BIOPHYSICAL STUDIES ON DNA MICROMECHANICS AND BACTERIAL NUCLEOID ORGANIZATION

QU YUANYUAN

(*B.Sc., ZJU*)

A THESIS SUBMITTED

FOR THE DEGREE OF Ph.D. OF SCIENCE

DEPARTMENT OF PHYSICS

NATIONAL UNIVERSITY OF SINGAPORE

2013

DECLARATION

I hereby declare that this thesis is my original work and it has been written by me in its entirety.

I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

_____

Qu Yuanyuan

13 August 2013

ACKNOWLEDGEMENTS

It's a great pleasure to take this opportunity to thank those who had helped and supported me along the way.

Firstly and foremost, I would like to express my sincere gratitude to my supervisor, Dr. Yan Jie, for his invaluable guidance, persistent support and encouragement, consistent trust throughout my entire Ph.D. period. He settles a stress-free and relaxed atmosphere in the lab that makes me four years study full of joy and happiness.

Much thanks also goes to Dr. Fu Hongxia and Dr. Zhang Xinghua, for their providing the corresponding single-molecule experiment data for my theoretical studies. Especially, I am grateful to Dr. Zhang Xinghua, who not only taught me how to do experiment when I just joined the lab, but also provided valuable suggestions for this thesis.

I am also greatly indebted to Mr. Lim Ci Ji, for his help and advice during our collaboration on the Lsr2 project. I am much appreciated that he generously put in great amount of time to commend and criticize on the thesis section 1.3, especially during his own busy period.

It is so lucky for me to be companioned with and supported by all the group members in the lab ---- Li You, Xu Yue, Li Yanan, Lee Sin Yi, Lim Ci Ji, Wong Wei Juan, Yao Mingxi, Yuan Xin, Le Shimin, Chen Hu, Zhang Xinghua, Chen Jin, Zhao Xiaodan, Cong peiwen, Saranya, Ranjit. This warm family is definitely the biggest treasure I have dug out of my four years life in Singapore.

Last but not the least, I would like to thank my parents and my friends, for their understanding, consideration and support during my four years study. Especially for my boyfriend, Mr. Chen Yuchen, I am deeply appreciated for his trust, companion, understanding and patience along the way.

TABLE OF CONTENT

SUMMARY

Deoxyribonucleic acid (DNA), the most fundamental building block of life, is a long linear polymer that stores the genetic codes for all living organisms. DNA is often described as the right-handed anti-parallel double helix structure, so-called B-DNA, by Watson-Crick base-pairing interaction. However, it can change to various structures to perform its multiple cellular functions. For examples, melted DNA bubble forms during transcription, and the left-handed helical Z-DNA exists in *vivo*, playing a role in transcriptional regulations.

Despite that many DNA structures have been identified under various conditions, how many new structures that DNA can form is still not clear. Due to the fundamental importance of DNA, discovering possible new DNA structures has been one of the hot topics in biophysics research. As mechanical force is now believed ubiquitous in cells, there has been an increasing need to understand the micromechanics of DNA and to probe possible new DNA structures that can be induced by force. In this regard, one of my research focuses is to understand DNA structural transitions induced by DNA tension.

Related to my studies on DNA micromechanics, I am also interested in addressing another important question regarding how a long genomic DNA can be packaged into cells by proteins, and how these DNA packaging proteins affect gene transcription. By now, the mechanism of DNA packaging in bacteria is not well understood. The packaged genomic DNA in bacteria is called nucleoid, which is a long circular DNA (up to a few mega bases) organized by a set of abundant DNA binding proteins called nucleoid-associated proteins (NAPs). Besides bacterial genome DNA packaging, these proteins also affect DNA replication and gene transcription globally. In order to gain insights to the mechanisms of bacterial DNA packaging and gene transcription regulation by NAPs, I investigated the interaction between DNA and Lsr2, an important DNA binding protein in the pathogenic bacteria *Mycobacterium tuberculosis* (*MTB*)

that is believed to play a critical role in both *MTB* genomic DNA packaging and controlling the pathogenesis of *MTB*.

Therefore, two topics, including force-induced DNA structural transitions and the interaction between DNA and the *MTB* protein Lsr2, were respectively investigated during my Ph.D. research. These studies involved extensive theoretical and single-molecular experimental approaches, which addressed several outstanding questions in the field.

For DNA micromechanics, I theoretically investigated the stability of different force-induced DNA structures identified in recent experiments, and elucidated the kinetics of their transitions from one structure to another. Further, using a novel full-atom steered molecular dynamics simulation strategy, an elongated double-stranded DNA structure was produced, which is a possible candidate for the mysterious S-DNA structure.

A combination of single-DNA stretching experiment and AFM imaging was employed to study the Lsr2-DNA interaction. I found that Lsr2 cooperatively binds to DNA and forms a rigid Lsr2 nucleoprotein complex at a single DNA level, which restricts DNA accessibility and also mediate tight DNA condensation. These results provide mechanistic insights into the two functions of Lsr2, including gene silencing by DNA access restriction, and genomic DNA packaging by DNA condensation.

