Scanning Probe Microscopy and Computational Simulation of Polymers and Biopolymers

A Dissertation presented

by

Jingyun Zhang

to
The Department of Physics
In partial fulfillment of the requirements for the degree of
Doctor of Philosophy
In the field of
Physics
Northeastern University
Boston, Massachusetts
April, 2011
2011

Jingyun Zhang

ALL RIGHTS RESERVED
Scanning Probe Microscopy and Computational Simulation of Polymers and Biopolymers

by

Jingyun Zhang

ABSTRACT OF DISSERTATION

Submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Physics in the Graduate School of Arts and Sciences of Northeastern University, April 2011
Abstract

In this thesis, polymer and biopolymer structure and dynamics were studied with scanning probe microscopy and computational simulation. In part 1, chapters 1-4, an image charge model was employed to simulate the effect of local dielectric response in electric force microscopy (EFM) of a glass polymer film. Local dielectric spectroscopy of the polymer films was studied with both simulation and EFM experiments. Part 2, chapters 5-12, scanning probe microscopy or atomic force microscopy (AFM) techniques and computational simulations were applied to study DNA flexibility enhancement by HMGB proteins. The study relies on a large amount of statistical works. HMGB protein DNA bending mechanism was under study. The experimental results show that HMGB proteins follow a static kink bending model. Protein aggregation structures: oligomers, cooperative binding modes and rigid filaments were also observed when the concentration of the protein was increased. Such structures made the analysis of the bends more complex. Finally, we conducted a computational simulation to quantify the bends by HMGB proteins. A Monte Carlo simulation was employed to generate bulk amount of simulated DNA molecules. Chi square statistical test was applied to quantify the bends by comparing the simulation with the AFM experiments. Appendix A, B and C are unpublished results or incomplete work on exploratory research. This work is preliminary, but some constructive ideas for future work have been produced.
Acknowledgements

It is a great pleasure to thank many people who help me make this thesis possible. This thesis represents the end product of many years of work.

It would be difficult to overstate my gratitude to my advisor Professor Nathan Israeloff. He provided me with the insight and resources necessary to pursue my Ph.D. He is always willing to help me out with difficulty on my road to pursue the degree. In my thesis writing he provided very sound advice and support.

I would also thank the members of my thesis committee, Professor Mark Williams and Professor Latika Menon for their useful suggestion and invaluable contribution to this thesis.

I also must special thank Dr. Micah McCauley, my collaborator. Without his help, I would have never finished my work. In addition, I would like to thank Professor Srinivas Sridhar for letting me use AFM in IGERT group and Dr. Dattatri Nagesha for his technical assistance.

The help and support from other people I worked with, side by side, during my graduate work ought to be recognized so I extend my sincere gratitude to all of them: Professor L. James Maher, III, Dr. Hassan Oukris, Dr. Philp S.Crider, M.R.Majewski and Dr.Yuhan Sun.

I would also like to thank my parents for their unconditional and unselfish support through my education. At last, I must thank my wife for her support, understanding and sacrifice for completing this thesis.
Contents

ABSTRACT ................................................................................................................. 3

ACKNOWLEDGEMENTS ......................................................................................... 5

CONTENTS ............................................................................................................... 7

Part 1 Local dielectric spectroscopy of polymer films

Chapter 1 Introduction ............................................................................................. 13

Chapter 2 Experimental methods ........................................................................... 15

Chapter 3 Theoretic model for computer simulation ............................................. 17
  3.1 Image charge force model .................................................................................. 17
  3.2 Capacitance model ............................................................................................ 20

Chapter 4 Results .................................................................................................... 21

Part 2 Visualizing protein-DNA complexes with AFM

Chapter 5 Introduction ............................................................................................. 33

Chapter 6 Worm-like chain (WLC) model ............................................................... 41
  6.1 DNA structure and WLC model ...................................................................... 41
  6.2 DNA deposition Model- 3D-2D transformation .............................................. 43
  6.2.1 Sample deposition process- 2D equilibration vs. kinetic trapping .......... 43
Chapter 11 Analysis of DNA flexibility enhancement by HMGB proteins by computational simulation

11.1 The motivation for simulation

11.2 DNA WLC model for simulation

11.3 Box-Muller algorithm-Gaussian random number generator

11.4 Knuth Algorithm-Poisson random number generator

11.5 Monte Carlo method to generate DNA chains

11.6 Analysis and simulation results for bare DNA chain

11.7 Analysis and simulation results for protein/DNA complexes

11.8 $\chi^2_{(k-1)}$ statistical analysis to determine the optimal protein-induced DNA bend angle

Chapter 12 Conclusion and discussion

12.1 Enhancement of DNA flexibility by HMGB proteins-Part 1

12.2 Enhancement of DNA flexibility by HMGB proteins-Part 2

12.3 Analysis of DNA flexibility enhancement by HMGB proteins by computational simulation

Appendix A. Non-equilibrium mechanics of active cytoskeletal networks

Appendix B. Local fluctuation dissipation spectroscopy for polymer (PVAC)

Appendix C. DNA relaxation process simulation
Part 1

Local dielectric spectroscopy of polymer films
Chapter 1

Introduction

Dielectric spectroscopy is a key tool for probing molecular dynamics in materials [1] such as glassy polymers, super-cooled liquids, ferroelectrics, epoxies, etc. The frequency spectrum of the complex susceptibility \( \varepsilon(\omega) = \varepsilon' - i\varepsilon'' \) contains information about the dynamics of dipolar and ionic degrees of freedom. It is also increasingly important to probe material properties with nanometer-scale resolution and in ultrathin materials or near interfaces. Recently there have been a number of experiments which demonstrate a significant reduction in the glass transition of ultrathin (20 nm) polymer films. The effects are mostly dramatic in freestanding high molecular weight polystyrene films [2]. In other cases where the films are constrained by a substrate, the shifts in \( T_g \) can be significantly smaller or reversed. In some materials such as polyvinyl acetate (PVAc), a \( T_g \) shift is undetectable [3] or small [4] depending on the technique, though changes in mechanical moduli were observed [3]. Standard dielectric susceptibility measurements [4] may have interfacial effects from one [5] to two electrodes such as a dramatic reduction in the glassy dielectric response. The mechanisms for the shifts in \( T_g \) are still debated but have focused on polymer chain confinement effects near a free surface [6] and/or the effect of a cooperative length scale [7] in glassy systems. Here we report the results of experiments in which a noncontact scanning probe microscopy technique is used to measure local frequency-dependent dielectric susceptibility in polymer films. The technique probes to the depth of 20 nm below a free
surface of bulk films and captures, with high resolution, both $T_g$ shifts and changes in the dynamics. In PVAc, a small reduction in $T_g$ and a narrowing of the spectrum of relaxation times are observed. Also from computational simulation, we compared the simulated results with experiment and further confirmed the validation of experimental results.

Fig 1.1 RHK UHV350 Scanning Probe Microscopy (SPM)
Chapter 2

Experimental methods

Electric force microscopy (EFM) is a scanning probe microscopy (SPM) method in which local electrostatic forces are probed. EFM has been used to image localized charges [8] on surfaces and dielectric constant variations [9]. In EFM a conducting cantilever is used so that dc and ac bias voltages can be applied to the tip - tip radius $R$ (25 nm). We use an ultrahigh-vacuum SPM with a variable temperature stage (RHK UHV 350); see Fig 2.1. Temperature was measured with a small thermocouple clamped to the sample surface. A noncontact frequency-modulation imaging mode is used. In this mode, the cantilever is oscillated at its resonance frequency, $f_{\text{res}} \sim 70$ kHz, using a Nanosurf EasyPLL phase-locked loop. Resonance frequency shifts ($f_{\text{res}}$) due to tip-sample interaction forces are detected with very high resolution using a Nanosurf EasyPLL detector, and this is used as a feedback parameter for controlling the tip-sample distance $z$. The samples are thin films (thickness $\sim 1$ $\mu$m of PVAc $M_w=167 000$), which have been prepared by spinning from a toluene solution, onto Au on glass, and annealed near $T_g=308$ K for 24 h vacuum. Typical dielectric susceptibility behavior was found, when measured in identically prepared bulk samples with a top (AuPd) electrode, using a Novocontrol dielectrometer. The Au film forms a back electrode which is grounded.
Fig 2.1 EFM tip-sample configuration
Chapter 3

Theoretic model for computer simulation

We used computer simulation to understand the mechanism of the tip-sample interaction to predict and compare with the experimental local dielectric spectroscopy.

3.1 Image charge force model

We used the infinite image charges methods to simulate the interaction between the SPM tip and the locally polarized sample (PVAC) surface. The local polarized PVAC surface can be substituted by infinite image charges.

Fig 3.1 Schematic graph of image charge model
The image charges are in dielectric media and in vacuum. The image charges in dielectric media are generated by the same order of charges in vacuum doing planar mirror imaging at the material and vacuum interface iteratively.

\[
\begin{align*}
Q_0' &= -(\varepsilon-1)/(\varepsilon+1) \cdot Q_0 \\
Q_1' &= -(\varepsilon-1)/(\varepsilon+1) \cdot Q_1 \\
\vdots \end{align*}
\]

Using the iterative process:

\[
Q_i' = -(\varepsilon-1)/(\varepsilon+1) \cdot Q_i
\]

Also:

\[
\begin{align*}
\Delta d_0' &= -z = -d_0 \\
\Delta d_1' &= -d_1 \\
\Delta d_i' &= -d_i \\
\end{align*}
\]

(eq 3.1)

And image charges in vacuum are generated by the previous order of charges in dielectric media doing spherical imaging relative to the sphere of the tip iteratively.

\[
\begin{align*}
Q_1 &= -R/(d_0-d_0') \cdot Q_0' \\
Q_2 &= -R/(d_1-d_1') \cdot Q_1' \\
\vdots \\
Q(i+1) &= -R/(d_i-d_i') \cdot Q_i'
\end{align*}
\]

Using the iterative process:

\[
\begin{align*}
\Delta d_1 &= d_0 - R^2/(d_0-d_0') \\
\Delta d_2 &= d_0 - R^2/(d_1-d_1') \\
\vdots \\
\Delta d(i+1) &= d_0 - R^2/(d_i-d_i') \\
\end{align*}
\]

(eq 3.2)

And, when tip oscillates, some of charges can follow the tip to oscillate while the others can’t because of complex dielectric response from PVAC film. Therefore, image charges can be divided into two categories: Fast response \(Q_3\) group and Static \(\Delta Q\) group.
The image charges ($Q_3$ group) which can follow the tip’s oscillation are in $\varepsilon=3$ dielectric material space, while the others image charges (Static $\Delta Q$ group) are in $\varepsilon=K$ dielectric material space. Therefore, the summation of interaction forces of these charges are:

$$F = \sum_{i,j} \frac{Q_{3(i)} Q_{3(j)}}{(d_{3(i)} - d_{3(j)})^2} + \sum_{i,j,k} \frac{\Delta Q_{(k,i)} Q_{3(j)}}{(d_{3(j)} - d_{(k,i)})^2} + \sum_{i,j,k} \frac{\Delta Q_{(k,i)} Q_{3(j)}}{(d_{3(j)} - d_{(k,i)})^2} + \sum_{i,j,k,m} \frac{\Delta Q_{(k,i)} \Delta Q_{(j,m)}}{(d_{(j,m)} - d_{(k,i)})^2}$$

(eq3.3)

Then, the variation of force can be connected to $V_{2\omega}$

$$V_{2w} = \frac{f_0}{4K} \left( \frac{dV_{bias}}{df} \right) \ast \left( \frac{dF}{dz} \right)$$

(eq3.4)

$$A = \left( \frac{dV_{bias}}{df} \right)$$

Where $A$ is the gain factor. $V_{bias}$ is applied voltage to tip. $df$ is feedback frequency to control the AFM to work in frequency modulation mode. $f_0$ is resonance frequency of AFM cantilever and $K$ is Hook constant for cantilever.

By calculating the second harmonic signal $V_{2\omega}$, we can get the local dielectric spectroscopy of the PVAC. When $K$ is a complex number, $\varepsilon=\varepsilon'+i\varepsilon''$, $F$ will be a complex number and therefore $V_{2\omega}$ will be complex number too.

In the experiment, the real and imaginary components of $V_{2\omega}$ can be directly measured. Therefore, experiment and simulation results can be compared.
3.2 Capacitance model

It is also possible to use a capacitance model to get $V_{2\omega}$,

$$C = \left( \sum_i Q_{3(i)} + \sum_{j,k} \Delta Q_{(j,k)} \right) / V_{bias}$$

$$F = \frac{1}{2} \left( \frac{dC}{dz} \right)^2 V_{bias}^2$$

$$V_{2w} = \frac{f_0}{4K} \left( \frac{dV}{df} \right) \left( \frac{dF}{dz} \right) = \frac{f_0}{4K} \left( \frac{dV}{df} \right) \left( \frac{d^2C}{dz^2} \right) V_{bias}$$  

(eq.3.5)
Chapter 4

Results

First, we can compare relation between $V_{2\omega}$ and $z_0$ (equilibrium position where tip interacts with PVAC surface) by image charge model for various $\varepsilon_p(\varepsilon)$ from 3.0 to 10.0, which represents various temperatures from room temperature (295K) to high temperature (322K) (Fig 4.1). Then at room temperature (295K). We tried to fit experimental results with our models. We find both image charges model, capacitance method and experimental result agree well. (Fig 4.2)

![Image Chart](image.png)

Fig 4.1 The second harmonic signal vs. $z_0$ for dielectric constant $\varepsilon_p(\varepsilon)$ from 3.0 to 10.0;
Different \( \varepsilon(\varepsilon) \) represents various temperatures from room temperature (295k) to high temperature (322K).

![Graph of \( V'2\omega \) vs. \( z_0 \) at a tip radius of 35 nm](image)

Fig 4.2 Compare of (Second Harmonic signal vs. \( z_0 \)) by image charge methods, capacitance methods and experiment at room temperature (295K).

A sphere image-charge model gives a result for tip radius, \( R=28\pm7 \) nm, and for the measurements that follow we determined that \( z_0=15\pm3 \) nm by fitting image charge method and capacitance method to experimental result at room temperature (295K).

