex and Its Roles in Chromosome Biology.” *Genes & Development* 22 (22): 3089–3114. https://doi.org/10.1101/gad.1724308.

Phillips, Rob, Jane Kondev, Julie Theriot, Hernan G. Garcia, and Nigel Orme. 2012. *Physical Biology of the Cell*. *Physical Biology of the Cell*. Garland Science. https://doi.org/10.1201/9781134111589.

Pollard, Thomas D. 2010. "A Guide to Simple and Informative Binding Assays." Edited by Douglas Kellogg. *Molecular Biology of the Cell* 21 (23): 4061–67. https://doi.org/10.1091/mbc.e10-08-0683.

Puchner, Elias M., and Hermann E. Gaub. 2009. "Force and Function: Probing Proteins with AFM-Based Force Spectroscopy." *Current Opinion in Structural Biology* 19 (5): 605–14. https://doi.org/10.1016/j.sbi.2009.09.005.

Ramanoudjame, Gomathi, Mei Du, Kimberly A. Mankiewicz, and Vasanthi Jayaraman. 2006. "Allosteric Mechanism in AMPA Receptors: A FRET-Based Investigation of Conformational Changes." *Proceedings of the National Academy of Sciences* 103 (27): 10473–78. https://doi.org/10.1073/pnas.0603225103.

Razo-Mejia, M., J. Q. Boedicker, D. Jones, A DeLuna, J. B. Kinney, and R. Phillips. 2014. "Comparison of the Theoretical and Real-World Evolutionary Potential of a Genetic Circuit." *Physical Biology* 11 (2): 026005. https://doi.org/10.1088/1478-3975/11/2/026005.

Reisner, Walter, Jason P Beech, Niels B Larsen, Henrik Flyvbjerg, Anders Kristensen, and Jonas O Tegenfeldt. 2007. "Nanoconfinement-Enhanced Conformational Response of Single DNA Molecules to Changes in Ionic Environment." *Physical Review Letters* 99 (5): 058302. https://doi.org/10.1103/PhysRevLett.99.058302.

Reisner, Walter, Jonas N Pedersen, and Robert H Austin. 2012. "DNA Confinement in Nanochannels: Physics and Biological Applications." *Reports on Progress in Physics* 75 (10): 106601. https://doi.org/10.1088/0034-4885/75/10/106601.

Rhee, Sangkee, Robert G. Martin, Judah L. Rosner, and David R. Davies. 1998. “A Novel DNA-Binding Motif in MarA: The First Structure for an AraC Family Transcriptional Activator.” *Proceedings of the National Academy of Sciences* 95 (18): 10413–18. https://doi.org/10.1073/pnas.95.18.10413.

Riehn, Robert, Manchun Lu, Y.-M. Wang, Shuang Fang Lim, Edward C. Cox, and Robert H. Austin. 2005. “Restriction Mapping in Nanofluidic Devices.” *Proceedings of the National Academy of Sciences* 102 (29): 10012–16. https://doi.org/10.1073/pnas.0503809102.

Riehn, Robert, Walter Reisner, Jonas O. Tegenfeldt, Yan Mei Wang, Chih-Kuan Tung, Shuang-Fang Lim, Edward Cox, et al. 2008. “Nanochannels for Genomic DNA Analysis: The Long and the Short of It.” In *Integrated Biochips for DNA Analysis*, 151–86. New York, NY: Springer New York. https://doi.org/10.1007/978-0-387-76759-8_12.

Riggs, Arthur D, Suzanne Bourgeois, and Melvin Cohn. 1970. “The Lac Represser-Operator Interaction.” *Journal of Molecular Biology* 53 (3): 401–17. https://doi.org/10.1016/0022-2836(70)90074-4.

Rohs, Remo, Xiangshu Jin, Sean M. West, Rohit Joshi, Barry Honig, and Richard S. Mann. 2010. “Origins of Specificity in Protein-DNA Recognition.” *Annual Review of Biochemistry* 79 (1): 233–69. https://doi.org/10.1146/annurev-biochem-060408-091030.