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

DNA = deoxyribonucleic acid

dsDNA = double-stranded DNA

ssDNA = single-stranded DNA

bp = base pair

NN = nearest-neighbor

PCR = polymerase chain reaction

FJC = freely jointed chain

WLC = worm-like chain

AFM = atomic force microscopy

NAP = nucleoid-associated protein

E. coli = *Escherichia coli*

MTB = *Mycobacterium tuberculosis*

HNS = histone-like nucleoid structuring protein

HU = heat-unstable nucleoid protein

IHF = integration host factors

FIS = factor for inversion stimulation

Lrp = leucine-responsive regulatory protein

Dps = DNA-binding protein from starved cells

CbpA = curved DNA-binding protein A

StpA = suppressor of td mutant phenotype A

APTES = (3-Aminopropyl) triethoxysilane

PBS = phosphate-buffered saline

BSA = bovine serum albumin

KMC = kinetic Monte Carlo

SMD = steered molecular dynamics

MD = molecular dynamics

AMBER = Assisted Model Building and Energy Refinement

CHARMM = Chemistry at HARvard Macromolecular Mechanics

EMSA = electrophoretic mobility shift assay

CHAPTER 1 Introduction

1.1. Background of the study

Among the diverse biological entities, bacteria are the most abundant and essential organisms on earth. They are present in most habitats even in acidic hot springs and radioactive waste (1), and they are critical participators in nutrient recycling of ecosystems.

Bacteria usually have similar components, including cell membrane, cytoplasm, nucleoid, ribosome, flagellum and so on as illustrated in Figure 1.1.1. The most important structure of a bacterial cell is the nucleoid, an irregularly-shaped region which stores all or most genetic materials called chromosome (2). The bacterial chromosome is a well-organized and highly-compacted structure containing a piece of chromosomal DNA and many nucleoid-associated proteins (NAPs). DNA is one of the most important and essential macromolecules for all living organisms and even some viruses, because it encodes all the genetic information they use to function, respond and evolve. NAPs are helpful and crucial in DNA organization and packaging, which makes it possible to put a millimeter's long chromosomal DNA into the nucleoid, a volume hundreds of times smaller than the DNA unconstrained volume. Moreover, NAPs are critical for gene regulation, a process in which a cell decides which gene is to be expressed and when.

As bacteria always live in a complicated and crowded environment, they have to sense many physical aspects of external environment and internal interactions as to respond appropriately for proper cellular functions. Force is one such factor ubiquitous in cell growth, motion, differentiation and metabolism. For instance, outside the bacterial cell, the interactions between a bacterial cell and extracellular matrices or adjacent cells are usually in the presence of force; inside

the bacterial cell, RNA polymerases will directly exert force up to 30 pN on DNA during the transcription process (3).

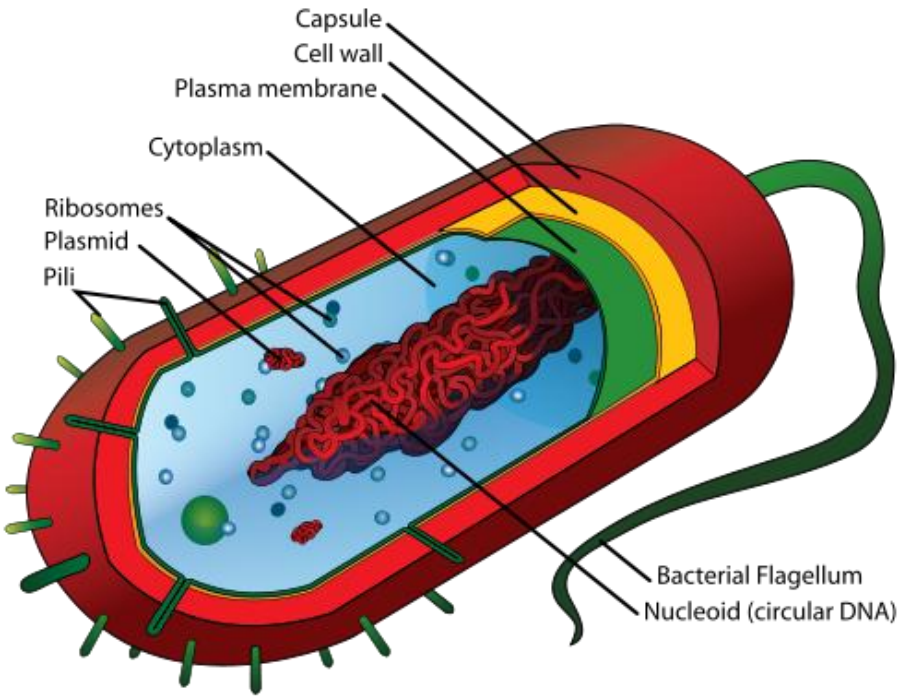


Figure 1.1.1 Cartoon illustration of a typical bacterial structure. This picture is adopted from: http://en.wikipedia.org/wiki/File:Average_prokaryote_cell-_en.svg.

Therefore, the study regarding the force that influences on DNA structure, property and the interaction between DNA and DNA binding proteins is undoubtedly very important and essential, which is also the premise of our correct and complete understanding of proper cell function and development.