Second, we obtained \( V'2\omega \) (real component) signal and \( V''2\omega \) (imaginary component) vs. frequency at various \( z_0 \) position, which represents the equilibrium position at various temperatures (Fig 4.3a,b).
Fig 4.3a: $V'\omega$ (real component) signal(a) and $V''\omega$ (imaginary component)(b) vs. frequency at various $z_0$ position, which represents to the equilibrium position at various temperatures.
Third, we can achieve good fits using $z_0 = 12$ nm, $d = 40$ nm, and $R = 17$ nm. We found, however, that using the bulk values of $\varepsilon$ gives smaller and broader imaginary component peaks than the experiment, as shown in Fig 4.4. We can get reasonable fits to the data at 314 K (Fig 4.4) and other temperatures only if we assume a narrower distribution of characteristic rates than in bulk. Using a Havriliak-Negami susceptibility model [10] with width and asymmetry exponents, $\alpha = 0.94$ and $\gamma = 0.60$ works best. In contrast, the bulk data can be fitted with $\alpha = 0.84$ and $\gamma = 0.5$. The Fig 4.4 inset plot of peak frequency versus temperature with Vogel-Fulcher fits shows that the local dynamics are shifted lower in temperature by 5–7 K and $T_g$ is shifted lower by 4±1 K [11].
Fig 4.4 Local real and imaginary components of $V_{2\omega}$ vs. frequency at 314 K for $df=40$ Hz and $V_{bias}=0.6$ Vrms. Also shown are the model outputs using bulk PVAc susceptibility measurements (322 K) (dashed lines) as inputs or susceptibility inputs which give the best fit (solid Lines). Inset: peak frequency of $\tan \theta = \varepsilon''/\varepsilon'$ for local and bulk measurements vs. temperature with Vogel-Fulcher fits (lines).
Next, the imaginary components of $V_{2\omega}$ measured for several temperatures near $T_g$, when normalized by peak value and peak frequency, can be collapsed quite well onto a single curve as shown in figure 4.5. This can be compared to the model calculation using the bulk $\varepsilon$ data as an input. Clearly this does not fit the measurements well, i.e. the bulk normalized peak is broader than the local normalized peak. Since we have contributions from both near surface and bulk layers we consider a two-component model. We find we can fit (Fig 4.5) the data best with an 85% contribution from a near-surface narrow distribution with $\alpha = 0.94$ and $\gamma = 0.56$, and a 15% bulk contribution whose peak is shifted one decade lower in frequency. This model can account for the narrowing near the peak but less narrowing in the low frequency wing [12].
Fig 4.5 Peaks in the imaginary component of local response \((V_{2\omega})\), normalized to peak frequency and amplitude, are plotted vs. frequency for several temperatures for \(\delta f = 55\text{Hz}\), \(V_{\text{bias}} = 1\text{Vrms}\). Model calculations (lines), using the bulk susceptibility data, and best fit susceptibility data, are shown on for comparison.
In conclusion, we have demonstrated a method for probing frequency-dependent dielectric spectroscopy on nanometer length scales with noncontact SPM. Imaging spatial-temporal variations[13] in dielectric susceptibility may also be possible. Shifts in $T_g$ and changes in dynamics relative to bulk behavior can be detected with very high resolution. We see a 4 K reduction in $T_g$ and a narrowing of the distribution of relaxation times in a 20 nm thick free-surface layer of bulk PVAc. We see no reduction in the glassy component of the dielectric response, as was seen in ultrathin polymers confined between [4] or on [5] metallic electrodes.
References


Part 2
Visualizing protein-DNA complexes with AFM
Chapter 5

Introduction

The effect of protein binding on DNA flexibility is very important for us to understand in the context of DNA transcription and replication. Transcription, or RNA synthesis, is the process of creating an equivalent RNA copy of a sequence of DNA. Both RNA and DNA are nucleic acids, which use base pairs of nucleotides as a complementary language that can be converted back and forth from DNA to RNA in the presence of the correct enzymes. During transcription, a DNA sequence is read by an RNA polymerase, which produces a complementary, anti-parallel RNA strand. Translation is the first stage of protein biosynthesis (part of the overall process of gene expression). In translation, messenger RNA (mRNA) produced in transcription is decoded to produce a specific amino acid chain, or polypeptide, that will subsequently fold into an active protein. Translation occurs in the cell's cytoplasm, where the large and small sub-units of the ribosome are located, and bind to the mRNA. The ribosome facilitates decoding by inducing the binding of tRNAs with complementary anticodon sequences to that of the mRNA (Fig 5.1).
Fig 5.1 Transcription and translation process in Prokaryote and Eukaryote Cells

(Figure by Micah J. McCauley)
In eukaryotic cells, Nucleosomes form the fundamental repeating units of eukaryotic chromatin (Fig 5.2).

Fig 5.2 chromatin structure: Histone protein are strongly alkaline proteins found in eukaryotic cell nuclei, which package and order the DNA into structural units called nucleosomes. They are the chief protein components of chromatin, act as spools around which DNA winds, and play a role in gene regulation.
HMGB proteins are a class of small proteins which widely exist in eukaryotic cells and sequence-non-specifically bind to DNA [1-3]. A common feature of these proteins is that they contain several duplications of an 80 residue “HMG” box domain such as NHP6A from budding yeast [4-8] and wild type human HMGB2 [9-13]. These proteins are found in relatively high concentration in eukaryotic cells with a ratio of one protein to five to ten nucleosomes in chromatin.

Among the striking features of HMGB proteins are their ability to bind to DNA and to induce a strong DNA bend [4] (fig 5.3). HMGB proteins appear to alter DNA structure through formation of hinges with enhanced flexibility [5-7]. The biological functions of HMGB proteins are not very clear yet. Current evidence shows HMGB proteins play multiple roles in chromatin structure and function [8]. Johnson and his collaborators first found that HMGB proteins can replace prokaryotic HU proteins in the role of bending DNA [9]. Research also has disclosed other important functions of these HMGB proteins, such as enhancing transcription activation through facilitating DNA binding and catalysis [10]. For example, NHP6A from budding yeast has been proven to facilitate Gal4p binding to DNA [11]. HMGB proteins act as a key component which is acting as a template for the transcription process in yeast [12]. HMGB proteins also play an important role in enhancing protein-protein interaction by bending DNA [13, 14].
Fig 5.3 Proposed HMGB protein function (Figure by Micah J. McCauley)
Four types of HMGB were under study in this work: wild type human HMGB2, NHP6A from budding yeast and artificially modified mutants of NHP6A [4]: Mutant 1 and Mutant 2. All four have sequences of approximately 100 amino acids (fig 5.4a). The principal difference between them is that NHP6A contains a long cationic leader sequence of 16 amino acids, which is absent HMGB2. Mutant 1 and Mutant 2 have 5 and 9 amino acid cationic leader sequences respectively (fig 5.4a). As shown in fig 5.4b, a common feature of HMGB group proteins is that they have an L-shaped portion with three α-helices (fig 5.4b). The L-shaped region touches the minor groove of DNA base pairs to widen the minor groove and partially insert one or more amino acid side chains between stacked base pairs [3], thus bending DNA significantly. Fig 5.4b shows molecular models of sequence-nonspecific architectural proteins: mammalian HMGB domain HMGB2 [15] and Nhp6A [16]. NHP6A contains a long cationic leader sequence (Fig 5.4b, right, magenta) which is located at the major groove to asymmetrically neutralize the DNA negatively charged phosphate backbone. The DNA bending by HMGB proteins may be involved in three types of interactions: minor groove widening, amino acid wedging, and asymmetric charge neutralization. Favorable electrostatic forces play a dominant role in HMGB nonspecific DNA binding and bending [26].
Fig 5.4 Proteins analyzed in this work. (a) Sequence alignments of wild-type (WT) HMGB2, chemical modified HMGB2 proteins (Mutant 1 and Mutant 2) and Yeast NHP6A protein. The length of N-terminal leader: NHP6A > Mutant 2 > Mutant 1 > HMGB2. (Figure by Emily M. Bystry) (b) Molecular models of sequence-nonspecific architectural proteins: mammalian HMGB domain HMGB2 [PDB entry 1ckt [15]], yeast HMGB protein Nhp6A [PDB entry 1j5n [16]]. Double strand DNA is colored with yellow and green, and protein is shown as a magenta ribbon. The cationic N-terminal leader of Nhp6A is the magenta strand indicated by the arrow. (Figure by RCSB PDB)
We want to understand the role of the cationic leader (N-terminal leader) of NHP6A and its mutants on DNA binding strength and bending ability. Thus, In NHP6A, Mutant 1, Mutant 1 and HMGB2, we can study the effect of systematic reduction of the size of the cationic leader. Recent research shows the cationic leader can enhance the protein binding ability [4]. Understanding how these HMGB proteins enhance DNA flexibility is very important for us to make clear the biological functions of these proteins. Previous research results suggest a static kink bending mechanism for HMGB proteins [2, 5]. Due to HMGB proteins sequence non-specific binding and bending on DNA, bulk ensemble-averaged measurements are needed to quantify protein-induced DNA bends. X-ray crystallography [17, 18] can give detailed information about the bends of protein-DNA crystals. However, in these structures two extra nonduplexed bases were incorporated, making an independent assessment of protein-induced bending impossible. Single-molecule techniques are well suited to resolve this issue. Atomic force microscopy (AFM) and optical tweezers are two of these techniques. In this study we use AFM to characterize HMGB proteins bending DNA mechanism at the single molecule level and compare with the results from optical tweezers and other experiments.
Chapter 6

Worm-like chain (WLC) model

6.1 DNA structure and WLC model

![Structure of DNA](image)

Fig 6.1 Schematic graph of DNA structure (Figure by Agilent Company and Dr. W Travis Johnson)
DNA consists of two long polymers of simple units called nucleotides, with backbones made of sugars and phosphate groups joined by ester bonds (negative charged). These two strands run in opposite directions to each other and are therefore anti-parallel. Attached to each sugar is one of four types of molecules called bases (A, T, C, G). It is the sequence of these four bases, along the backbone, that encodes information. The DNA chain is 22 to 26 Ångströms wide (2.2 to 2.6 nanometres), and one nucleotide unit is 3.3 Å (0.33 nm) long.

Schellman in 1974 (Schellman in 1974, Biopolymers) proposed a worm-like chain (WLC) model for DNA bending:

In two dimension, the bending angle probability density function for a separation is:

\[ N(\theta) = Ae^{-p\theta^2/2l} \]

where \( A \) is the normalized constant and \( p \) is the persistence length of DNA

Fig 6.2 Worm-like chain model for DNA
Persistence length $p$ is a basic mechanical property, quantifying the stiffness of a long polymer. Informally, for pieces of the polymer that are shorter than the persistence length, the molecule behaves rather like a flexible elastic rod, while for pieces of the polymer that are much longer than the persistence length, the properties can only be described statistically, like a random walk. Formally, $p$ is defined as the length over which correlations in the direction of the tangent are lost. We will use this parameter to quantify the flexibility of DNA polymer chain.

6.2 DNA deposition Model - 3D-2D transformation

6.2.1 Sample deposition process- 2D equilibration vs. kinetic trapping

An advantage of AFM (see Chapter 7) is that structures can be directly observed [19-22] which can complement other indirect measurements such as these obtained by optical tweezers, or magnetic tweezers [6, 23, 24]. A disadvantage is that the three-dimensional structures need to be condensed onto a two-dimensional substrate, typically mica. However, with proper preparation, substrate effects on properties such as persistence length have been successfully modeled and minimized. When depositing double-stranded DNA (dsDNA) or protein-nucleic acid complexes onto a mica surface, the result of the deposition depends on the nature and strength of the molecule-surface interactions. Bustamante et. al.[19] and Rivetti et. al.,[25] proposed that there exist two extremes in this process: equilibration and kinetic trapping. Equilibration occurs if the molecules can approach the substrate and are able to find an equilibrium configuration on
the mica surface during the deposition process. However, if kinetic trapping occurs during the deposition process, parts of the dsDNA or protein-DNA complex stick to the mica surface as they make contact, thereby getting trapped at their contact sites.

2D equilibration
Fig 6.3 Schematic graph of 2D equilibration vs. 3D kinetic trapping for bare DNA
For protein binding to DNA, the binding free energies for protein-DNA interactions are within 15-30 times the free energy $k_B T$ [19]. Fig 6.4 shows how protein DNA complexes deposit on a substrate surface for two deposition conditions. It is clear that in the 2D equilibration scenario the protein-induced DNA bend angle is unchanged, whereas in the 3D kinetic trapping case the angle may be altered by a strong molecule surface interaction force. Therefore, in order to obtain correct protein-induced DNA bend angles, the experiment should be done in 2D equilibration conditions[19].

**Protein-DNA complex in 2D equilibration**

\[
G_{\text{Bind}} = 15 - 30k_B T
\]
Fig 6.4 Schematic graph of 2D equilibration vs. Kinetic trapping for protein DNA complexes at protein binding site.
6.2.2 Surface dynamics for 2D equilibration

If two-dimensional equilibration occurs, it is nonetheless found that imaged molecules retain a persistence length quite close to that found in three dimensions in solution [19, 25], suggesting that molecular flexibility and local bend angles are negligibly affected by substrate effects. Whether equilibration occurs is determined by the characteristic time \( t \), required by the molecules to equilibrate on the mica surface relative to the deposition time before drying of the solvent. The characteristic time can be written as:

\[
 t = \frac{t_0}{\{1 - f_b [1 - \exp(\frac{\Delta G_{bind}}{k_B T})]^{-1}\}} 
\]

Here \( f_b \) is fraction of the monomers on the dsDNA polymers interacting with positively charged ion binding sites on the mica surface; \( \Delta G_{bind} \) is the binding energy of the molecule to each surface site; \( k_B \) is the Boltzmann constant and \( T \) is the deposition temperature, \( t_0 \) is the characteristic time for the dsDNA polymer to access its equilibrium configurations with free diffusion on the two dimensional surface.
\[ t = \frac{v^N}{N \rho_0} \]  

(6.2)

This equation is based on the assumption that a polymer molecule is made up of \( N \) identical units; \( v \) is number of degrees of freedom for a single unit; \( \rho_0 \) is transition rate of one unit among its degree of freedom in the absence of binding interactions with the surface.

Bustamante and Revetti [25] also pointed out that typically, for 5994 base-pair (bp) dsDNA fragments, the characteristic time, \( t \), is about 5 minutes and for shorter dsDNA fragments, as little as 1 minute is required to attain equilibrium. For the PBR322 dsDNA fragments with 4361 bp that we study here, the characteristic time \( t \) should be well under 5 minutes.
6.3 WLC model in deposition

Because DNA can find its equilibrium on 2D a substrate surface by thermal energy, the bending angle distribution $N(\theta)$ only occurs in one plane $\theta$.

If 3D kinetic trapping occurs, the whole process can be considered as 2 steps:

1. The sum of two independent 2D bending angle distribution density function in $\theta$ and $\phi$:

   $$N(\theta,\phi)=N(\theta) \cdot N(\phi)$$

2. Projection from 3D to 2D
6.4 WLC model in 2D equilibration

Here because we deposit DNA in 2D equilibration condition, the model WLC model for 2D equilibration case will be discussed.

Fig 6.7 Schematic graph of DNA in WLC model
First we define $\vec{s}$ and $\vec{s'}$ are two positions on the polymer chain. $u(\vec{s})$ and $u(\vec{s'})$ are the unit vector tangent to the polymer chain at position $\vec{s}$ and $\vec{s'}$. $l$ is the contour length of the polymer from position $\vec{s}$ to position $\vec{s'}$

$$l = |\vec{s} - \vec{s'}|$$ (6.3)

The bending angle between position $\vec{s}$ and $\vec{s'}$ is $\theta$. Therefore the scalar product of $u(\vec{s})$ and $u(\vec{s'})$ is $\cos(\theta)$

$$<\vec{u}(\vec{s}) \cdot \vec{u}(\vec{s'}) > = <\cos(\theta)>$$ (6.4)

From the worm like chain model in two dimensions, the probability function of the polymer chain bending angle $\theta$ is as follows:

$$N(\theta) = Ae^{-p\theta^2/2l}$$ (6.5)

Where $A$ is the normalizing constant, $p$ is the persistence length of the polymer.

Therefore, the mean cosine bending angle can be deduced in this way:

$$<\cos(\theta)> = \int_{-\infty}^{+\infty} \cos(\theta)N(\theta)d\theta = e^{-l/2p}$$

$$\Rightarrow <\vec{u}(\vec{s}) \cdot \vec{u}(\vec{s'}) > = e^{-l/2p}$$ (6.6)
And if we define the $R$ as the end to end distance of the overall polymer and $L$ is the whole contour length of the polymer, the relationship between them and persistence length of the polymer is:

$$< R^2 > = \int_0^L \int_0^L < \tilde{u}(s) \cdot \tilde{u}(s') > d\tilde{s}' d\tilde{s} = \int_0^L \int_0^L e^{-|s-s'|/2\rho} d\tilde{s}' d\tilde{s}$$

$$\Rightarrow < R^2 >= \frac{4p}{L} \left( 1 - \frac{2p}{L} \left( 1 - e^{-L/2\rho} \right) \right) \quad \text{(6.7)}$$

6.5 Protein induced DNA bending mechanisms

Two biophysical mechanisms were proposed for protein induced DNA bending, a static kink model and a flexible hinge model. The advantage of using AFM is that we can not only obtain the mean protein induced bend angle but also the angle distribution. The shape of the distribution can give us information about the mechanism of protein-induced flexibility in DNA. For example, an almost flat distribution of bend angles was found for the case of HU proteins bound to DNA, which provided strong evidence that protein binding formed a flexible hinge in that case [23], as opposed to a static kink (fixed angle) model.
The effect of HMGB binding on the flexibility of double-stranded DNA (dsDNA) was previously studied in optical tweezers experiments [26, 27]. It was found that dsDNA force-extension curves were strongly altered by the presence of HMG proteins. DNA Persistence length is a mechanical property quantifying its flexibility. In the presence of several different HMG proteins, the shape of the force-extension curves at low force indicated that the DNA persistence length was significantly reduced at high
protein concentrations relative to that observed in the absence of protein. The change in persistence length as a function of protein concentration was used to determine protein equilibrium association constants and the average protein-induced DNA bend angle. The model used to determine the average bend angle assumed that the bound protein induced a random, flexible hinge. However, no direct information about the flexibility of the protein-bond site or distribution of angles could be obtained.