Sfeir, Agnel, Settapong T. Kosiyatrakul, Dirk Hockemeyer, Sheila L. MacRae, Jan Karlseder, Carl L. Schildkraut, and Titia de Lange. 2009. “Mammalian Telomeres Resemble Fragile Sites and Require TRF1 for Efficient Replication.” *Cell* 138 (1): 90–103. https://doi.org/10.1016/j.cell.2009.06.021.

Sharma, Rajhans, Sriram KK, Erik D. Holmstrom, and Fredrik Westerlund. 2020. “Real-Time Compaction of Nanoconfined DNA by an Intrinsically Disordered Macromolecular

Counterion." *Biochemical and Biophysical Research Communications* 533 (1): 175–80. https://doi.org/10.1016/j.bbrc.2020.06.051.

Shi, Zhubing, Haishan Gao, Xiao-chen Bai, and Hongtao Yu. 2020. "Cryo-EM Structure of the Human Cohesin-NIPBL-DNA Complex." *Science* 368 (6498): 1454–59. https://doi.org/10.1126/science.abb0981.

Shore, David, and Robert L. Baldwin. 1983. "Energetics of DNA Twisting." *Journal of Molecular Biology* 170 (4): 957–81. https://doi.org/10.1016/S0022-2836(83)80198-3.

Smisek, David L., and David A. Hoagland. 1989. "Agarose Gel Electrophoresis of High Molecular Weight, Synthetic Polyelectrolytes." *Macromolecules* 22 (5): 2270–77. https://doi.org/10.1021/ma00195a048.

Smith, Hamilton O. 1979. "Nucleotide Sequence Specificity of Restriction Endonucleases." *Science* 205 (4405): 455–62. https://doi.org/10.1126/science.377492.

Starr, Barry D., Barbara C. Hoopes, and Diane K. Hawley. 1995. "DNA Bending Is an Important Component of Site-Specific Recognition by the TATA Binding Protein." *Journal of Molecular Biology* 250 (4): 434–46. https://doi.org/10.1006/jmbi.1995.0388.

Stormo, Gary. 2013. *Introduction to Protein-DNA Interactions : Structure, Thermodynamics, and Bioinformatics*. Cold Spring Harbor Laboratory Press.

Streng, Diana E., Shuang Fang Lim, Junhan Pan, Alena Karpusenka, and Robert Riehn. 2009. "Stretching Chromatin through Confinement." *Lab on a Chip* 9 (19): 2772. https://doi.org/10.1039/b909217j.

Strobl, Gert. 2007. *The Physics of Polymers*. *The Physics of Polymers: Concepts for Understanding Their Structures and Behavior*. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-68411-4.

Sun, Xiaoyin, Takao Yasui, Takeshi Yanagida, Noritada Kaji, Sakon Rahong, Masaki Kanai, Kazuki Nagashima, Tomoji Kawai, and Yoshinobu Baba. 2016. “Identifying DNA Methylation in a Nanochannel.” *Science and Technology of Advanced Materials* 17 (1): 644–49. https://doi.org/10.1080/14686996.2016.1223516.

Sweryda-Krawiec, Beata, Halagowder Devaraj, George Jacob, and James J. Hickman. 2004. “A New Interpretation of Serum Albumin Surface Passivation.” *Langmuir* 20 (6): 2054–56. https://doi.org/10.1021/la034870g.

Takai, Kaori K., Tatsuya Kibe, Jill R. Donigian, David Frescas, and Titia de Lange. 2011. “Telomere Protection by TPP1/POT1 Requires Tethering to TIN2.” *Molecular Cell* 44 (4): 647–59. https://doi.org/10.1016/j.molcel.2011.08.043.

Tegenfeldt, Jonas O., Christelle Prinz, Han Cao, Steven Chou, Walter W. Reisner, Robert Riehn, Yan Mei Wang, et al. 2004. “The Dynamics of Genomic-Length DNA Molecules in 100-Nm Channels.” *Proceedings of the National Academy of Sciences* 101 (30): 10979–83. https://doi.org/10.1073/pnas.0403849101.