1.2. Literature review on DNA micromechanics

1.2.1 DNA structure

DNA is the short name for deoxyribonucleic acid, a double-stranded helix containing two long polymers running in opposite direction to each other and winding with each other. These two long polymers consist of simple units called nucleotides, which include four different types, namely guanine (G), adenine (A), thymine (T) and cytosine (C). The nucleotides are composed of backbones, phosphate groups and the nucleobases (G, A, T, C) attached to backbones. Each type of the nucleobase on one strand can only form stable hydrogen bonds with just one type of the nucleobase on the other strand, which means that guanine (G) can only stably pair with cytosine (C) forming three stable hydrogen bonds, and adenine (A) can only stably pair with thymine (T) forming two stable hydrogen bonds. This pairing rule is called complementary base pairing, which makes the double helix maintain a regular helical structure independent of the nucleotides sequence. This structure of DNA was firstly solved by Francis Crick and James D. Watson in 1953 (4), so the complementary base pair (G-C and A-T) is also called Watson-Crick base pair. The detailed structure of DNA molecule is illustrated in Figure 1.2.1.

Besides the B-DNA structure as shown in Figure 1.2.1, there are several different kinds of DNA structures needed to execute multiple cellular functions, such as single-stranded DNA (ssDNA) involved in transcription and DNA replication (5), left-handed double-stranded Z-DNA involved in relaxing the supercoiling stress (6), and four helices G-quadruplex structures involved in maintaining the chromosomal stability and gene regulation (7) as illustrated in Figure 1.2.2. Besides these known structures, recently a novel double-stranded "S-DNA" was identified in mechanically stretched DNA at forces above ~ 65 pN (8-15).

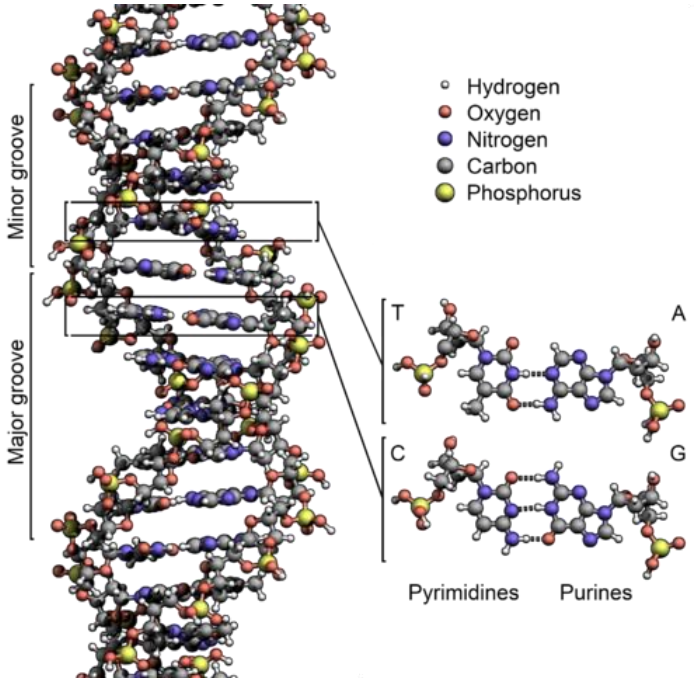


Figure 1.2.1 DNA structure at the atomic-level. This picture is adopted from: http://en.wikipedia.org/wiki/File:DNA_Structure%2BKey%2BLabelled.pn_NoBB.png.

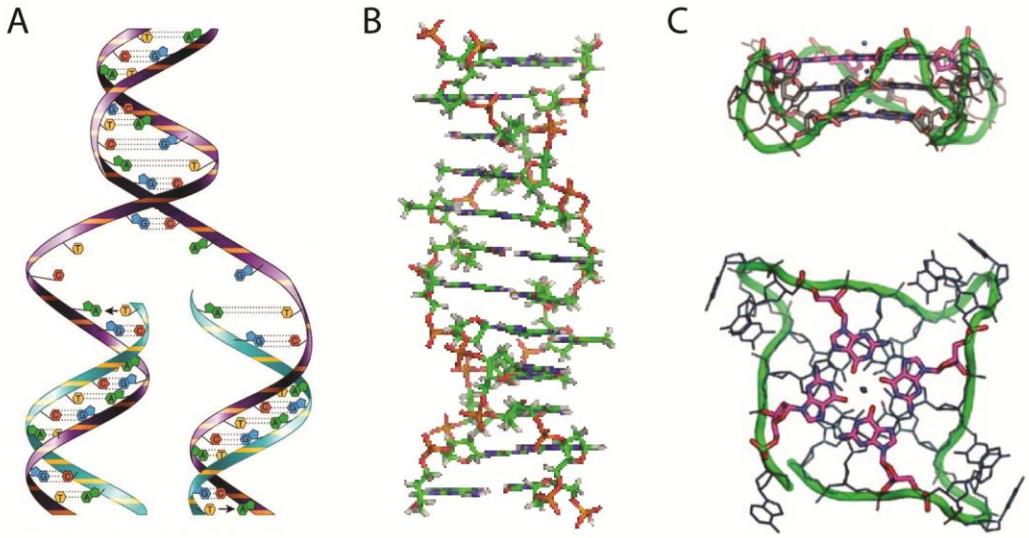


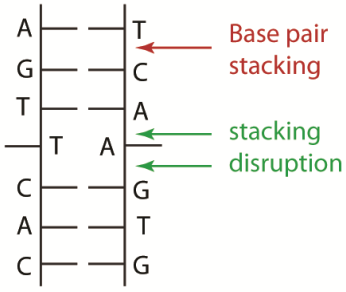
Figure 1.2.2 Alternative DNA conformations. (A) ssDNA involved in DNA replication and transcription. The picture is adopted from: http://classconnection.s3.amazonaws.com/423/flashcards/592423/jpg/dna_replication1315082012045.jpg (B) Z-DNA structure. The picture is adopted from: http://upload.wikimedia.org/wikipedia/commons/f/f5/Z-DNA_orbit_animated_small.gif (C) The structure of G-quadruplex. This picture is adopted from: http://www.intechopen.com/source/html/16928/media/image9.jpeg.