6.6 Modified WLC model for protein-DNA complexes

Now we can further modify the worm like chain model with the protein binding case. When there are proteins binding to the homogeneous dsDNA polymer, singularities forms. In this model, the proteins divide the homogeneous to several pieces and each piece $l_i$ is still the homogenous structure. And each piece $l_i$ still maintains the same persistence length $p$ as the previous whole dsDNA polymer. We suppose each protein has a diameter of $\delta$ and there are average $N = \langle N_p \rangle$ proteins binding to a single dsDNA molecule. $N_p$ is the number of protein bound to each DNA molecule. And for the i’th protein which binds to dsDNA at position $k$ on the chain, it has an induced bending angle $\beta_i$. (fig 6.9)
Because each protein we used is only \(~10\) nm, only a couple of proteins bind to each individual dsDNA molecule which has a contour length \(~1.5\) μm (1500 nm). Even though the protein can divide the dsDNA into several pieces, we observed that each piece still has a contour length at least 300 nm long. Therefore, the length of protein can be neglected and can be only considered as a singularity.

When \(\delta \to 0\), the scalar product of the unit vector at the protein binding site is:

\[
<\ddot{u}(s) \cdot \ddot{u}(k_i - \delta/2) > <\ddot{u}(k_i - \delta/2) \cdot \ddot{u}(k_i + \delta/2) > <\ddot{u}(k_i + \delta/2) \cdot \ddot{u}(s') > \\
\delta \to 0 \\
= \cos \beta_i <\ddot{u}(s) \cdot \ddot{u}(s') >
\]

(6.8)
And the contour length of the divided pieces by the proteins can be written as:

\[ l_i = |\vec{k}_i - \vec{k}_{i-1}| \]  \hspace{1cm} (6.9)

Now we consider the relationship of the overall end to end distance of heterogeneous protein DNA complex to protein-induced bend angle and these divided pieces. The integral in equation (6.10) should be divided to several sections at the protein binding sites:

\[ < R^2_\beta > = (\int_0^{\tilde{k}_1} + \int_{\tilde{k}_1}^{\tilde{k}_2} + \ldots + \int_{\tilde{k}_N}^{\tilde{l}})(\int_0^{\tilde{k}_1} + \int_{\tilde{k}_1}^{\tilde{k}_2} + \ldots + \int_{\tilde{k}_N}^{\tilde{l}}) < \vec{u}(s) \cdot \vec{u}(s') > ds' ds \\
\]  \hspace{1cm} (6.10)

The equation (6.10) can be classified as two categories (eq. 6.11): the first category depicts the self-correlation of the each individual piece divided by the protein. And the second category gives the correlation between different two divided pieces. Here we only consider the correlation between two adjacent pieces \((i+1=j)\) and neglect others.

\[ = (\int_0^{\tilde{k}_1} \int_0^{\tilde{k}_1} + \int_{\tilde{k}_1}^{\tilde{k}_2} \int_{\tilde{k}_1}^{\tilde{k}_2} + \ldots + \int_{\tilde{k}_N}^{\tilde{l}} \int_{\tilde{k}_N}^{\tilde{l}}) < \vec{u}(s) \cdot \vec{u}(s') > ds' ds \\ + 2 \sum_{i,j=0}^{N} \int_{\tilde{k}_i}^{\tilde{k}_{i+1}} \int_{\tilde{k}_j}^{\tilde{k}_{j+1}} < \vec{u}(s) \cdot \vec{u}(s') > ds' ds \\
\]  \hspace{1cm} (6.11)
Then we can get the result as follows:

$$
= 4pL \left(1 - \frac{2p}{L} \sum_{i=1}^{N} (1 - e^{-l_i/p}) - \cos \beta_i (1 - e^{-l_i/p})(1 - e^{-l_i/p}) \right)
$$

Furthermore, if $l_i > 6p$, the term $e^{-l_i/p}$ can be neglected due to approaching zero. For the protein DNA complexes we observed, most of the protein divided pieces are more than 300nm ~ 6 times the persistence length $p$~50nm. Therefore, we neglect the $e^{-l_i/p}$

Finally, the equation (6.11) can be written as:

$$
\approx 4pL \left(1 - \frac{2pN}{L} (1 - \langle \cos \beta \rangle) \right)
$$

(6.12)
6.7 Correction to modified WLC model for the case of cooperative protein binding

In addition to individual protein binding to DNA, there sometimes exists cooperative protein binding to DNA to form a small rigid filaments on DNA. Especially for NHP6A protein bound to DNA, since protein-DNA complexes can only be seen at higher protein concentrations (x=1:9) compared to other three kinds of proteins cases, the cooperative binding can’t be avoided. One or more small rigid sections can be seen on some DNA. Therefore, these small rigid sections must be taken into account into the WLC model (fig 6.10).

Fig 6.10: Schematic map showing the cooperative binding and how to model it by using WLC model.

Here, we assume there only exists one straight contiguous section of cooperatively bound proteins on one individual DNA and its length $l_f$ which is much smaller than the
DNA contour length $L$ can be considered as a small perturbation which resides somewhere in an $L-l_f$ contour length DNA, see Fig 6.10.

Suppose the DNA has an entry angle of $\beta_{f1}$ and an exit angle of $\beta_{f2}$ to the rigid section. The rigid section is composed of several individual proteins which have diameter of $\delta$. And $\delta$ is very small compared to $l_f$. Then we found that as $\delta \rightarrow 0$, the correlation of adjacent DNA strand to the rigid section in the presence of protein is that the projection of protein induced bend angle times the correlation of adjacent DNA strand to the rigid section.

$$\lim_{\delta \rightarrow 0} <\bar{u}(s) \cdot \bar{u}(l_f - \delta / 2) >= \cos(\beta_{f1}) <\bar{u}(s) \cdot \bar{u}(l_f) >$$
$$\lim_{\delta \rightarrow 0} <\bar{u}(l_f) \cdot \bar{u}(l_f + \delta / 2) >= \cos(\beta_{f2}) <\bar{u}(l_f) \cdot \bar{u}(s') >$$

(6.13)

Where $\bar{u}(s)$ is a constant when $s$ is within the rigid fragment, which is different from the $\bar{u}(s)$ at other places on DNA. Therefore, it is very important to separate the integral in the rigid fragment from the entire DNA.

The entire integral can be divided into three pieces:

Self correlation of other parts of DNA; self correlation of this rigid fragment and correlation of the two above.
\[<R^2_{p}> = (\int_{k_1}^{k_2} + \int_{k_2}^{k_3} + \ldots + \int_{k_{i-1}}^{k_i} + \ldots + \int_{k_N}^{k_1}) <\bar{u}(s) \cdot \bar{u}(s')> ds' ds
\]

\[= (\int_{l_{f_1}}^{l_{f_2}} + \int_{l_{f_2}}^{l_{f_3}} + \ldots + \int_{l_{f_N}}^{l_{f_1}}) <\bar{u}(s) \cdot \bar{u}(s')> ds' ds
\]

\[+ 2(\int_{l_{f_1}}^{l_{f_2}} + \int_{l_{f_2}}^{l_{f_3}} + \ldots + \int_{l_{f_N}}^{l_{f_1}}) <\bar{u}(s) \cdot \bar{u}(s')> ds' ds
\]

\[+ \int_{l_{f_1}}^{l_{f_2}} \int_{l_{f_1}}^{l_{f_2}} <\bar{u}(s) \cdot \bar{u}(s')> ds' ds
\]

(6.14)

The self correlation of the this rigid fragment

\[\int_{l_{f_1}}^{l_{f_2}} \int_{l_{f_1}}^{l_{f_2}} <\bar{u}(s) \cdot \bar{u}(s')> ds' ds = \int_{l_{f_1}}^{l_{f_2}} \int_{l_{f_1}}^{l_{f_2}} 1 ds' ds = l_f^2
\]

(6.15)

The self correlation of other parts of DNA

\[\int_{l_{f_1}}^{l_{f_2}} \int_{l_{f_1}}^{l_{f_2}} 2 <\bar{u}(s) \cdot \bar{u}(s')> ds' ds = \int_{l_{f_1}}^{l_{f_2}} \int_{l_{f_1}}^{l_{f_2}} \frac{L}{2} ds' ds
\]

\[\approx 4p(L-l_f)(1-\frac{2pN}{L-l_f}(1-<\cos \beta>))
\]

(6.16)

The correlation of the two above, only correlations between the adjacent fragments are considered and other correlation can be neglected due to the reason specified, and

\[e^{-l_f/2}p << 1 \text{, So } e^{-l_f/2}p \text{ can be neglected.}
\]

61
By summarizing the result above, \( < R^2_\beta > \) can be corrected to the follows:

\[
< R^2_\beta > \approx 4 p(L - l_f)(1 - \frac{2pN}{L-l_f}(1 - < \cos \beta >)) + 4 pl_f (\cos \beta_{f_1} + \cos \beta_{f_2}) + l_f^2
\]

\[
\approx 4 pL(1 - \frac{2pN}{L}(1 - < \cos \beta >)) + 4 pl_f (\cos \beta_{f_1} + \cos \beta_{f_2} - 1) + l_f^2
\]

Therefore, compared with equation (6.12), an additional term:

\[4 pl_f (\cos \beta_{f_1} + \cos \beta_{f_2} - 1) + l_f^2,\]

should be added to the mean square DNA end to end distance.
Chapter 7

Material and Methods

7.1 Sample preparation

Freshly cleaved mica is used as the substrate to deposit dsDNA and protein-nucleic acid complexes. Mg\(^{2+}\), a divalent cation, was used in the deposition buffer to promote the adhesion of negatively charged DNA to the mica surface and prevent binding of HMGB proteins [28, 29] because freshly cleaved mica is negative charged [19, 28].

Fig 7.1 Schematic graph showing how to immobilize DNA on mica surface (Figure by Agilent Company and Dr. W Travis Johnson)
DNA:

Isolated plasmid pBR322 (Fermentas) was linearized by digestion with PvuII (Fermentas) followed by phenol extraction. The DNA was diluted with 10mM Tris–HCl or 10mM Hepes (pH 8.0), 10mM MgCl2 to 231nM per base pairs(bp) to avoid aggregation[30].

Five types of HMGB proteins:

HMGB1, HMGB2, Yeast NHP6A, Mutant 1 and Mutant 2 were used to bind DNA.

All of these proteins are belonging to HMGB proteins, the differences among them are structure differences, which are introduced in chapter 5.

Proteins were purified in[6] and initial protein solutions are:

HMGB2: 376nM, NHP6A: 36uM, Mutant 1: 64um, Mutant 2: 47uM with 20mM- Hapes, PH7.5; 5% glycerol; 100mM KCL; 1mM EDTA and 1mM DT.

Because the initial protein concentrations are so high, the above protein solutions were diluted by 10mM Tris–HCl or 10mM Hepes(pH 8.0), 10mM MgCl2 buffer to 200-1000 times lower. The diluting makes the final protein solutions ready to react with DNA are HMGB1 18nM, 1.9nM for HMGB2, 72nM for NHP6A, 64nM for Mutant 1 and 47nM for Mutant 2.
Sample deposition involved the following steps

(i) Diluted DNA and protein solution above were mixed and reacted for approximate two minutes. Optimal samples for imaging and analysis required moderate concentrations of both protein and DNA. We can define x, the ratio of the concentration of protein solution to the concentration of DNA base pairs to depict the mixed solution concentration of Protein-DNA complexes.

(ii) 5ml the reaction solution was deposited on freshly cleaved muscovite mica (Ted Pella Inc.) surface for 10 mins to obtain DNA equilibrium[19, 31].

(iii) The surface was rinsed with 5ml distilled water and then air dried for 10–15 min, after which excess water was removed by careful blotting.

(iv) Samples were then imaged immediately.

7.2 Experimental Setup

Atomic Force Microscopy (AFM) consists of a cantilever with a sharp tip (probe) at its end that is used to scan the specimen surface (fig 8.2a). The cantilever is typically silicon or silicon nitride with a tip radius of curvature on the order of nanometers. When the tip is brought into proximity of a sample surface, forces between the tip and the sample lead to a deflection of the cantilever according to Hooke's law. Depending on the situation,
forces that are measured in AFM include mechanical contact force, van der Waals forces, capillary forces, chemical bonding, electrostatic forces, magnetic, etc. Typically, the deflection is measured using a laser spot reflected from the top surface of the cantilever into an array of photodiodes. If the tip was scanned at a constant height, a risk would exist that the tip collides with the surface, causing damage. Hence, in most cases a feedback mechanism is employed to adjust the tip-to-sample distance to maintain a constant force between the tip and the sample. Traditionally, the sample is mounted on a piezoelectric tube, which can move the sample in the z direction for maintaining a constant force, and the x and y directions for scanning the sample. Alternatively a 'tripod' configuration of three piezo crystals may be employed, with each responsible for scanning in the x,y and z directions. This eliminates some of the distortion effects seen with a tube scanner. The AFM can be operated in a number of modes, depending on the application. In general, possible imaging modes are divided into static (also called contact) modes and a variety of dynamic (or non-contact) modes where the cantilever is vibrated such as tapping mode and non-contact mode (attractive force mode).
Fig 7.2a Setup of AFM (Figure by Agilent Company and Dr. W Travis Johnson)

AC Mode Imaging

- Piezoelectric transducer shakes cantilever holder at or near resonant frequency of cantilever
- Dynamic in x, y, & z
- Intermittent contact
- Interaction with sample reduces oscillation amplitude. Reduction in amplitude used as feedback signal
- Soft surfaces stiffened by viscoelastic response
- Impact forces predominately vertical
- Large vertical force, small lateral force
- Higher lateral resolution than contact mode

Fig 7.2b AC mode imaging (Figure by Agilent Company and Dr. W Travis Johnson)
We used a Pico-Plus Atomic Force Microscopy (AFM) from Agilent Technology (Fig 7.2a). The AFM was operated in tapping or intermittent contact mode in air (Fig 7.2b). Tapping mode has been used widely for imaging soft biological samples. In this mode, the cantilever is driven at a fixed frequency near its resonance frequency as it scans the sample. The tip is allowed to make transient contact with the sample surface at the bottom of the oscillation, which reduces its oscillation amplitude. The amplitude is used as a height feedback control parameter. The height, controlled by a piezo-crystal voltage, is extracted during scanning to form topography images. In addition, the phase of the oscillations relative to drive is also used to form images. The resolution in tapping mode can be nearly as good as in contact-mode, which is much more damaging to soft samples. Background cantilever thermal noise is inversely proportional to the resonance frequency [32]. Also, a cantilever with higher resonance frequency allows for a faster scanning rate. It should also be taken into account that polymer surfaces become stiffer at higher frequencies. This property further reduces the possibility of sample damage when using a high resonant frequency cantilever. Therefore, cantilevers with the highest resonant frequency are preferable. Nanosensors silicon supershar AFM tips (SSS-NCHR) were employed (resonance frequency 190-330 kHz; spring constant 40 N/m). Tips typically have 2nm radius of curvature, limiting lateral imaging resolution to 15nm under the best conditions. The scan range was either 2 um x 2um at 256 x 256 pixels or 512 x 512 pixels. The scan rate was typically 4 lines per second. Both topography and phase images were analyzed. DNA contours were traced semi-automatically by using ImageJ.
software from NIH with the NeuronJ tracing plug-in [33, 34]. A tracing step from 1 pixels to 10 pixels can be selected to trace DNA strand. Proteins were indentified and protein heights were measured by the Gwyddion software (AFM image processing software) by carefully setting a height threshold. Then protein induced DNA bend angle was measured by using the tracing contours from image J software to locate the entry and exit strands where DNA are bound by proteins.