Tegenfeldt, Jonas O., Christelle Prinz, Richard L. Huang, Robert H. Austin, Stephen Y. Chou, Edward C. Cox, James C. Sturm, and Han Cao. 2004. “Micro- and Nanofluidics for DNA Analysis.” *Analytical and Bioanalytical Chemistry* 378 (7): 1678–92. https://doi.org/10.1007/s00216-004-2526-0.

Valle, Francesco, Mélanie Favre, Paolo De Los Rios, Angelo Rosa, and Giovanni Dietler. 2005. “Scaling Exponents and Probability Distributions of DNA End-to-End Distance.” *Physical Review Letters* 95 (15): 158105. https://doi.org/10.1103/PhysRevLett.95.158105.

Verdun, Ramiro E., and Jan Karlseder. 2007. “Replication and Protection of Telomeres.” *Nature* 447 (7147): 924–31. https://doi.org/10.1038/nature05976.

Veverka, Pavel, Tomáš Janovič, and Ctirad Hofr. 2019. “Quantitative Biology of Human Shelterin and Telomerase: Searching for the Weakest Point.” *International Journal of Molecular Sciences* 20 (13): 3186. https://doi.org/10.3390/ijms20133186.

Wang, Y. M., Jonas O. Tegenfeldt, W. Reisner, R. Riehn, X.-J. Guan, Ling Guo, Ido Golding, Edward C. Cox, James Sturm, and Robert H. Austin. 2005. “Single-Molecule Studies of Repressor-DNA Interactions Show Long-Range Interactions.” *Proceedings of the National Academy of Sciences* 102 (28): 9796–9801. https://doi.org/10.1073/pnas.0502917102.

Watson, J. D., and F. H.C. Crick. 1953. “Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid.” *Nature* 171 (4356): 737–38. https://doi.org/10.1038/171737a0.

Wendt, Kerstin S., Keisuke Yoshida, Takehiko Itoh, Masashige Bando, Birgit Koch, Erika Schirghuber, Shuichi Tsutsumi, et al. 2008. “Cohesin Mediates Transcriptional Insulation by CCCTC-Binding Factor.” *Nature* 451 (7180): 796–801. https://doi.org/10.1038/nature06634.

Westra, K.L., and D.J. Thomson. 1995. “The Microstructure of Thin Films Observed Using Atomic Force Microscopy.” *Thin Solid Films* 257 (1): 15–21. https://doi.org/10.1016/0040-6090(94)06351-6.

Wood, Kathleen H., Brian S. Johnson, Sarah A. Welsh, Jun Y. Lee, Yue Cui, Elizabeth Krizman, Edward S. Brodkin, et al. 2016. “Tagging Methyl-CpG-Binding Domain Proteins Reveals Different Spatiotemporal Expression and Supports Distinct Functions.” *Epigenomics* 8 (4): 455–73. https://doi.org/10.2217/epi-2015-0004.

Zhang, C., P. G. Shao, J. A. van Kan, and J. R. C. van der Maarel. 2009. “Macromolecular Crowding Induced Elongation and Compaction of Single DNA Molecules Confined in a Nanochannel.” *Proceedings of the National Academy of Sciences* 106 (39): 16651–56.

https://doi.org/10.1073/pnas.0904741106.

Zhang, Yi, H.-H. Ng, Hediye Erdjument-Bromage, Paul Tempst, Adrian Bird, and Danny Reinberg. 1999. “Analysis of the NuRD Subunits Reveals a Histone Deacetylase Core Complex and a Connection with DNA Methylation.” *Genes & Development* 13 (15): 1924–35. https://doi.org/10.1101/gad.13.15.1924.

Zhu, Ping, Hongshan Guo, Yixin Ren, Yu Hou, Ji Dong, Rong Li, Ying Lian, et al. 2018. “Single-Cell DNA Methylome Sequencing of Human Preimplantation Embryos.” *Nature Genetics* 50 (1): 12–19. https://doi.org/10.1038/s41588-017-0007-6.