The sequence of the nucleotides contains the genetic information that DNA encodes. DNA is a perfect carrier for genetic information for many reasons. First, as the backbone of the DNA is made of deoxyribose sugar, it is stable and resistant to cleavage. Second, the complementary base pairing ensures the accuracy of DNA replication to pass down the genetic information. Last, the double-stranded helices structure makes it a build-in duplicate of the encoded genetic information.

1.2.2 **DNA base pair stability**

As mentioned in section 1.2.1, the complementary base pairing of the nucleotides is based on the formation of the stable hydrogen bonds, and DNA with high GC-content is more stable than DNA with high AT-content (16). However, on the contrary to intuitive belief, the hydrogen bonds between the nucleotides do not notably stabilize DNA, while the stacking interaction, including dispersion attraction, short-range exchange repulsion and electrostatic interaction (17), between the adjacent base pairs is the main factor stabilizing DNA (16,18). The DNA base pair stability is often described by the DNA melting temperature (T_m), the temperature at which half of the double-stranded DNA (dsDNA) base pairs are unpaired in single-stranded DNA (ssDNA) state. Typically, the T_m of DNA is in the range of 40 – 100 ^{o}C depending on DNA sequence, DNA length, and salt concentration (19).

The DNA stability is mainly determined by nearest neighboring base pair stacking, which is usually referred as the nearest-neighboring (NN) model (20). The four different bases define 16 different sequence-dependent stacking, among which only 10 are non-redundant ones (red, right panel in Fig. 1.2.3). Therefore, in the NN model, the sequence-dependent base pair stacking energies are described for the following adjacent base pair combinations: AA/TT, AT/TA, TA/AT, CA/GT, GT/CA, GA/CT, CG/GC, GC/CG and GG/CC.



	A-T	T-A	G-C	C-G
A-T	A-T	T-A	G-C	C-G
	A-T	A-T	A-T	A-T
T-A	A-T	T-A	G-C	C-G
	T-A	T-A	T-A	T-A
G-C	A-T	T-A	G-C	C-G
	G-C	G-C	G-C	G-C
C-G	A-T	T-A	G-C	C-G
	C-G	C-G	C-G	C-G

Figure 1.2.3 NN model describing the DNA stability. Ten non-redundant base pairs are highlighted in red in the table.

The application of the NN model to DNA can be traced back to 1960s, a few years after the DNA structure was revealed. Tinoco and coworkers are the pioneers in predicting the stacking energy using NN model. Early in 1962, DeVoe and Tinoco theoretically calculated the free energy of the nearest-neighbor base-base interactions of DNA helix both in vacuum and in solution (16). Followed by this, a series theoretical studies based on NN model by Tinoco *et al.* came out regarding the sequence-dependent stability of DNA and RNA molecules (21-24). A few years later, with the determination of the sequence for φX174 DNA by Sanger *et al.* in 1977 (25), Lyubchenko *et al.* made a direct comparison between the theoretical and the experimental DNA melting profiles for the first time in 1978 (26). Subsequently, a few more papers did the comparison but failed to get a good agreement between the theoretical prediction and the experimental observation for different DNAs (27-32). After decades of corrections and improvements, there are several sets of NN parameters for predicting sequence-dependent DNA stability in the literature (19,33-40). Among those studies, different designs of DNAs and salt conditions are chosen, different methods to determine thermodynamics and different ways to present data are employed in different research groups, which brought up great confusion between different sets of NN parameters.

In 1998, SantaLucia *et al.* (20) summarized seven sets of the NN parameters in the literature (19,33,34,36,38-40) and found that they were actually in quite

good agreement with each other. Furthermore, a single set of NN parameters showing the entropy (ΔS) and enthalpy (ΔH) changes of ten NN base pairs during the thermal melting process were provided in this paper as shown in Table 1.2.1. Therefore, the free energy of ten NN base pairs stacking under different temperature could be obtained by following equation: $\Delta G = \Delta H - T\Delta S$. Furthermore, the salt dependence of the free energy was also derived with an empirical equation:

$$\Delta G(I) = \Delta G(I_0) + m \times \ln (I/I_0)$$

where I represents the monovalent ionic (e.g., Na^+) strength, and I_0 = 1 M, denotes the standard ionic strength, $m = 0.114$ kcal/mol is the salt-dependent correction parameter that is the same for all NN base pairs. Since then, this set of NN parameters has been widely used in predicting the DNA folding and hybridization (41,42), estimating DNA melting temperature for specific DNA design (43), probing design for array-based experiment (44) and predicting DNA structural transition under force (45).

Table 1.2.1 Unified oligonucleotide ΔH° and ΔS° NN parameters in 1 M NaCl. Table is adopted from SantaLucia *et al.* (20), where signs are reversed to denote the energy cost.