Fig 7.3 Nanosensors silicon supersharp AFM tip (SSS-NCHR) (Nanosensors Inc.)
Chapter 8

AFM experimental results for bare DNA

8.1 DNA contour length and local bend angle

Bare PBR322 dsDNA for New England Biolabs, which has 4361 base pairs was initially imaged (fig 8.1). The distribution of contour lengths obtained from 250 molecules by image tracing (fig 8.2a) had a mean of 1.48 μm, and standard deviation 0.11 μm, giving 0.34±0.03 nm per base pair, which agrees very well with the theoretical value of 0.34 nm per base pair of B-form dsDNA [35]. The local bend angle is defined as the angle between the pair of adjacent line segments formed with three neighboring points along the tracing contour extracted by image J software, as shown in Fig. 1b. Bend angles were measured for all points on the bare DNA molecules. The distribution of local bend angles was Gaussian like [36], with a mean of 0° and σ ~22° (see Fig. 8.2b). However, resolution effects give an apparent width of DNA strands of approximately 15 nm by assuming the DNA cross section as a Gaussian distribution and the width at σ. (see fig 8.1d), but the theoretical width is only 2 nm. This discrepancy may come from two aspects from AFM experiment: One aspect is due to the tip radius. Most of tips have more than 10 nm tip radius. Thus convolution effect may play an important role in limiting the observed DNA strand size to be more than 10 nm. The other aspect may come from the liquid layer covered on top of DNA stand. According to theoretical value,
the cross section of DNA strand should be near circular and it should has a 2 nm height as well, but actually we measured DNA stand height is only 0.2-0.3 nm. Such a 15 nm width causes measurement error in the angle measurement, particularly for small segment sizes. The image tracing algorithm statistically picking a point along DNA contour with a Gaussian distribution for DNA cross section affects this error somewhat, hence the error is found by fitting simulated DNA contours. We estimated this effect contributes an additional $\sigma' \sim 8^\circ$ for 10 nm segments. Therefore, the intrinsic DNA bend angle distribution should have a width: $\sigma_0 = \sqrt{\sigma^2 - \sigma'^2} \sim 20.5^\circ$. For smaller segment sizes the angle error increases, as we will see in analysis of persistence length discussed below.
Figure 8.1. (a) shows a 2 µm image of the bare PBR322 dsDNA, which has contour length ~1.5 µm. (b) shows a 3D view of a single dsDNA molecule from (a). (c) shows how the local bend angle was calculated by the angle between two adjacent line segments drawn between three adjacent points (step size 10 nm) along the contour. (d) shows resolution effects (convolution effect) give an apparent width of DNA strands of approximately 15 nm(w), which causes measurement error in the angle measurement, particularly for small segment sizes.
Figure 8.2(a) shows the distribution of the contour lengths of bare dsDNA and its Gaussian fit. (b) shows the local bend angle distribution for bare dsDNA at segment step size 10 nm and a fit to a Gaussian distribution.
8.2 Persistence length determination

In order to determine the persistence length, the mean-squared end-to-end distance, \( R^2 \), of individual dsDNA strands was measured. Because dsDNA is a semi flexible polymer [37, 38] the Worm Like Chain (WLC) model [39, 40] can be used. With the assumption of equilibration, we used the two-dimensional WLC model to determine the persistence length [19, 25] (Eq.6.7):

\[
\langle R^2 \rangle = 4pL \left[ 1 - \frac{2p}{L} \left( 1 - e^{-L/2p} \right) \right]
\]

Where \( \langle R^2 \rangle \) is the mean squared end to end distance, \( p \) is the persistence length of the polymer, \( L \) is the polymer’s contour length. The mean squared end to end distance \( \langle R^2 \rangle \) was measured to be 0.32±0.3um². We further applied Monte Carlo simulation to generate 10,000 the same contour length DNA chain. Then we compared the \( R \) distribution obtained by the above two methods (please see Chapter 10).

Using the above equation, we find the persistence length is 57±6nm. This agrees very well with Rivetti’s [25] result, which gives 56nm using the WLC model at this contour length and agrees well with values found in solution [41].

We also used another method to obtain the persistence length of the dsDNA by measuring the local angle distribution of the bare dsDNA. This method was first
proposed by Landau and Lifshitz in 1958 [42]. Other research groups such as Rivetti et.
al. in 1996 [19, 25] and Abels et. al. in 2004 [43] deduced their experimental results with
a similar method. By tracing the dsDNA contours, we calculated the local bend angle $\theta$ of
two consecutive segments at different step size $l$. Using the probability distribution
function for the worm like chain model in two dimensions it can be shown that (Eq.6.6)

$$\langle \cos(\theta) \rangle = e^{-l/2p}$$

The above equation can be transformed to the following equation by take the
natural logarithm on both sides:

$$-\ln(<\cos \theta>) = l/2p$$

From a linear curve fit of this equation to the data, the slope $l/2p$, is extracted.
Angle measurement errors at small $l$, and excluded volume effects at large $l$ mean that a
good linear fit is obtained only at intermediate values [25, 44]. We find step sizes of 12
nm < $l$< 26 nm are optimal for determining $p$. The persistence length obtained by this
method was $p = 49.5$ nm with error (+3.5,-3.5) nm, which agrees within errors with the
value found by the end to end distance method, and with 48 nm expected from solution
studies. This illustrates the near equivalence of locally measured and globally measured
polymer properties. In fact, the locally measured value should be more reliable than end
to end distance method because by choosing a limited range of the step size we have
avoided the excluded volume effect that occurs for longer polymers.
Surface effects must be considered when analyzing the protein-induced DNA bend angles. Bustamante et. al. and Revetti, et. al. [19] [25] pointed out that the analysis is reliable only if the deposition process itself does not affect the conformation of the protein-induced DNA complexes. Under equilibration conditions, the binding free energies for the proteins to DNA are typically 15-30 times larger than the DNA-surface binding energies which are of order $k_B T$. Therefore, DNA can attain accessible surface configurations via thermal energy without altering the protein-induced DNA bend angles. Alternatively, for kinetic trapping conditions, as discussed above, the surface binding may be strong enough to alter the protein-induced DNA complex conformation, including
the protein-induced DNA bend angle. Thus, bend angle analysis is meaningful when conditions are favorable for equilibration, as was the case for the samples studied here.
Chapter 9

Enhancement of DNA flexibility by HMGB proteins-Part 1

9.1 Protein-induced DNA bend angle distributions

In figure 9.1, images of DNA with HMGB proteins bound are shown. As can be seen in the height profile in Fig. 9.1b, the bound proteins are significantly taller than the bare DNA, by ~ 0.5 nm typically, and thus they appear as light colored blobs in the images. More images are shown in figure 9.2, along with three-dimensional images, which clearly show the proteins as relatively higher spikes. Thus, protein binding sites were identified visually by this significant height difference. The protein–induced DNA bend angles were then measured at the protein binding sites and the distribution of these angles was determined.
Fig 9.1 In (a) we see the image for 3 individual HMGB protein-DNA complex’s molecules. In (b) height information for the two lines shown on the image: slice 1 crosses the only the dsDNA without protein. We can see the peak height for an image slice across the dsDNA is no more than 0.1 nm. Slice 2 crosses two proteins and we can see the peak heights for protein are more than 0.55 nm.
Figure 9.2 (a) and (c) shows single DNA molecules selected from fig 9.1a. (b) and (d), which show three-dimensional images of these molecules showing proteins as spikes.
Fig. 9.3 shows the distribution of protein-site bend angles for (a) HMGB (Box A) and (b) HMGB (Box A+B). From these two distributions, the analysis shows for HMGB (Box A) protein, that it induces an average bend angle of 78° with a standard deviation of 23°. And HMGB (Box A+B) exhibits an average bend angle of 67° with a standard deviation of 21°. These moderately broad distributions are much narrower than those found for HU binding [23]. For comparing these average angles with other measurements we can determine, using the standard error, that the statistical error bar on the averages was ±1.3°.
Fig 9.3 HMGB(Box A) (a) and HMGB(Box A+B) (b) protein site bend angle distributions are shown
9.2 Using the WLC model to estimate average protein-induced bend angles

A second method, based on the global average flexibility, was also used to obtain the average local bend angle. Since the structure of a protein-DNA complex is heterogeneous, the WLC model (equation 6.7) [45] for a homogeneous structure is no longer valid. We may quantify the relationship between the persistence length, \( p \), of the bare DNA, the end to end distance of the protein-induced DNA complexes \( <R^2> \), and the protein-induced bending angle, \( \beta \), by using a model based on the Worm Like Chain model with intrinsic bends developed by Rivetti [46] for the 3D case. In that model, the distribution of local protein-induced bend angles and the background \( p \) for bare DNA are known inputs. Then the average \( R^2 \) can be calculated. We have adapted this approach and derived a relation for 2D (Eq. 6.12). However we use only the average of protein-site cosine of the bend angle, the average number of bound proteins, \( Np \), and \( p \). Then the average \( R^2 \) can be found (see chapter 6):

\[
\langle R^2_\beta \rangle \approx 4pL \left( 1 - \frac{2p <Np>}{L} \left( 1 - <\cos \beta > \right) \right)
\]

We counted the number of proteins bound to each individual dsDNA molecule. The distributions of \( Np \) for the two proteins are shown in Fig 9.4. The average number of proteins bound per DNA molecule is 3 for HMGB (box A)-DNA complexes and 4 for HMGB (Box A+B) -DNA complexes. A Gaussian fit gave \( \sigma \approx 1 \) for both cases. We are then able to use the measured \( \langle R^2_\beta \rangle \) and equation 6.12 to calculate the average bend angle. For the HMGB (Box A) protein DNA complexes \( \langle R^2_\beta \rangle = 0.27 +/- 0.04 \mu m^2 \) and for
the HMGB (Box A + B) protein DNA complexes $\langle R^2 \rangle = 0.26+/0.05 \mu m^2$. By this method we found the mean protein-site bend angle $<\beta> = 82\pm8^\circ$ and $74\pm16^\circ$ for HMGB (Box A) and HMGB (Box A+B), respectively.

(a)

![Graph](image1.png)

(b)

![Graph](image2.png)

Fig 9.4(a)(b) Distribution of the number of binding proteins per dsDNA molecule.
Chapter 10

Enhancement of DNA flexibility by HMGB proteins-Part 2

10.1 Protein aggregating structures-protein oligomerization and cooperative binding

It is well known that proteins can aggregate to form oligomers. Oligomers widely exist in nature. A variety of mechanisms exist behind protein oligomerization [47-49]. Oligomers were observed in our AFM experiments. Both oligomers and individual proteins were observed to bind to DNA and enhance DNA flexibility. From high resolution AFM images, the relation between protein oligomer sizes and their corresponding ability to enhance DNA flexibility was studied. A protein can also bind to DNA and enhance the probability of other proteins binding to DNA in the vicinities of this protein. This is called cooperative binding [50], a mechanism for saturated protein binding to DNA. This phenomenon is typically seen at higher protein concentrations for HMGB (protein to DNA base pairs ratio in solutions x>1:10) [23, 50-52].

10.2 Experimental results

We first used bare DNA to confirm the DNA can be well equilibrated on the substrate surface (fig 10.1a). Then, we mixed the protein with DNA solution to x=1:68 for HMGB2, x=1:37 for Mutant 1 and x=1:50 for Mutant 2. In these concentrations the protein-DNA complexes are easily been seen. Fig 10.1b,c,d,e show 2um x 2um scale AFM image for HMGB2, Mutant 1 and Mutant 2 binding to DNA respectively. Higher
order protein oligomers were observed. However, for HMGB 2, the protein size is more homogeneous and small. This phenomenon can be confirmed by the protein height distributions shown in Fig 10.3. For NHP6A, if the protein concentration is very low \( x \approx 1:100 \), binding to DNA could not be observed as was the case for HMGB2. Therefore, we had to increase the \( x \) until 1:9 in order to see protein binding to DNA. Fig 10.1c shows a 2 \( \mu \)m x 2 \( \mu \)m scale AFM image for this \( x \). Various size oligomers bound to DNA were observed. In addition, cooperative binding regions of filamentary structure were observed at this concentration (fig 10.2).
Fig 10.1a,b,c,d,e: 2μm x 2μm AFM image of bare DNA and Protein-DNA structures at protein concentration for HMGB2 (x=1:68), NHP6A (x=1:9), Mutant 1 (x=1:37) and Mutant 2 (x=1:50) protein respectively. x is the ratio of concentration of protein solution.
to concentration of DNA base pairs to depict the mixed solution concentration of protein-DNA complexes.

Fig 10.2: 2 μm x 2 μm AFM image showing cooperative binding exists for NHP6A protein at lower protein concentration (x=1:9).
10.3 Protein height distribution

Fig 10.3 shows protein height distributions for HMGB 2, yeast NHP6A, Mutant 1 and Mutant 2. A weighted moving average was generated and optimized at 7 or 9 points to smooth the discrete raw data. The distribution exhibits several strong peaks, which were fit with a sum of Gaussians by minimizing $\chi^2$ (fig 10.3e,f,g,h) [48, 49, 53].

From the height distributions, it is notable that for HMGB 2 the height distribution is weighted heavily toward the low-height end, as compared with the other three. This suggests that the cationic leader, which is not present in HMGB2, may play a role in promoting protein aggregation in the other three proteins. The tendency toward aggregation would cause fewer single proteins to be present in solution. This could explain the lack of bound proteins at low concentration observed for NHP6A.

From the Gaussian fits (fig 10.3), peak position were found at 1.1 nm, 2.0 nm, 3.1 nm, 4.1 nm and 7.2 nm for HMGB2; 1.2 nm, 2.1 nm, 2.9 nm, 3.7 nm, 4.5 nm and 5.1 nm for NHP6A; 1.6 nm, 2.1 nm, 2.9 nm, 3.7 nm, 4.5 nm and 5.1 nm for Mutant 1 and 0.9 nm, 2.6 nm, 3.3 nm and 4.2 nm for Mutant 2. However, for Mutant 2 the expected 1.8 nm peak in the protein height distributions absent. Fig 10.3i further show the relation between average numbers of protein oligomer aggregation and length of N-terminal leader in four types of proteins. As the length of N-terminal leader increases, the average numbers of protein oligomer aggregation increases. However, the Mutant1 shows an exception and has a highest average numbers of oligomer aggregations. Next, a relationship between protein height and oligomer volume ($v_0$) was studied and a power
function was used to fit the height vs. volume data. For four types of protein, the volumes all have a near power of three relationships with their heights by the height volume relationship from Fig 10.4 and protein volumes at their corresponding protein featured heights can be deduced in Table 10.1.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HMGB 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>h (nm)</td>
<td>1.1</td>
<td>2.0</td>
<td>3.1</td>
<td>4.1</td>
<td>7.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$v_0$ (nm³)</td>
<td>135</td>
<td>541</td>
<td>1487</td>
<td>3071</td>
<td>5652</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>1</td>
<td>4</td>
<td>11</td>
<td>23</td>
<td>89</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>NHP6A</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>h (nm)</td>
<td>1.2</td>
<td>2.1</td>
<td>2.9</td>
<td>3.7</td>
<td>4.5</td>
<td>5.1</td>
<td></td>
</tr>
<tr>
<td>$v_0$ (nm³)</td>
<td>144</td>
<td>711</td>
<td>1925</td>
<td>3966</td>
<td>7050</td>
<td>10242</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>1</td>
<td>5</td>
<td>13</td>
<td>27</td>
<td>48</td>
<td>71</td>
<td></td>
</tr>
<tr>
<td><strong>Mutant 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>h (nm)</td>
<td>1.8</td>
<td>2.9</td>
<td>4.1</td>
<td>4.7</td>
<td>5.4</td>
<td>6.1</td>
<td>6.7</td>
</tr>
<tr>
<td>$v_0$ (nm³)</td>
<td>376</td>
<td>2021</td>
<td>6580</td>
<td>10967</td>
<td>17941</td>
<td>26242</td>
<td>36415</td>
</tr>
<tr>
<td>n</td>
<td>1</td>
<td>5</td>
<td>17</td>
<td>29</td>
<td>47</td>
<td>69</td>
<td>96</td>
</tr>
<tr>
<td><strong>Mutant 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>h (nm)</td>
<td>0.9</td>
<td>1.8*</td>
<td>2.6</td>
<td>3.3</td>
<td>4.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$v_0$ (nm³)</td>
<td>287</td>
<td>2295</td>
<td>6643</td>
<td>13221</td>
<td>25029</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>1</td>
<td>8</td>
<td>23</td>
<td>46</td>
<td>87</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

h: featured height of protein oligomer; $v_0$: protein volume; n: number of proteins to form protein oligomer

*: the expected data but not got from the Gaussian fit for protein height distribution analysis.