Sequence	ΔH° kcal/mol	ΔS° cal/(K•mol)
AA/TT	7.9	22.2
AT/TA	7.2	20.4
TA/AT	7.2	21.3
CA/GT	8.5	22.7
GT/CA	8.4	22.4
CT/GA	7.8	21.0
GA/CT	8.2	22.2
CG/GC	10.6	27.2
GC/CG	9.8	24.4
GG/CC	8.0	19.9

However, these NN parameters mentioned above are all measured based on thermal DNA denaturation experiments. In 2010, Huguet *et al.* (46) obtained the NN parameters from single-molecule stretching experiment and further improved the salt dependent parameter m_i for individual base pair as listed in Table 1.2.2. This set of NN parameters provided additional information for studying DNA conformational changes under external force or tension in different solution conditions.

Table 1.2.2 Force measurements of optimal enthalpies and entropies of ten NN base pairs in 1 M NaCl and salt dependence for individual base pair. This table is adopted from Huguet *et al.* (46), where the sighs are reversed to denote energy cost.

Sequence	ΔH_i kcal/mol	ΔS_i cal/(K•mol)	m_i kcal/mol
AA/TT	7.28	20.28	0.145
AT/TA	4.63	11.62	0.117
TA/AT	8.31	25.06	0.091
CA/GT	8.96	24.48	0.091
GT/CA	5.80	14.46	0.099
CT/GA	5.21	12.89	0.070
GA/CT	8.16	22.46	0.155
CG/GC	9.66	24.43	0.132
GC/CG	10.10	25.96	0.079
GG/CC	8.57	22.30	0.063

1.2.3 **DNA conformation under force**

As a well-defined long polymer with unique mechanical properties, the conformation of DNA under force has attracted attentions from many polymer physicists. Further, as DNA is mechanically folded in cells, the mechanical response of DNA also has important physiological implications. Therefore, it is necessary to investigate DNA elasticity and conformational changes under force.

In 1992, the first direct mechanical measurements of the elasticity of single DNA molecule were carried out by Smith *et al.* (47). In this experiment, they measured the DNA force-extension curve for the first time, which could be well fitted by the worm-like chain (WLC) model, describing DNA as an inextensible elastic rod with a finite bending rigidity (48,49). In the WLC model, for a given conformation of a stretched DNA with a contour length of *L*, its conformational free energy is described by:

$$\frac{E}{k_B T} = \int_0^L \frac{A}{2}\left(\frac{d\hat{t}}{ds}\right)^2 ds - \frac{fz}{k_B T}$$

where $\hat{t}(\mathrm{s})$ is the tangent vector, s is the arc length, $\frac{d\hat{t}}{ds}$ is the local curvature of the DNA, z is the end-to-end extension of the DNA and f is the tensile force applied to the DNA. The parameter *A*, which is often referred as the bending persistence length, describes the DNA bending rigidity. From this model, at high force ($f \gg 0.08\ pN$), the force-extension curve is derived to be :

$$\frac{fA}{k_\mathrm{B}T} = \frac{1}{4(1 - z/L)^2}$$

while at low force ($f \ll 0.08\ pN$), it is :

$$\frac{fA}{k_\mathrm{B}T} = \frac{z}{L}$$

A direct interpolation of the two force limits leads to the Marko-Siggia formula:

$$\frac{fA}{k_\mathrm{B}T} = \frac{z}{L} + \frac{1}{4(1 - z/L)^2} - \frac{1}{4}$$

which approximates the force responses of DNA over a wide force range from 0 to 20 pN. Fitting the force-extension curve measured by Smith *et al*. (47) with the Marko-Siggia formula leads the bending persistence length *A* of DNA to be ~ 50 nm (48,49).

The Gibbs free energy per base pair at constant force can be obtained by integration of the extension

$$w_B(f) = -\int_0^f df' x_B(f')$$

where $x_B = h_B\left(1 - \sqrt{\frac{k_BT}{4Af}}\right)$ for $f \gg 0.08\,pN$ or $x_B = \frac{fAh_B}{k_BT}$ for $f \ll 0.08\,pN$ is the DNA extension per base pair and $h_B = 0.34$ nm is the B-DNA contour length per base pair. The negative sign represents that the tensile force in DNA is opposite to the extension direction, which ensures that free energy decreases as force increases.

The above model assumes DNA is an inextensible polymer, which is only valid at forces < 20 pN. In fact, DNA has certain stretchability with a spring constant per base pair of $f_B/h_B \approx 1200/0.34\,pN/nm$, where the constant force f_B is often referred to the stretching elastic constant (45). With this correction, a more general WLC model (extensible WLC model) becomes:

$$x_B = h_B\left(1 - \sqrt{\frac{k_BT}{4Af}} + \frac{f}{f_B}\right)$$

Further raising the stretching force to ~ 65 pN, an abrupt increase of the DNA extension from 1.05 times to 1.7 times of the DNA contour length was reported (8,9). This abrupt transition is usually referred as the overstretching transition, whose nature has been debated for 17 years since it was discovered and was elucidated very recently (8-15,45,50-53). The details of this transition will be thoroughly elaborated in the following section.

Compared to dsDNA, ssDNA, with a longer contour length per nucleotide h_{ss} ~ 0.56 nm, is 50-100 times more flexible than dsDNA with a persistence length A_{ss} smaller than 1 nm. Therefore, it tends to collapse into small random coils at force $< k_BT/A_{ss}$ ~ 4 pN and can be stretched to longer extension at high force than B-DNA.

The force response of ssDNA is not so trivial to characterize. Under moderate ionic strength (e.g., 150 mM Na^+), the ssDNA force-extension curve can be well-characterized by a modified FJC model:

$$x_{ss} = h_{ss}\left[\coth\left(\frac{fb}{k_BT}\right) - \frac{k_BT}{fb}\right]\left(1 + \frac{f}{S}\right)$$

where $h_{ss} = 0.56$ nm is the contour length per nucleotide, $b = 1.5$ nm is a Kuhn length in the model, and $S = 800$ pN is the stretch modulus of ssDNA (8). However, the modified FJC fitting fails at low ionic strength (e.g., 2.5 mM Na^+) as shown in Figure 1.2.4 (8).

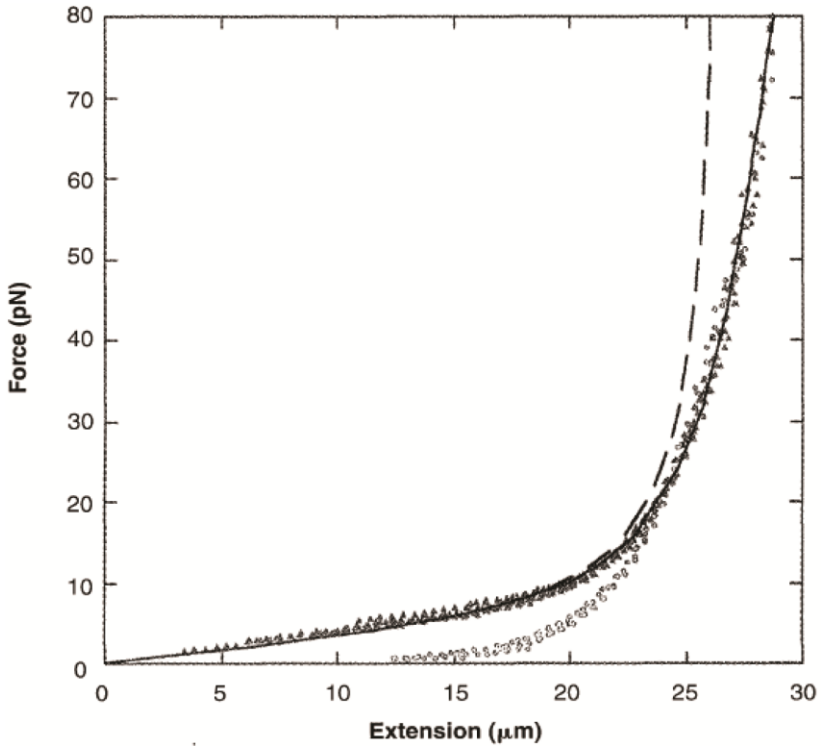


Figure 1.2.4 Force-extension behavior of ssDNA at moderate ionic strength (black dots, 150 mM Na^+) and low ionic strength (grey dots, 2.5 mM Na^+). The dashed line represents the fitting of FJC; while the solid line represents the fitting of FJC with a stretch modulus (modified FJC model). Discrepancy lies between different salt conditions at low forces. This figure is extracted from Smith *et al.* (8).

It has been shown that the elastic response of ssDNA is highly dependent on the ionic strength and sequence composition (54-57). Under high ionic strength,

local association of hydrophobic groups, formation of hydrogen bonds between base pairs and formation of hairpins will result in the collapse of ssDNA under force < 2 pN (54,58,59). While under low ionic strength, the long-range electrostatically repulsion will magnify and influence the ssDNA conformations. A well-accepted unified theoretical model to describe elastic behavior of ssDNA with different sequence composition under different ionic strength is still missing.

In 2004, Cocco *et al.* (45) proposed a phenomenological analytic formula which fits the salt dependent ssDNA elasticity reasonably well over a wide monovalent salt concentration range (2 mM ~ 1 M Na^+):

$$x_{ss} = h\left(\frac{a_1 \ln\left(f/f_1\right)}{1 + a_3 e^{-f/f_2}} - a_2 - \frac{f}{f_3}\right)$$

with the following fitting parameters: $h = 0.34\,nm$, $a_1 = 0.21$, $a_2 = 0.34$, $f_1 = 0.0037\,pN$, $f_2 = 2.9\,pN$, $f_3 = 8000\,pN$ and $a_3 = 2.1\ln(M/0.0025)\,/\,\ln(0.15/0.0025) - 0.1$. The parameter a_3 depends on NaCl concentration M (in Mol/litre) (45). With this expression, the Gibbs free energy per nucleotide at constant force can be obtained:

$$w_{ss}(f) = -\int_0^f df' x_{ss}\left(f'\right)$$

1.2.4 **The debate over DNA overstretching**

In 1996, a DNA overstretching transition was found to occur in a narrow force range slightly above 60 pN, which leads to DNA elongation by ~ 1.7- fold (8,9). The debate over the nature of this transition lasts about 17 years until very recently it was elucidated.