Table 10.1: Protein oligomer featured heights with their correspondent zero height base volumes and number of individual protein of oligomer.
Fig 10.3a,b,c,d: Protein height distribution graph for HMGB2, NHP6A, Mutant 1, and Mutant 2 protein respectively and a move average is employed to smooth the discrete distributions. Fig 10.3e,f,g,h: Multiple Gaussian curves to fit protein height distribution graph from fig 10.3a,b,c,d: Featured heights are 1.1 nm, 2.0 nm, 3.1 nm, 4.1 nm and 7.2 nm for HMGB2 protein (fig 10.3a,e); 1.2 nm, 2.1 nm, 2.9 nm, 3.7 nm, 4.5 nm and 5.1 nm for NHP6A protein (fig 10.3b,f); 1.6 nm, 2.1 nm, 2.9 nm, 3.7 nm, 4.5 nm and 5.1 nm for Mutant 1 protein (fig 10.3c,g); 0.9 nm, 2.6 nm, 3.3 nm and 4.2 nm for Mutant 2 protein (fig 10.3d,h). However, it is surprising that for Mutant 2 protein the expected 1.8 nm peak in the protein height distribution chart does not show in the Gaussian fit (fig 10.3h). Fig 10.3i: relation between the average numbers of protein oligomer aggregation and the length of the N-terminal leader in four types of proteins.
Fig 10.4a,b,c,d: Relation between protein zero height base volume with their heights and its power fit for HMGB2, NHP6A, Mutant 1, and Mutant 2 protein respectively. All four graphs show protein zero height base volume increases by the power of around three with its height.
10.4 Protein-induced DNA bending angle distributions and analysis

An advantage of AFM is that the protein-induced DNA bend angle can be directly visualized [19, 20]. Fig 10.5 is protein induced DNA bend angle distributions for HMGB2, yeast NHP6A, Mutant 1 and Mutant 2 respectively. Protein induced DNA bend angles were measured directly at the protein binding site. The statistics are based on typically 250 bend angles for each case. The bend angle distributions for HMGB 2 and NHP6A have a mean angle of 69.9±2.3° and 59.5±2.5° with a standard deviation of 30.3° and 36.5° respectively. Mutant 1 and Mutant 2 are quite similar. Mutant 1 induces a mean angle of 42.3±2.3° with a standard deviation of 34.9° whereas Mutant 2 induces a mean angle of 36.6±2.1° with a standard deviation of 36.6°. However, both Mutant proteins’ distributions are different from those of HMGB2 and yeast NHP6A. The Mutant proteins’ distributions show a narrow peak at near 0° and a second peak at ~40°. This suggests a portion of Mutant proteins only cover the DNA and do not induce any bend angles on DNA while the remaining mutant proteins show ability to induce a distribution of DNA bend angles. A Gaussian distribution was centered on 0° used to fit and exclude these from the distribution (Fig 10.5c,d). We found after angle adjustment, Mutant 1 and Mutant 2 induce a mean bend angle of 57.8±2.4° and 54.6±2.5° with a standard deviation of 30.3° and 34.2°, which are very similar to NHP6A. The bend angle distributions support a static kink model with a variation of DNA local flexibility and an additional variance due to protein binding and bending by HMGB proteins.
Additionally, using the protein height distributions from Fig 10.3 we can study bend angle distributions for different oligomer sizes by setting certain height thresholds. For HMGB2, thresholds can be set at 1.7 nm, 2.5 nm, 3.5 nm and 4.6 nm which correspond to minima in fig 10.3a. Fig 10.5a shows protein bend angle distributions in different oligomer height ranges. Similar analysis was applied to NHP6A, Mutant 1 and Mutant 2 and the results are shown in Fig 10.5b-d. The average protein induced DNA bend angles at different oligomer heights can be easily measured (fig 10.5e). We found for NHP6A, Mutant 1 and Mutant 2, the average bend angle increases as the oligomer size increases. However, for HMGB2, the above relationship is reversed. The above results demonstrate that for HMGB 2, the smaller size oligomers and single proteins have more ability to enhance DNA flexibility whereas for NHP6A, Mutant 1 and Mutant 2, the larger size oligomers have more ability to enhance DNA flexibility. The mechanism behind this phenomenon is unknown, but the cationic leader must play a role in changing their bending abilities, although this effect does not vary monotonically with the size of the cationic leader.
(a) HMGB 2 protein induced DNA bend angle distribution

(b) HMG yeast NHP6A protein induced DNA bend angle distribution
HMG mutant 1 protein induced DNA bend angle distribution after adjustment
(d) HMG mutant 2 induced DNA bend angle distribution after adjustment
(d-1) Mutant 2 protein height below 1.8nm

(d-2) Mutant 2 protein height between 1.8nm and 2.3nm

(d-3) Mutant 2 protein height between 2.3nm and 3.2nm

(d-4) Mutant 2 protein height between 2.3nm and 3.2nm

(d-5) Mutant 2 protein height above 3.8nm
Fig 10.5a,b,c,d: Protein induced DNA bend angle distributions for HMGB2, NHP6A, Mutant 1, and Mutant 2 protein respectively. Fig 11.5a-1 to 5, Fig 11.5b-1 to 6, Fig 11.5c-1 to 7 and Fig 11.5d-1 to 5: Protein induced DNA bend angle distributions cataloged by different protein featured heights for HMGB2, NHP6A, Mutant 1, and Mutant 2 protein respectively. Fig 11.5e: Relation between protein featured heights and their correspondent average induced DNA bend angle distributions for HMGB2, NHP6A, Mutant 1, and Mutant 2 protein respectively. The results show that for NHP6A and Mutant 1 and Mutant 2 protein, the bigger size protein has more ability to bend DNA whereas for HMGB2 protein, the smaller size protein has more ability to bend DNA.
Furthermore we are more interested in the variance of the bend angles, because it gives us information about additional induced flexibility at the protein binding sites and the influence of N-terminal leaders. From the bend angle distributions, we confirmed that HMGB protein-induced DNA bend β follows a static with flexibility of DNA and an additional Gaussian variance $\sigma'$ due to protein binding and bending. Thus the total Gaussian variance $\sigma$ is equal to $\sqrt{\sigma^2 + \sigma_{DNA}^2}$. Fig 10.6a shows this model, an important thing needed to be noted is that the static kink $<\beta>$ can be deflected in two directions in plane. Therefore, the actual $\beta$ distribution is bi-Guassian whose peaks are located at $<\beta>$ and $-<\beta>$ (Fig 10.6b). However, in the experimental angle measurement, it is very difficult to distinguish the orientations of the $\beta$ because DNA molecules are randomly distributed on surface and only the absolute values of the angles are recorded. Therefore, in order to fit the measurement angle distributions with the model, we need convert the angle in the model to that of the measurement. In this case, the angles on the left side of the vertical axis at $\beta=0$ were mirrored to the right side (Fig 10.6c). Because the distributions should be symmetric, the final summed distribution remains the same as the right side distribution. Such a projection will cause the experimental measured $<\beta>$ will be slightly larger than the model (off peak) as illustrated in Fig 10.6c.
Fig 10.6a: the proposed protein-bend DNA mechanism the protein-induced DNA bend angle $\beta$ follows a static kink model $<\beta>$ with a flexibility of DNA with a Gaussian variance $\sigma$. The kinks can bend in two directions in plane and the actual $\beta$ distribution is bi-Gaussian whose peaks are located at $<\beta>$ and $-<\beta>$. Fig 10.6b: the probability density function (PDF) function of the $\beta$. Fig 10.6c the probability density function (PDF) after the projection (the summed distribution after the angles on the left side of vertical axis at $\beta=0$ were mirrored to the right side). The experimental measured $<\beta>$ will be slightly larger than the model (off peak).
Then, we fit the measurement by the model, a $\chi^2$ fit was employed. The chi-square test statistic [54] is of the form

$$
\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}.
$$

(10.1)

where $E_i$ is the expected value, $O_i$ is the observed value and the function of $(\sigma, <\beta>)$, and $k$ is the number of the histogram bins.

A random variable $\chi^2$ is said to have a chi-square distribution with $m$ degrees of freedom if it is the sum of the squares of $m$ independent standard normal random variables (the square of a single standard normal random variable has a chi-square distribution with one degree of freedom).

The standardized counts $\frac{(O_i - E_i)}{\sqrt{E_i}}$ for $k$ possibilities are approximately normal, but they are not independent because one of the counts is entirely determined by the sum of the others (since the total of the observed and expected counts must sum to $n$). This results in a loss of one degree of freedom, so it turns out the distribution of the chi-square test statistic based on $k$ counts is approximately the chi-square distribution with $m = k - 1$ degrees of freedom.
The error of the $\chi^2$ fit [55] is defined as

$$ \text{error} = \sqrt{2\left(\frac{\partial \chi^2(\sigma)}{\partial \sigma^2} \right)_{\sigma=m}^{-1}} \quad (10.2) $$

where $\chi^2(\sigma_m) = \chi^2_{\text{min}}$

Then we used this test to find the best fit variance $\sigma_m$ and its errors. Fig 10.7a-h show the results of the fitting for four types of HMGB proteins. And the best fit $\sigma_m$ for four types of HMGB proteins have been summarized in Fig10.7i. A similar analysis can be applied to find the best fit $<\beta>$ as well. The results are summarized in Table 10.1.

The best fit variance $\sigma_m$ will be further employed to determine the $\sigma^*$ for the other $\chi^2$ test to find optimal $<\beta>$ in Chapter 11. The $\sigma_m$ we obtained from $\chi^2$ fitting depends on the length of N-terminal leader and the protein oligomer size. In order to find N-terminal leader induced bend angle flexibility $\sigma_n$, the protein oligomer size effect induced bend angle flexibility must be eliminated from $\sigma_m$. We assume $\Delta\sigma_\beta$ is the bend angle flexibility due to the size effect and it is:

$$ \Delta\sigma_\beta = \left| \left(\frac{d <\beta>}{dh}\right)\sigma_b \right| \quad (10.3) $$

108
where \( \frac{d \langle \beta \rangle}{dh} \) can be obtained from the slope of Fig 10.5e and \( \sigma_h \) is the standard deviation from the protein oligomer height distributions (Fig 10.3).

Then the N-terminal leader induced bend angle flexibility can be approximated as follows:

\[
\sigma_N \approx \sqrt{\sigma^2_m - \Delta \sigma^2_\beta}
\]  

(10.4)

The results are further summarized in Fig 10.7g and Table 10.1. By comparing the four types of HMGB proteins, we can conclude that the longer N terminal leader in the protein the higher the induced flexibility in protein binding and bending.
Length of N-terminal leader (# of amino acids)
Fig 10.7a-d: $\chi^2$ fit for the four types of HMGB proteins by comparing the model in fig 10.6 and the measurements. And the best fit variance $\sigma_m$ and its errors were determined by the test and Eq.10.2. Fig 10.7e-h: the comparison of the histogram between the best fit variances $\sigma_m$ and the measurement for four types of HMGB proteins. Note: Due to the difficult to fit the mutants, only the right shoulder of the distribution in the histograms was fitted. Fig 10.7i: the relation between the length of the N-terminal leader in HMGB protein and the best fit variance $\sigma_m$. Fig 10.7g: the relation between the lengths of the N-terminal induced bend angle flexibility $\sigma_m$. The relation shows the longer N terminal leader in the protein the higher induced flexibility in protein binding and bending.
<table>
<thead>
<tr>
<th></th>
<th>HMGB2</th>
<th>NHP6A</th>
<th>Mutant 1</th>
<th>Mutant 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFM local angle</td>
<td>69.9±2.3°</td>
<td>59.5±2.5°</td>
<td>57.8±2.4°</td>
<td>54.6±2.5°</td>
</tr>
<tr>
<td>measurement &lt;β&gt;</td>
<td>σ=30.3°</td>
<td>σ=36.5°</td>
<td>σ=30.3°</td>
<td>σ=34.2°</td>
</tr>
<tr>
<td>Model fit for AFM local angle measurement</td>
<td>64.5±2.0°</td>
<td>54.5±2.6°</td>
<td>45.5±2.2°</td>
<td>37.5±2.5°</td>
</tr>
<tr>
<td>Best fit</td>
<td>26.0±1.7°</td>
<td>38.0±2.0°</td>
<td>28.0±2.6°</td>
<td>34.0±3.0°</td>
</tr>
<tr>
<td>σₘ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N-terminal leader induced</td>
<td>25.0±1.0°</td>
<td>37.8±1.8°</td>
<td>25.6±1.5°</td>
<td>33.4±2.8°</td>
</tr>
<tr>
<td>σₙ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 10.1: mean protein induced DNA bend angle <β> by experimental local angle measurement and model fitting.
10.5 \(<R^2> method to infer mean protein-induced DNA bend angle\)

We used a second method to estimate the average protein induced bend angle, which avoids possible errors in local bend angle measurements. This method is based on Worm Like Chain (WLC) model [37, 38] and was used by Revetti [31, 46] to model DNA deposited on a 2D mica surface. Revetti further developed a modified WLC model for DNA bound with proteins [46] in 3D. In [5], this modified WLC model was adapted for protein DNA complexes which are deposited on mica surfaces and are able to reach a 2D equilibrium state (eq.6.12). This method requires measuring the global protein-DNA complex end to end distance and counting the number of proteins binding to each individual DNA.

\[<R^2_\beta> \approx 4pL(1-\frac{2p<Np>}{L}(1-<\cos \beta>))\]

\(<R^2_\beta>\) is protein-DNA complexes mean squared end to end distance, \(p\) is bare DNA persistence length, which is 57±6nm for 4361 bp PBR322 DNA [5, 19]. And \(L\) is the contour length of bare DNA, which is 1.48±0.11um [5]. \(<Np>\) is the number of proteins bound and \(\beta\) is the protein induced DNA bend angle.

Using equation (6.12), a \(<\beta>\) can be determined by measuring the \(<R^2_\beta>\) and \(<Np>\). Table 10.3 shows the results for the four types of proteins.
<table>
<thead>
<tr>
<th></th>
<th>HMGB2</th>
<th>NHP6A</th>
<th>Mutant 1</th>
<th>Mutant 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$&lt;R_{\beta}^2&gt;$ ($\mu$m$^2$)</td>
<td>0.283±0.031</td>
<td>0.201±0.027</td>
<td>0.291±0.040</td>
<td>0.316±0.026</td>
</tr>
<tr>
<td>$&lt;N_p&gt;$</td>
<td>2.9±0.3</td>
<td>5.0±0.4</td>
<td>4.0±0.2</td>
<td>2.4±0.1</td>
</tr>
<tr>
<td>$&lt;\beta&gt;$ (°)</td>
<td>74.3±5.5</td>
<td>92.8±1.7</td>
<td>55.7±21.8</td>
<td>51.4±25.2</td>
</tr>
</tbody>
</table>

Table 10.3: Protein-DNA complexes mean squared end to end distance, number of protein or protein oligomer bound per DNA and the correspondent average induced DNA bend angle deduced by using equation (6.12).
For HMGB2, the mean angle obtained by the two methods agrees well and for Mutant 1 and Mutant 2, the mean angle obtained by two methods agrees within errors. However for NHP6A the protein induced bend angle determined by direct measurement is 59.5±2.5°, with a standard deviation of 36.5° whereas the global method gives a 92.8±1.7°, resulting in poor agreement between two methods. In addition to individual protein and protein oligomer binding to DNA, there also exists a cooperative binding structure which produces a rigid DNA segment. For NHP6A, because protein binding can only be seen at higher protein concentration(x=1:9) compared to the other three types, cooperative binding would be enhanced. In fact, one or more small rigid fragments can be seen on some NHP6A images. Therefore, these small rigid fragments must be taken into account in the model (fig 6.10).

To account for this effect, an extra term $4pl_f (\cos \beta_{f_1} + \cos \beta_{f_2} - 1) + l_f^2$ (eq. 6.18) must be added to the right side of equation 2, where $p$ is the persistence length of bare DNA; $l_f$ is the length of rigid fragment; $\beta_{f_1}$ and $\beta_{f_2}$ are the two bend angles the rigid fragment with its two connecting DNA segments. Therefore, the NHP6A protein induced bend angle deduced by mean squared end to end distance should be revised. We estimate that ~1/4 of the NHP6A are bound to DNA via cooperative binding. The average rigid fragment length was estimated to be 0.110±0.005um and the average $\beta_{f_1}$ and $\beta_{f_2}$ ~73.8±10.6°, all of these lead to a calculated $<\beta>$ of 81.8±5.2°, which is closer to the directly measured mean bend angle 59.5±2.5° with a standard deviation of 36.5°.Still,
there remains a discrepancy which seems to require a more sophisticated model to understand.

10.6 High protein concentration- rigid fragment forms

As the protein concentration increases, the number of proteins bound to DNA increases, as well as the tendency to cooperatively bind and form rigid filaments. By comparing the AFM images protein for low and high x, we saw a gradual transition in the number and size of proteins cooperatively bound to DNA. Fig 10.8a shows a proposed mechanism for how proteins cooperatively bind and how rigid fragments form [50]. Fig 10.8b shows the 2 μm x 2 μm AFM image of NHP6A/DNA at higher protein concentration(x=1:3). In the image, rigid filament structures can clearly be seen. These saturated regions are often seen at higher protein concentration. And for HMGB2, Mutant 1 and Mutant 2, similar phenomenon can be seen at concentration ratios x=1:10 for HMGB2, x=1:3.7 for Mutant 1 and x=1:5 for Mutant 2 as well.
Fig 10.8a: Schematic map showing how proteins make the cooperative binding and how regional rigid fragments form on DNA. Fig 10.8b: Regional rigid fragments at higher protein concentration on 2umx2um AFM image for NHP6A (x=1:3).
Chapter 11

Analysis of DNA flexibility enhancement by HMGB proteins by computational simulation

11.1 The motivation for simulation

In an AFM experiment, there are two methods to obtain protein induced DNA bend angle. One method is directly visualizing DNA induced bend angle from the AFM image. Because the DNA strand is a semi-flexible structure, even with some advanced tracing software such as Image J [33, 34], the entry and exit strands, which are used to construct an angle, can be very hard to measure. The bend angle measured by this method is more or less biased by human eye or by software. The other method relies on end to end distance \((R)\) measurement and worm like chain model [5]. Compared to the angle measurement, the end to end distance measurement is easier to implement because as long as DNA chains are well distributed and not crossing with each other on AFM image, the end to end distance is very easy to measure by just indentifying two ends of DNA chain. With the end to end distance statistics, the protein induced bend angle can be inferred from modified worm like chain model [5, 46]. Our previous work [5] derived a closed form formula from this model. However, such a formula is highly approximated and has many limitations [5]. Therefore, in order to precisely quantify the model and
employ the model to find protein induced bend angles, computational simulation is necessary.

In this work, we applied a Monte Carlo method to simulate a large number of DNA chains or DNA-protein complex chains and obtain statistics of these chains. Then, we proposed an \( \chi^2 \) statistical test first disclosed in Dame’s work [56]. The test is based on the compare of end to end distance (\( R \)) histogram between Monte Carlo simulation for DNA worm like chain model in 2D equilibrium and AFM experiment (Fig 11.1). The test provides us a sophisticated way to determine optimal protein induced bend angles and make an optimal fit.

Fig 11.1 2 \( \mu \)m x 2 \( \mu \)m AFM image of HMGB protein bound DNA complexes
11.2 DNA WLC model for simulation

The worm like chain (WLC) model in 2D equilibrium [57] assumes the probability density function for the bend angle ($\theta$) of adjacent segments of DNA chain follows a Gaussian distribution (Fig 11.2a):

$$N(\theta) = Ae^{-p\theta^2/2l}$$

Where $A$ is normalized constant, $p$ is the persistence length of DNA, $l$ the segment length of the DNA chain and the variance $\sigma_{DNA} = \sqrt{\frac{l}{p}}$

From previous AFM experiment, we found HMGB proteins binding almost follows a static kink model [2, 5]. When there is protein bound to a location on DNA chain, we assume a fixed angle of $\beta$ will be added to the DNA curvature (Fig 11.2b). In addition, an additional variance $\sigma'$ may be incurred [57]. We derived a closed form formula for the modified worm like chain model in our previous publication [5]. The original mean squared end to end distance equation is the sum of a covariance matrix (see chapter 6).

$$< R^2_\beta >= (\int_{\tilde{k}_1}^{L} + \int_{\tilde{k}_1}^{L} \ldots + \int_{\tilde{k}_N}^{L})(\int_{0}^{\tilde{k}_1} + \int_{0}^{\tilde{k}_1} \ldots + \int_{0}^{\tilde{k}_N})< \vec{u}(s) \cdot \vec{u}(s') > ds' ds \equiv$$
where $\tilde{k}_1$ is the position where protein binds to DNA chain, $L$ is the length of protein-DNA complex, $R_\beta$ is the end to end distance of protein-DNA complex, $\tilde{u}(\tilde{s})$ and $\tilde{u}(\tilde{s}')$ are the tangent vectors at position $\tilde{s}$ and $\tilde{s}'$ on DNA chain. 

However, the equation (Eq.6.12)

$$< R^2_\beta > \approx 4 p L (1 - \frac{2p < Np >}{L} (1 - < \cos \beta >))$$

we derived in [5] is just a approximation and can lead to obvious deviation when the amount of proteins bind to DNA increases because we only keep the diagonal elements and the dominant term of the sub-diagonal elements. Therefore, a more quantitative and precise computational simulation method is needed.
Fig 11.2a Worm like chain model for DNA; Fig 11.2b Worm Like Chain model for DNA with protein bound
11.3 Box-Muller algorithm-Gaussian random number generator

The bend angle ($\theta$) of adjacent segments of a DNA chain follows a Gaussian distribution. We can simulate the bend angles by using a Gaussian random number $N(0, \sigma_{DNA})$

The Box-Muller algorithm [58] is a method of generating pairs of independent standard Gaussian distributed (zero expectation, unit variance) random numbers, given a source of uniformly distributed random numbers.

Suppose $U_1$ and $U_2$ are independent random variables that are uniformly distributed in the interval $(0, 1)$. Let

$$Z_0 = \sqrt{-2 \ln U_1 \cos(2\pi U_2)}$$

And

$$Z_1 = \sqrt{-2 \ln U_2 \sin(2\pi U_1)}$$

Then $Z_0$ and $Z_1$ are independent random variables with a Gaussian distribution of standard deviation of 1 ($N(0,1)$).
11.4 Knuth algorithm-Poisson random number generator

In our research, we assume the number of proteins binding to per DNA \((Np)\) follows a Poisson distribution.

The Poisson distribution probability density function (PDF) is:

\[
P(k) = \frac{\lambda^k e^{-\lambda}}{k!}
\]  

(11.2)

where \(k\) is the number of occurrences of an event and \(\lambda\) is a positive real number, equal to the expected number of occurrences that occur during the given interval.

The idea behind the Knuth algorithm [59] is that the time between arrivals in a Poisson process is exponentially distributed. Count how many arrivals there are in an interval by simulating the times between arrivals and adding them up until the time sum spills over the interval. We can draw \(u_i (i=1,2,3\ldots)\) from uniform distribution \(U(0,1)\), until \(u_1 u_2 u_3 \ldots u_k u_{k+1} < e^{-\lambda}\), Then \(k\) is the generalized Poisson random number.

11.5 Monte Carlo method to generate DNA chains

Monte Carlo method [60] is a class of computer algorithms which rely on repeated random sampling to compute the results. In our simulation, we are concerned about the convergence of the simulation data. We can assume we repeat sampled a series of random variable \(X_i (i=1,2,3\ldots N)\), which can represent a series of sampling DNA
contours for the Monte Carlo simulation. Each random variable is assumed to have the
same variance and expectation value.

\[ \text{Var}(X_i) = \sigma^2 \]
\[ E(X_i) = \bar{X} = (X_1 + X_2 + X_3\ldots + X_N) / N \]

\[ \text{Var}(\bar{X}) = (\text{Var}(X_1) + \text{Var}(X_2) + \text{Var}(X_3) + \ldots + \text{Var}(X_N)) / N^2 \]
\[ = N\sigma^2 / N^2 = \sigma^2 / N = (\sigma / \sqrt{N})^2 \]  (11.3)

Therefore, the convergence rate is \( \frac{1}{\sqrt{N}} \).

11.6 Analysis and simulation results for bare DNA chain

By putting all of the above together, we can apply the Monte Carlo method to simulate real DNA chains. First, we need to select an appropriate length of segment to construct DNA chains. Fig 11.3a shows the simulation results for averaging 1,000 DNA chain in different segment lengths. From the graph, we can see that the segment lengths from 5 nm to 20 nm are good choices because these points shows consensus. Therefore, we chose a fixed segment length of 10 nm in our simulation.

Next, we generated various numbers of randomly generated DNA contours from 100 to 500,000. Fig 11.3b shows the \(<R^2>\) of DNA chains that we constructed, which have the same contour length as PBR322 DNA converges to a constant value of
3.22x10^5 \text{ nm}^2, which is very close to our AFM experiment value of 3.2±0.3x10^5 \text{ nm}^2 [5].

Fig 11.3c further confirmed the variance of $<R^2>$ Monte Carlo simulation follows a $\frac{1}{\sqrt{N}}$ relationship with the number of paths or averages.
Fig 11.3a) The dependence of $<R^2>$ on segment length in Monte Carlo simulation for PBR322 4361bp DNA. Each point is based on 1,000 simulated protein bound DNA molecules; b) $<R^2>$ vs. the # of simulated DNA contours in a Monte Carlo simulation for PBR322 4361bp DNA; c) Convergence rate for error of $<R^2>$ Monte Carlo simulation for PBR322 4361bp DNA
11.7 Analysis and simulation results for protein/DNA complexes

Then proteins were added to bare DNA chain in simulation based on the modified WLC model. Fig 11.4 shows the relationship between $<R^2>$, mean protein induced DNA bend angle $<\beta>$ and mean # of protein bound per DNA chain $<Np>$ in 1,000 simulated protein bound-DNA molecules for each $<R^2>$. It is very clear that when $<\beta>$ or $<Np>$ increases, $<R^2>$ decreases. From the modified WLC model, the approximate equation(eq.3) only is valid for the case of $Np \leq 8\[5\]. The simulation can provide us a more accurate and precise estimate for large $<Np>$.
Fig 11.4 simulated protein bound PBR322 4361bp DNA $<R^2>$ with $<Np>$ from 0 to 100 and mean protein induced angle $<\beta>$ from 0° to 170°. Each $<R^2>$ is based on 1,000 simulated DNA molecules.
The modified WLC model equation only can provide us a mean squared end to end distance \(<R^2>\). It lacks a description of detailed end to end distance \(R\) distribution. The simulation has an edge in finding this distribution. Fig 11.5a shows a relationship between end to end distance \(R\) distribution and \(<Np>\). The mean protein induced angle \(<\beta> = 90^\circ\) was assumed. We can see that when \(<Np>\) increases, the \(R\) distribution shifts to left (low value) and higher peak.

Similar result can be observed as \(<\beta>\) increases (Fig 11.5b). In this figure, we assumed \(<Np> = 5\). The result confirms and explains the behaviour of the mean squared end to end distance \(<R^2>\) (Fig 11.4) in a very good detail.
Fig 11.5a) Simulated protein bound PBR322 4361bp DNA R Histogram with $\langle N_p \rangle$ from 0 to 160. Each Histogram is based on 10,000 simulated protein bound DNA.
molecules and mean protein induced angle $\beta$ of $90^\circ$ is assumed; b) Simulated protein bound PBR322 4361bp DNA $<R>$ Histogram with $<Np>$ from 0 to 160. Each histogram is based on 10,000 simulated molecules and $<Np>=5$ is assumed.

11.8 $\chi^2_{(k-1)}$ statistical analysis to determine the optimal protein-induced DNA bend angle

Dame first published this test method to determine the optimal $\beta$ [56]. In his work, he did a control experiment and simulation limited to one protein bound at a specific location on DNA with a fixed angle $\beta$. In this simulation, we hypothesized that the experimental histogram bins are the expected values and simulation histogram bins as observed bins. By changing the $\beta$ and $\sigma'$, we will find the best fit by the $\chi^2$ test. From the simulation, we have 15 bins in histogram, therefore the degree of freedom for chi-squared test is 14. We will accept the hypothesis at $\chi^2$ distribution cumulative probability P values bigger than 0.1, which is statistically insignificant between the experimental $R$ histogram and the fitted simulation ones. Such $P$ values correspond to $\chi^2$ less than 21.1 [58].

The following graphs Fig 6 show $\chi^2$ from $0^\circ < \beta < 170^\circ$ and $0^\circ < \sigma' < 30^\circ$. Every slice in the direction of $\beta$ has a minimum value. From the AFM experimental result, a total variance was obtained for model fitting of the four types of protein-DNA complexes from the experimental $\beta$ measurement from chapter 10,
Therefore the variance of $\sigma' = \sqrt{\sigma_m^2 - \sigma_{DNA}^2}$. Thus, we can calculate the additional variances caused by protein bending for four types of protein DNA complexes (Table 11.1).

<table>
<thead>
<tr>
<th></th>
<th>HMGB2</th>
<th>Mutant 1</th>
<th>Mutant 2</th>
<th>NHP6A</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma'$</td>
<td>9°</td>
<td>13°</td>
<td>23°</td>
<td>29°</td>
</tr>
</tbody>
</table>

Table 11.1 Additional variance $\sigma'$ caused by protein bending by applying

$$\sigma' = \sqrt{\sigma_m^2 - \sigma_{DNA}^2}$$

where $\sigma_m$ is total variance of protein induced DNA bend angle by model fitting from AFM local bend angle distributions. $\sigma_{DNA}$ is the variance of DNA local bend angle.
(a) HMGB2
(b) NHP6A

c) Mutant 1
d) Mutant 2
Fig 11.6a,b,c,d) $\chi^2(k-1)$ fit for four HMGB group protein bound DNA chains with mean protein induced bend angle $<\beta>$ from 0° to 170° and additional variance $\sigma'$ from 0° to 30°. The $<N_p>$ is set to the values obtained from the AFM. Fig11.6e, f, g, h) $\chi^2(k-1)$ fit slice from a,b,c,d respectively by the experimental found $\sigma'$ (Table 11.1). Optimal angles were obtained by this method and recorded in Table 12.2. Each $\chi^2$ point represents the comparing result of 1,000 simulated molecules with experiment.
Then a slice for corresponding to $\sigma'$ from AFM experiment was obtained for four types of protein-DNA complexes (Fig 11.6). Optimal angles were determined from these graphs by choosing angles with $\chi^2$ below 21.1 (Table 12.2). The results agree very well with experimental values.

After the optimal angles for each type of protein-DNA complex was found. An optimal angle was selected to fit the experimental histogram. Fig 11.7a shows a comparison result of the experiment with simulation of end to end distance $R$ histogram of bare DNA. Fig 11.7b-e shows the fits of four types of protein-DNA complexes.

(a) Bare DNA
Fig 11.7a) Histogram of experimental obtained values of $R$ of PBR322 4361bp DNA and the corresponding fits; Fig b-e) Histogram of experimental obtained values of $R$ of HMGB protein bound PBR322 4361bp DNA and the corresponding fits by selecting a point of $\chi^2 < 21.1$. The $< Np >$ is set to the values obtained from the AFM experiment (see chapter 10)
Chapter 12

Conclusion and discussion

12.1 Enhancement of DNA flexibility by HMGB proteins-Part 1

Bare DNA was imaged with AFM and its flexibility was characterized through analysis of the persistence length. It was evaluated with two methods, one based on end-to-end distances and the other by locally measured bend-angles. Both methods assumed the validity of the two-dimensional WLC model. The agreement between persistence lengths found by the two methods and with bulk solution-based bulk measurements validated the method and its assumptions. This gives us confidence that the molecules we prepared were well equilibrated on the substrate and that substrate effects on local bend angles was weak.