Figure 1.2.5, which is the original figure extracted from the paper by Smith *et al.* (8), describes the observation of the transition and its complicated kinetics. The authors used an optical tweezers to stretch a phage λ-DNA (48502 bp with a

contour length ~ 16 μm) by controlling the distance between the two DNA ends while measuring the force (i.e., end-to-end extension). The force was applied to the two opposite strands of DNA in that experiment (Top panel in Fig. 1.2.5). A sharp rising in force was observed when the extension approximates the contour length of B-DNA (i.e., x_B ~ 0.34 nm), and a structural reorganization was observed when the tension reached ~ 65 pN, indicated by a flat force plateau. The transition ends when the extension became ~ 1.7 times of the contour length of B-DNA, marked by sharp rising in force that represents the force response of the new elongated DNA. The transition is reversible, as when the authors reduced the extension, they observed drop in force until the DNA returned to B-DNA at force slightly below 40 pN.

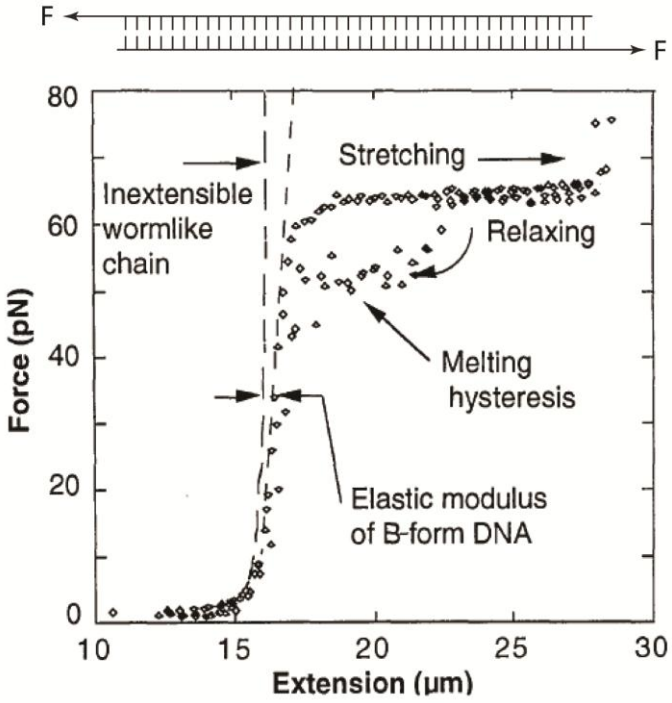


Figure 1.2.5 Stretching of λ-DNA in 150 mM NaCl, 10 mM Tris, 1mM EDTA, pH 8.0. The top panel shows the DNA construct applied in this experiment where force was applied to the two opposite strands. This figure is extracted from Smith *et al.* (8).

There are several facts observed in this transition: 1) the DNA tether was not broken after the transition, making it difficult to explain it by DNA melting into two separated ssDNA strands. Otherwise the tether would break since the force

was applied to two opposite strands. 2) During the force-decrease relaxation, there was a hysteresis in force-extension curve, but only involving around half of the transition plateau. The complicated kinetics led to a 17 years of debate on the nature of the transition. Three possible transitions that may lead to different elongated DNA structures have been proposed, namely, an ssDNA under tension, DNA bubbles consisting of two parallel, separated ssDNA (2ssDNA) under tension, and a hypothesized new form of base-paired double-stranded DNA named S-DNA as shown in Figure 1.2.6. The focus of the debate is whether the mysterious S-DNA exists.

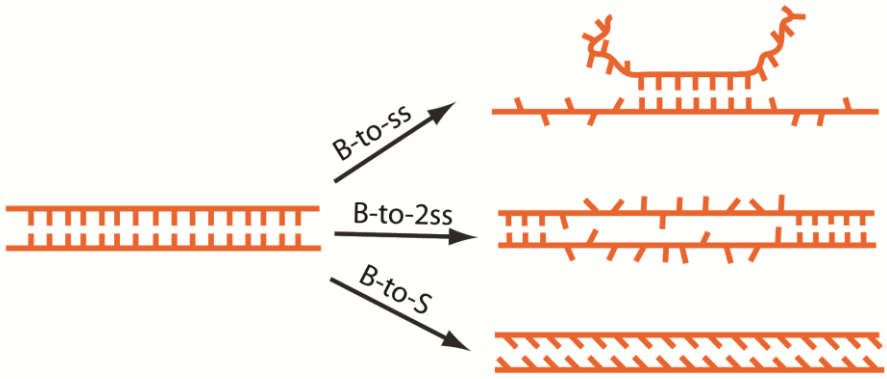


Figure 1.2.6 Three possible transitions to three different elongated DNA structures.

Some groups argued that, due to the one-dimensional nature of the transition, DNA melting alone was sufficient to explain the observations (52,60). By a salt-dependent transition force study (51), the authors concluded that the transition was mainly through an internal melting mechanism, leading to formation of two parallel ssDNA (2ssDNA), where two ssDNA strands are in close proximity. The most direct experimental evidence supporting the DNA melting picture was published in 2009, which directly visualized a partially overstretched DNA using fluorescence-labeled ssDNA-binding proteins and dsDNA binding dye (53). This work clearly showed peeling ssDNA (1ssDNA) mixed with a dsDNA region that can be explained as B-DNA. On the contrary, several other groups argued for the existence of a new form of dsDNA, the S-DNA, during DNA overstretching transition. Assuming its existence, Cocco *et al.* predicted how salt concentration,

force and DNA sequence, may regulate the selection between strand-peeling to 1ssDNA, internal melting to 2ssDNA, and the B-to-S transition to the S-DNA (45). However, direct experimental evidences supporting the existence of S-DNA came only very recently by works from several groups since 2010 (10-15).