For the HMGB proteins bound to DNA, the induced bending angle was measured at the protein binding site directly from the scanned image. The mean bend angles obtained by this method were $78^\circ \pm 1.3^\circ$ for HMGB(Box A) and $67^\circ \pm 1.3^\circ$ for HMGB(Box A+B). We also used a variation of the WLC model to infer the mean protein induced bend angle from measurement of a global property: end-to-end distances. This approach gave $82 \pm 8^\circ$ for HMGB(Box A) and $74 \pm 16^\circ$ for HMGB(Box A+B), which was consistent with the locally measured mean angles. The distribution of the protein
induced locally measured bend angle was found to be moderate, with standard deviations of 23° and 22° respectively. These distributions are not consistent with a purely flexible-hinge model, in which there is no preferred bending angle, nor with a static kink model, in which each protein binds at the same angle.

It is interesting to compare these results with those found in single-molecule DNA stretching experiments [6]. In those experiments, force-distance curves were measured vs. protein concentration. Persistence lengths were derived from fits to a three-dimensional WLC model. The protein-induced bend angle was then inferred from the persistence length in fully saturated limit. The results for bend angle were 99° ± 9° for HMGB(Box A) and 77±7° for HMGB(Box A+B), are quite close (given the assumptions made) but slightly larger than the values found here. Significantly, in both measurements the single box protein produces a larger bend angle. One difference may arise from saturation effects on the local bend angle. We cannot easily measure local bend angles at high protein concentration because of depletion and aggregation effects.

<table>
<thead>
<tr>
<th>Method</th>
<th>HMGB (Box A)</th>
<th>HMGB (Box A+B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>X-ray crystallography</td>
<td>~111°</td>
<td>101.5± 9.1°</td>
</tr>
<tr>
<td>Optical tweezers</td>
<td>99±9°</td>
<td>77±7°</td>
</tr>
<tr>
<td>AFM &lt;R²&gt;</td>
<td>82±8°</td>
<td>74±16°</td>
</tr>
<tr>
<td>AFM local angle</td>
<td>78±1.3° (σ=23°)</td>
<td>67±1.3° (σ=21°)</td>
</tr>
</tbody>
</table>

Table 12.1 Average DNA bend angles induced by HMGB proteins based on different experimental methods
It is somewhat surprising that our current results do not agree with the flexible hinge model, in contrast to that observed recently for the *E. coli* HU protein, another nonspecific DNA bending protein that is believed to play a role in *E. coli* analogous to that of the HMGB proteins in eukaryotic systems. Although, in the HU case it was imaged by AFM under kinetic trapping conditions [23], which may have resulted in an altered distribution of bend angles. Bustamante et. al. have argued [19] that structural information obtained from AFM will only be reliable for equilibrated samples. Thus, while the mechanism for achieving protein-induced DNA flexibility may differ for the two proteins, they both achieve the same result, which is a significant reduction in overall flexibility. The result is also surprising because the single molecule studies, in which a significant increase in flexibility is induced on a single DNA molecule, also showed that HMGB proteins were unable to dissociate from DNA under certain solution conditions [26]. Therefore, HMGB proteins must induce DNA flexibility by a combination of the static kink model, in which proteins bind to DNA transiently and induce bends at different locations on the molecule, as well as by introducing local flexibility upon binding. Our results also show that an individual two box HMGB (box A+B) is not more effective at increasing DNA flexibility. This is consistent with the suggestion made by McCauley et al. [27] in which double box proteins in solution are more effective at increasing overall DNA flexibility due to its higher equilibrium DNA binding affinity relative to the single box protein, HMGB (boxA), and therefore greater number of proteins bound under a given set of conditions, not due to its induced DNA bending angle.
12.2 Enhancement of DNA flexibility by HMGB proteins-Part 2

In this work, four types of HMGB proteins and their abilities to bind to and bend DNA were studied. The experiment was performed on DNA equilibrated in 2D on a substrate.

Protein induced DNA bend angles were directly measured from AFM images. From the bend angle distributions, HMGB2 shows better bending ability than NHP6A. The two mutants of NHP6A are very similar, they induced weaker bend angles. For the two mutant proteins, some of them induced near 0° bend angle because of the rigid fragments formed by cooperative binding, while the remaining induce similar bend angle distributions as NHP6A protein. Optical tweezers experiment sensed weaker protein binding and bending for the two mutants as well. All bend angle distributions support a static kink model with a local flexibility of DNA and an additional variance due to protein binding and bending. In addition, it is very interesting that we found that for HMGB2, the smaller size protein oligomers has greater ability to bend DNA whereas for the other three, the larger size protein oligomers have stronger bending ability. A cationic leader of these proteins must play a role in changing their bending abilities. Next, the average protein bend angle was estimated by global methods, by measuring the mean squared end to end distance, the number of protein bound per DNA. The average bend angles obtained by different experimental methods were compared (Table 12.2).
Protein aggregation was observed. This result shows HMGB proteins can form various degrees of oligomers. From the height distributions, it is notable that for HMGB 2 the height distribution is weighted heavily toward the low-height end, as compared with the other three. The underlying reason is that cationic leader of the other three proteins may play a role in promoting protein aggregation.

We tried to increase the protein to DNA ratio (x) to very high value and saw the highly saturated protein DNA complexes. Cooperative bound complexes and rigid protein DNA fragments were observed. By comparing the AFM images for low and high x, we saw a gradual transition in the number and size of protein oligomers binding to DNA. This coincides with the results from other single molecule experiments such as optical tweezers experiments [6] and magnetic tweezers experiments [23]. And, a mechanism for this transition was proposed. As x increases, more and more proteins tend to aggregate together to form oligomers, or cooperatively bind to DNA.

In conclusion, we used AFM to study various protein structures and the mechanism of how protein enhances DNA flexibility. In addition, from low to high protein concentration, a mechanism for the cooperative binding transition was proposed. The insight into the structures and mechanisms of the HMGB protein binding and bending DNA process can help us to better understand the role of these protein-DNA complexes in genetic expression and DNA replication processes in the nucleus of eukaryotic cells.
12.3 Analysis of DNA flexibility enhancement by HMGB proteins by computational simulation

The effects of four types of HMGB protein bound to DNA have been evaluated in this work by simulation. We constructed a mathematical model based on modified WLC equilibrated on a two dimension surface [5, 19]. Compared to Dame’s work [56], in which he did a control experiment and simulation limited to one protein bound at a specific location on DNA with a fixed angle, our model is more complex and more closely mimics to the real biological system. The HMGB proteins that we studied bind to DNA nonspecifically. From the previous AFM experimental results [4], we inferred that HMGB proteins bend DNA by a static kink mechanism, and we assumed that the actual protein induced bend angle we observed is the combination angle of static kink plus a variation of DNA local bend angle and an additional bend due to protein binding. Furthermore, we observed multiple proteins bound to DNA in AFM experiments and assumed the \( N_p \) follows a Poisson process.

The simulation provides us a very useful tool to validate our experiment results and another powerful method to determine protein induced DNA bend angle. We employed the \( \chi^2 \) test in determining the optimal protein induced DNA bend angles. The \( \chi^2 \) test relied on the experimental data for the histogram of \( R \) and the validation of the assumptions of the simulation model. By comparing the protein induced bend angle
from both AFM and optical tweezers experiments [5, 6], the optimal angles obtained by the test agree with these obtained by experiments very well.

<table>
<thead>
<tr>
<th></th>
<th>HMGB2</th>
<th>NHP6A</th>
<th>Mutant 1</th>
<th>Mutant 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFM local angle</td>
<td>69.9±2.3°</td>
<td>59.5±2.5°</td>
<td>57.8±2.4°</td>
<td>54.6±2.5°</td>
</tr>
<tr>
<td>measurement</td>
<td>σ=30.3°</td>
<td>σ=36.5°</td>
<td>σ=30.3°</td>
<td>σ=34.2°</td>
</tr>
<tr>
<td>Model fit for AFM local angle measurement</td>
<td>64.5±2.0°</td>
<td>54.5±2.6°</td>
<td>45.5±2.2°</td>
<td>37.5±2.5°</td>
</tr>
<tr>
<td>Best fit σₘ</td>
<td>26.0±1.7°</td>
<td>38.0±2.0°</td>
<td>28.0±2.6°</td>
<td>34.0±3.0°</td>
</tr>
<tr>
<td>N-terminal leader induced σₙ</td>
<td>25.0±1.0°</td>
<td>37.8±1.8°</td>
<td>25.6±1.5°</td>
<td>33.4±2.8°</td>
</tr>
<tr>
<td>AFM &lt;R²&gt;</td>
<td>74.3±5.5°</td>
<td>81.8±5.2°</td>
<td>55.7±21.8°</td>
<td>51.4±25.2°</td>
</tr>
<tr>
<td>Optical tweezers</td>
<td>59±6°</td>
<td>54±5°</td>
<td>34±3°</td>
<td>38±3°</td>
</tr>
<tr>
<td>X ray Crystallography</td>
<td>~111°</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>χ² test</td>
<td>61°-125°</td>
<td>80°-108°</td>
<td>18°-59°</td>
<td>45°-72°</td>
</tr>
</tbody>
</table>

Table 12.2: protein-induced DNA bend angle by different simulated and experimental methods
More importantly, the simulation can give us insight into the mechanisms of distribution of angle broadening. The experiment can only measure protein-induced bend angles with several proteins bound. Due to the technique difficulty, it is very difficult to obtain a sample with one protein bound (one binding site) or with many proteins bound, because DNA will become highly saturated, and proteins may cooperatively bind to DNA to form rigid filaments. However, in the simulation, we can generate from one protein to over one hundred proteins to clearly observe the trend in $R$. However, simulations for $< Np > > 100$ may not be accurate because in that case proteins may cooperatively bind and a more sophisticated model and simulation is needed in order to take protein cooperative binding into account. We have already carried out a beginning level model for this case (see chapter 6). A more sophisticated simulation will be left for further work.
References


Appendix A

Non-equilibrium mechanics of active cytoskeletal networks

A.1 Introduction

Cells both actively generate and sensitively react to forces through their mechanical framework, the cytoskeleton, which are a non-equilibrium composite material including polymers and motor proteins. In this project we plan to use AFM to measure the dynamics and mechanical properties of a simple three-component model system consisting of myosin II, actin filaments, and cross-linkers. And Myosin is the motor of the system which can generate force via Hydrolysis of adenosine Triphosphate (ATP) to change the equilibrium of the sample system.

A.2 Theoretical background

The response function $\alpha(\omega)$ is defined as

$$\alpha(\omega) = u(\omega)/F$$

(A.1)

$u(\omega)$ is the displacement of tip oscillation and $F$ tip-sample force by the displacement.

And its connection with elastic modulus $G(\omega)$ is

$$\alpha(\omega) = \frac{1}{6\pi G(\omega)a}$$

(A.2)

where $G(\omega) = G^\prime(\omega) + iG^\prime\prime(\omega)$
According to the structural and hysteretic damping theorem, $G(\omega)$ can be expressed as

$$G(\omega) = G_0 \left( \frac{\omega}{\Phi_0} \right)^{-1} (1 + i\bar{\eta}) \Gamma(2 - x) \cos \frac{\pi}{2} (x - 1) + j\omega u $$

(A.3)

where $G_0, \Phi_0$ are scale factor for stiffness and frequency, $\Gamma$ denotes gamma function., $\bar{\eta}$ is the structural damping coefficient and $u$ is the Newtonian viscous term. $X$ denotes the effective temperature or noise level. In [1], it gives that $x > 1$ means the soft glass states, $x = 1$ means the solid state and $x$ approaching 1 means the element is under glass transition.

Because we use the soft glass model to simulate the living cell, $x > 1$ glass state can be used to characterize the living cell.

In our AFM noise spectra measurement, for the cantilever is in free oscillation, the average squared amplitude of tip vertical displacement fluctuations

$$\langle z^2 \rangle \sim \frac{k_BT}{k}$$

(A.4)

where $k_B$ is the Boltzmann constant and $k$ is the force constant of the cantilever. We will measure the AFM noise spectra, as was done in [2]. The $\alpha(\omega)$ normally is a complex number ,which can be written as $\alpha(\omega) = \alpha'(\omega) + i\alpha'(\omega)$

And in equilibrium state (No myosin in sample system), according to the FD(Fluctuation dissipation theorem)
\[
\alpha''(\omega) = \frac{\omega}{2k_B T} S_{eq}(\omega)
\]  
(A.5)

Where \( k_B \) is Boltzmann constant, \( T \) is the temperature and \( S_{eq}(\omega) \) is the Power spectral density at equilibrium state, which is defined as:

\[
S(\omega) = \int (u(t) * u(0))e^{i\omega t} dt
\]

(A.6)

And in the non-equilibrium state, the above FD theorem equation is not valid any more. So we are interested to research how the response function deviates from the right side of the above equation and its dynamics. Similar research done by F. C. MacKintosh’s group [3] revealed this deviation by AMR (Active Microrheology) and PMR(Passive Microrheology) methods.

A.3 Current results

In order to find how to determine the response function \( \alpha(\omega) \), we used a soft-cantilever(PPP-BSI-10,f=13KHz,k=0.01-0.5N/m) to test the behavior of the AFM on hard surface and on soft gel surface.
We observed the first harmonic resonance peak in AC mode (~2KHz in liquid cell) was suppressed when we change the deflection value to get the tip closer to the sample surface. (Fig A2)
Fig A2 Noise spectrum (first harmonic signal in AC mode)
Then we applied 50mv sine wave into the z input port of the AFM. The Sine wave can drive the scanner head to oscillate in the z direction. We measured the AFM deflection output in response to different applied frequencies. The experiment is conducted in contact mode on soft gel surface. With the feedback off at low frequency f<50Hz, the AFM deflection follows the sine wave input. (Fig A3a)

![Deflection vs. Frequency (Setpoint 1v, Feedback Off, Applied v=50mv, f<50Hz)](image)

**Fig A3a** Tip deflection vs. frequency (Setpoint 1v, feedback off, applied v=50mv, f<50Hz)

With the feedback on at low frequency f<50Hz, the feedback tried to stabilize the AFM cantilever but it can not keep up as the frequency increases. (Fig A3b)

![Deflection vs. Frequency (Setpoint 2v, Feedback On, Applied v=50mv, f<50Hz)](image)

**Fig A3b** Tip deflection vs. frequency (Setpoint 2v, feedback on, applied v=50mv, f<50Hz)
With the feedback on at frequency \( f_{>} = 50 \text{Hz} \), the feedback cannot stabilize the AFM cantilever and cause the AFM deflection to follow the sine wave input. (Fig A3c)

![Deflection vs. f(setpoint 2v applied v=50mv \( f_{>} = 50 \text{Hz} \) feedback on)](image)

Fig A3c Tip deflection vs. frequency (Setpoint 2v, feedback on, applied \( v=50\text{mv}, f_{>} = 50 \text{Hz} \))

In the experiment, we did not observe that there was a significant change of the AFM deflection output signal when feedback was off. Maybe that is because we used the AFM cantilever with a sharp tip (~30 nm), which can puncture inside the soft gel surface in contact mode or because the tip radius is too small to probe the appropriate length scales.
A.4 Future work

Therefore, in the future we propose to use a bead (10-20 μm) attached to the end of the cantilever will approach the soft gel surface [4-5]. (A bead (10-20 μm) will not puncture inside the soft gel surface). And, because the experiment is in liquid, it is better to do the experiment in MAC mode instead of regular AC mode.
References


Appendix B

Local fluctuation dissipation spectroscopy of a polymer (PVAC)

B.1 Introduction

In order to measure the local fluctuation dissipation spectroscopy (~10 nm scale), AFM need scan image for a small area (less than 500 nm x 500 nm) as fast as possible. An alternative way is by carefully manipulating AFM tip above a fixed point of the sample surface, local fluctuation dissipation spectroscopy may be detected. The goal of this would be to resolve the structure difference in different phases of the glass polymer.

B.2 Material and methods

5ml PVAC solution was deposited on 1cm diameter round Mica Surface and then was spin for 30 s. A thin film of PVAC, 1um thick was formed on Mica surface.

A Picoplus AFM was employed in the experiment. This AFM works in N2 gas environment. Budget sensors Tap300Al AFM tip was used in this experiment. AFM works in AC mode in air to scan images. The AFM is working in AM mode (Amplitude Modulation mode) compared to RHK UHV350 SPM, which works in FM mode (Frequency Modulation mode). In addition, in order to measure the local fluctuation dissipation spectroscopy (~10 nm scale), we need the AFM to scan for a small area (less than 500 nm x 500 nm) as fast as possible because speed is important because significant dissipation is expected only when the dynamics are comparable with the tip oscillation
frequency. In addition, we can carefully manipulate the AFM tip above a fixed point of the sample surface to detect local fluctuation dissipation spectroscopy. But the thermal drift of the tip may cause a limit to this method.

**B.2 Local fluctuation dissipation spectroscopy for polymer (PVAC) by AM (Amplitude modulation mode) AFM**

![Image of PVAC scan at 25 c](image-url)

*Fig B1a* PVAC 0.2 μm x 0.2 μm scan at 25 c
Fig B1b PVAC 0.2 μm x 0.2 μm scan at 25 c (2 min later)

Fig B1a and Fig B1b are 0.2 μm x 0.2 μm PVAC images scanned by AFM at room temperature (22.8 c), the time interval between two images is 2 min (120 s). The time for scanning one image is about half minute (32 s). AFM tip drift (possible in sample plane) is less than 8 nm/min. So the two scanned images have a maximum 16 nm drift. Compared to the 0.2 μm scanned size, 16 nm drift is small enough to be able to see changes of the drift. Therefore, a marker is needed to track the tip drift. By comparing with the two images, we found the two images are about the same. This result is consistent with that the PVAC is in the solid state at 25 c.
Fig B2a PVAC 0.4 μm x 0.4 μm scan at 45°

Fig B2b PVAC 0.4 μm x 0.4μm scan at 45° (2 min later)
Fig B2a and B2b are 0.4 μm x 0.4 μm PVAC scanned images at 45 c. The time interval between two images is 2 min as well. And the time for scanning each image is 0.5 min. By comparing the two images, we found they are quite different. The phase transition temperature (Tg) is about 35 c for PVAC. Therefore, at 45 c, PVAC is in liquid state and has fluctuation on its surface. We are very curious about how fast and on what length the fluctuations occur. Due to the limitation of AFM tip (10 nm) we used, the tip may resolve dynamics on length scales less than 10 nm. The fluctuation in this region is the combination effect of the fluctuations of several smaller regions (less than 10 nm), Compare to cooperatively rearrangement regions ~3 nm supersharp tip.

Another method to detect the fluctuation of the PVAC surface is holding the tip above the PVAC surface at a fixed point with feedback on. In this case, the vibration of the tip oscillation will reflect the fluctuation of the PVAC sample surface. But at first, we need to know the phase or amplitude relationship with z (vertical distance). Then, we can know vector displacement the tip is above the sample.
Fig B.3a Amplitude vs. z curve

Fig B.3b Phase vs. z curve

Fig B.3a is the amplitude vs. z relation with free amplitude 28 nm. The AFM tip oscillates at a free amplitude (28 nm) at the beginning, as the tip goes closer to the sample, the tip oscillation is suppressed by the tip-sample interaction.
There exists a transition region (from -0.12 μm to -0.18 μm) where the tip amplitude decreased gradually. After this region, the tip gets contact with the sample. Fig B.3b is the phase vs. z curve after we measured the amplitude vs. z curve. Because of the limitation of the instrument, we are unable to measure the two curves at the same time. For the transition region, there exists a hysteresis because the tip reaches contact to the sample surface before the amplitude vs. z curve was measured.

But the same transition region can be observed in the phase vs. z curve. When the tip reaches contact with sample, the signal becomes very noisy.

We are interested in the transition region in fig 2a and 2b. We can hold the AFM tip fixed at one point (oscillating at certain amplitude) in this region. When the AFM tip oscillates at a certain amplitude, we may find the relationship between the phase, the variance of the phase, and the dissipation of the energy of the tip-sample interaction and at the point where the maximum energy dissipation occurs [1]. Previous researchers such as Tamayo, Garcia and Cleveland [2-6] proposed a relationship between energy dissipation and phase:

\[ E_{\text{dis}} = \left( \sin \phi - \frac{A}{A_0} \right) \frac{A}{A_0} F_E \] (B.1)

Where \( E_{\text{dis}} \) is the dissipation energy of the tip-sample interaction, \( \phi \) is the phase shift, \( A \) is the oscillating amplitude, \( A_0 \) is the free amplitude, and

\[ F_E = \frac{\pi k A_0^2}{Q} \] (B.2)
where $k$ is the force constant and $Q$ is quality factor of the AFM cantilever.

**There are two ways to do this experiment**

One way is by changing the “setpoint= % of the free amplitude” in the AFM software, so in AC mode (tapping mode), how much the tip taps to the surface can be controlled and thus we can know where the tip is in the transition region. The disadvantage of this method is each time when the set point changes, the tip need to disengage with the sample. And when reengaging, the tip will likely interact with another area of the sample because of the hysteresis is the force which is caused by the high voltage used in the scanner head when engaging.

Another way is better, we can change the set point in the AFM software when the tip engages the sample while maintaining the tip in the servo range and stable. By this method, we can change the oscillation amplitude of the tip.
Fig B.4a Variation of phase vs. Amplitude at 25 c

Fig B.4b Variance of phase vs. Amplitude at 45 c
Fig B4a and B4b are the curves showing the variance of phase of the PVAC vs. AFM tip oscillating amplitude at 25 c and 45 c, the free amplitude is 28 nm.

We can see there exist certain oscillation amplitudes where the variance of the phase is largest both at 25 c and 45 c.

Martinez and Garcia showed 2006 [5] that there exists a maximum energy dissipation at certain oscillation amplitude.

We still don’t know how the energy dissipation correlates to the variance of the phase. But both results are qualitatively similar.

**B.3 FFT result of the time series phase data**

Fig B.5a and B.5b are two measurements for FFT result of the phase data vs. time at different temperatures.

Fig B.5a shows the FFT result for one measurement. The measurement at each temperature was done for 2 times. We can see the noise level at 31.5 c is apparently lower than other temperatures. And, at nearly 10 Hz, there exist a broad peak only at 31.5 c.

For other temperatures, the FFT noise level shifts to higher value. But from the FFT noise level difference seems very small between 38.1 c and 40.5 c.
We observed a broad peak at 10 Hz at 31.5 °C, the peak has a maximum value of 5E-6 V². Whereas for other temperatures, we didn’t find this kind of peak, but the noise levels at 10 Hz are all above 1E-5 V². Therefore, it is highly possible that the 10 Hz broad peak still exists for all other temperatures, but it is submerged in the higher noise.

Fig B5b shows the FFT result of the repeating experiment done in fig B4a. But this time we did the measurement for more temperatures.

For this measurement, the overall noise level is higher than previous measurement. The signal levels at 10 Hz are all over 1E-4 V². So, there has no way to observe the peak which magnitude is only the order of 1E-6 V².

From 32 °C to 33.5 °C, there exist a big upwards shift for the FFT noise level. This result is the same as Fig B4a, which indicates this result is reproducible.

And generally, the FFT noise shifts to higher values as the temperature rise. This result is the same as we found in fig B5a.
Fig B5a FFT result of the time series phase data by AFM
Fig B5b FFT result of the time series phase data by AFM (with more temperatures)

**B.4 Some conclusions**

The current research shows the PVAC fluctuating at temperatures above $T_g$ (35 c). But due to the limitations of the AFM tip (10 nm) and tip thermal drift, we have no way to resolve at a region smaller than 10 nm (need supersharp tip).

By holding the AFM tip at a fixed point above the PVAC sample surface, the response of the AFM tip represents the change of the fluctuation levels for local PVAC sample surface. Therefore, in this way, we can investigate the phase or variance of phase relationship to tip oscillating amplitude and FFT result of the time series phase data can
be used to examine the fluctuation of the sample. Next step we propose to use supersharp tip (~2 nm radius) tip to resolve the fluctuation.

References


Appendix C

DNA relaxation process simulation

C.1 3D DNA Chain Construction

DNA exhibits a 3D conformation in solutions. From the worm like chain model in 3D[1-3], the probability density function will follow a Gaussian distribution in the two polar angles : $\theta$ and $\phi$.

$$P(\theta) = \frac{1}{\sqrt{2\pi \sigma}} e^{-\theta^2/2\sigma^2}$$  \hspace{1cm} (C3.1)

$$P(\phi) = \frac{1}{\sqrt{2\pi \sigma}} e^{-\phi^2/2\sigma^2}$$

The following graph (Fig C3.1) shows how the 3D DNA chain evolves by stepping in segments of length $l$ along the contour. The chain begins from position $\vec{r}_0 = (0,0,0)$ in Cartesian coordinates $\vec{r} = (x,y,z)$ and the first step goes to position $(x_1,y_1,z_1)$. However, the second position $(x_2,y_2,z_2)$ can’t be obtained directly. Because each segment must follow the distribution in C3.1 in spherical coordinates, the change in position a $(\Delta x_2, \Delta y_2, \Delta z_2)$ in $(x',y',z')$ coordinate system. Then a 3D coordinate transformation needs to be applied to convert from $(x',y',z')$ to $(x,y,z)$. 

179
Fig C3.1 3D DNA chain

\[
\begin{pmatrix}
  x' \\
  y' \\
  z'
\end{pmatrix} =
\begin{pmatrix}
  \cos \theta_i \cos \varphi_i & \sin \varphi_i & -\sin \theta_i \cos \varphi_i \\
  -\cos \theta_i \sin \varphi_i & \cos \varphi_i & \sin \theta_i \sin \varphi_i \\
  \sin \theta_i & 0 & \cos \theta_i
\end{pmatrix}
\begin{pmatrix}
  x \\
  y \\
  z
\end{pmatrix}
\]
\[
\begin{bmatrix}
x_1 \\
y_1 \\
z_1
\end{bmatrix}
= \begin{pmatrix}
sin \theta_1 cos \varphi_1 \\
sin \theta_1 sin \varphi_1 \\
cos \theta_1
\end{pmatrix}^T \begin{pmatrix}
x \\
y \\
z
\end{pmatrix}
\]

\[
\begin{bmatrix}
\Delta x_2 \\
\Delta y_2 \\
\Delta z_2
\end{bmatrix}
= \begin{pmatrix}
sin \theta_2 cos \varphi_2 \\
sin \theta_2 sin \varphi_2 \\
cos \theta_2
\end{pmatrix}^T \begin{pmatrix}
x' \\
y' \\
z'
\end{pmatrix}
\]

After we go to step N to the end, we have constructed a 3D DNA chain. Finally we keep all the (x,y) coordinates to obtain a projection of the 3D coordinates in order to simulate kinetically trapped DNA[1, 3].

C.2 Langevin equation and Ornstein Uhlenbeck process

When DNA is deposited to substrate surface, we can assume the DNA is first kinetically trapped to the substrate surface and then relaxes on surface. Fig C3.2 shows the model. When dividing the DNA chain into thousands of small segments with length \( l \), we can abstract each small segment as a point mass connected by springs to its neighbors with Hooke constant \( K \). For point mass \( i \), we can decompose the motion of the mass to two directions, \( x_i \) is the direction normal into the segment, and \( y_i \) is the direction tangent to the segment. Point mass \( i \) is only correlated to its adjacent point masses \( i-1 \) and \( i+1 \). In
addition, when point mass $i$ moves, it undergoes a friction with friction constant $r$. Therefore, we can construct its equations of motion in $x_i$ and $y_i$ directions and they are Langevin equations

\begin{align}
  r \frac{dx^i(t)}{dt} &= -k_x(x^i(t) - u_x^i) + \varepsilon_{x}^i(t) \\
  r \frac{dy^i(t)}{dt} &= -k_y(y^i(t) - u_y^i) + \varepsilon_{y}^i(t)
\end{align}

Fig C.3.2 DNA relaxation model
Where $\xi_x^i(t)$ and $\xi_y^i(t)$ are white noise terms for point mass $I$, 

$r$ is the friction constant of the point mass with the surface, 

$k_x$ and $k_y$ are Hooke constants, which are equal to $k$, 

$u_x^i$ and $u_y^i$ are the equilibrium positions, where the point mass ultimately arrives. 

We further can write the equation as standard differential equation (SDE). It is an Ornstein-Uhlenbeck mean reverting process [4]: 

$$dx^i(t) = -\alpha_x(x^i(t) - u_x^i)dt + \sigma d\omega_{xt}$$  \hspace{1cm} (C3.4) 

Where $\alpha_x = \frac{k_x}{r}$ 

$$\sigma = \sqrt{\frac{2k_BT}{r}}$$ 

$\sigma$ is derived from the Stokes-Einstein equation $D = \sigma^2/2 = k_BT/r$ for the effective diffusion constant.

$\omega_{xt}$ is the Wiener process 

By constructing $f^i(t) = x^i(t)e^{\alpha_x t}$ and apply Itos lemma:

$$df^i(t) = dx^i(t)e^{\alpha_x t} + x^i(t)de^{\alpha_x t}$$ 

$$df^i(t) = e^{\alpha_x t}(\alpha_x u_x^i dt + \sigma d\omega_{xt})$$ 

$$x^i(t) = x^i(0)e^{-\alpha_x t} + e^{-\alpha_x t}\int_0^t \alpha_x u_x^i e^{\alpha_x s}ds + e^{-\alpha_x t}\int_0^t \sigma e^{\alpha_x s}d\omega_{xs}$$ 

$$x^i(t) = x^i(0)e^{-\alpha_x t} + u_x^i(1 - e^{-\alpha_x t}) + e^{-\alpha_x t}\int_0^t \sigma e^{\alpha_x s}d\omega_{xs}$$  \hspace{1cm} (C3.5)
\[ E(x^i(t)) = x^i(0)e^{-\alpha_t} + u_x^i(1-e^{-\alpha_t}) \]  

(C3.6)

\[ \text{var}(x^i(t)) = E(e^{-\alpha_t} \int_0^t \sigma e^{\alpha_x} d\omega) (e^{-\alpha_t} \int_0^t \sigma e^{\alpha_x} d\omega) \]

\[ = E(e^{-2\alpha_t} \int_0^t \sigma^2 e^{\alpha_x(s)} \delta(s) d\omega x_s d\omega) \]

\[ = e^{-2\alpha_t} \int_0^t \sigma^2 e^{2\alpha_x} ds \]

\[ = \frac{\sigma^2}{2\alpha_x} (1-e^{-2\alpha_t}) \]  

(C3.7)

As similar the derivation can be applied to \( y^i(t) \) also.

\( u_x^i \) and \( u_y^i \) are the final equilibrium positions, which can be confined to positions which the equilibrium requires. However, how to determine these positions needs further investigation.

The simulation results can be verified and calibrated by the AFM experiment. A simple way is to observe the end to end distance histogram at different times after DNA is deposited to the substrate surface. Such a simulation can help to make clear the entire DNA deposition and relaxation process [1, 3, 5, 6]. This provides an efficient guide to validate AFM experiment and devise novel experiment.

**C.3 Performance enhancement techniques of Monte Carlo Method**

The traditional Monte Carlo chain only has a convergence rate of \( 1/\sqrt{N} \). The convergence speed will be very slow if there are many dimensions of random variables
such as the problem above, for each point mass i, there are 2 directions and therefore two dimensions. However, if we need utilize n (let say n=100) such points mass to simulate the relaxation process of DNA chain, we need totally use 200 Gaussian random variables to do the simulation.

Therefore, we want the Monte Carlo simulation converge as fast as possible. Variance reduction techniques will become necessary in solving such a problem. One of the variance reduction techniques is called Quasi Monte Carlo method. This method relies on the random number generator to generate low discrepancy number to reduce variance. A traditional Quasi random number generator is linear congruential generator and a more advanced one is called “Sobol” sequence [7, 8].
References