From 2010 to 2011, two successive publications by Fu *et al.* (10,11) showed that DNA overstretching involves two transitions that are distinct in kinetics, namely, a slower hysteretic transition involving peeling off an ssDNA from the other (B-to-ss transition) and a faster non-hysteretic transition to an unknown DNA structure. It was also shown that the selection between these two kinetically distinct transitions is highly sensitive to changes in environmental factors and DNA sequence which influence DNA base pair stability. The overall trend is that when a factor change reduces the DNA base pair stability, it will favor the hysteretic melting transition over the non-hysteretic transition, highlighting the fundamentally distinct nature of DNA re-organization between these two transitions (Tab. 1.2.3).

Table 1.2.3 Experimental conditions affect overstretching transition.

	non-hysteretic transition	hysteretic transition
salt concentration	high	low
temperature	low	high
GC content	high	low

After the non-hysteretic transition, the resulting DNA structure has a unique force response, which is distinct from B-DNA, one ssDNA, two parallel ssDNA, or any combinations of these DNA forms (11-13,45). This overstretched DNA force response can be fitted with the extensible WLC model:

$$x_S = h_S\left(1 - \sqrt{\frac{k_B T}{4A_S f}} + \frac{f}{f_S}\right)$$

where x_S is the overstretched DNA extension per base pair, $h_S = 0.56\,nm$, with a persistence length $A_S = 11.6\,nm$ and a stretch modulus $f_S = 3030\,pN$ (11).

However, it remains unclear regarding how the DNA bases are organized in the unknown DNA structure.

Some insights to the unknown DNA structure can be inferred from some recent results reported by Zhang *et al.* (12) in 2012, in which entropy and enthalpy changes per base pair during both the hysteretic and the non-hysteretic DNA overstretching transitions were measured. It was achieved by measuring the temperature dependences of the respective transition forces $F_{\mathrm{ov}}(T)$ (Fig. 1.2.7). It has been shown that the entropy change ΔS per base pair is related to the slope $\partial F_{\mathrm{ov}}/\partial T$ by the equation: $\Delta S = -(\partial F_{\mathrm{ov}}/\partial T)\Delta x$, where Δx is the DNA extension change per base pair during the transition (60). They confirmed that the hysteric transition (B-to-ss transition) is associated with large positive entropy change of ~ 20 cal/(K•mol) due to the gained freedom from the dissociated DNA bases, which is consistent with the DNA thermal melting experiments (20). In contrast, the non-hysteretic transition was found to be associated with a small negative entropy change of ~ -3 cal/(K•mol), which strongly suggests that DNA re-arranges into a highly ordered, non-melted state during the non-hysteretic transition.

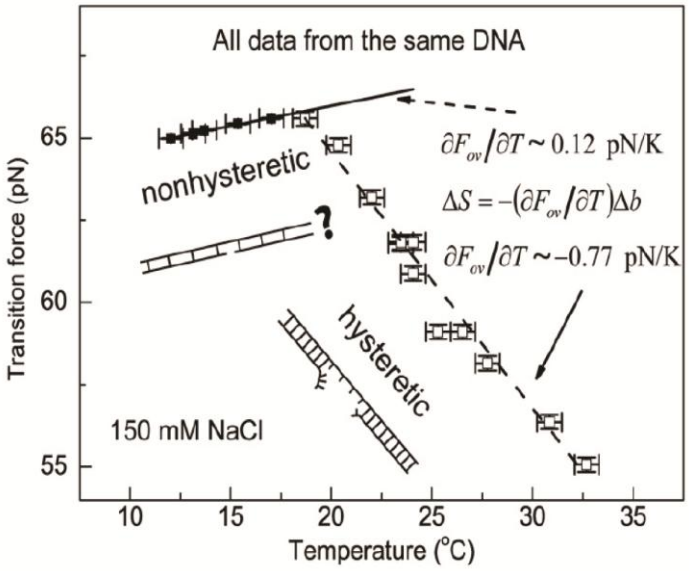


Figure 1.2.7 Temperature-dependent transition force measurement indicates two different transitions:1) B-to-ss hysteric transition associated with large positive entropy change; 2) non-hysteretic transition associated with negative entropy change. This figure is extracted from Zhang *et al.* (12).

In addition, the salt dependence of the transition forces $F_{ov}(I/I_0)$ was also measured, I, I_0 represents the monovalent ionic strength as defined in section 1.2.2, and $I_0 = 1$ M. Theoretical study indicated a linear relation exists between $F_{ov}(I/I_0)$ and $\ln((I/I_0)$ for $I \ll I_0$ by the equation: $\partial F_{ov}/\partial \ln((I/I_0) = v(k_BT/l_B)$, where $l_B \sim 0.71$ nm is the Bjerrum length in water at room temperature, v is the structural coefficient, which is predicted to be ~ 1.2 for transition with one strand peeling off another and ~ 0.5 for transition with two strands tightly associated within the Debye screening length (60) (more details will be shown in section 3.2.1). From their results (Fig. 1.2.8), two distinct values of structural coefficient v were obtained for two transitions, consistent with the former suggestion that the hysteretic transition leads to strand-peeling to 1ssDNA, while in the DNA from the non-hysteretic transition the two strands are in close proximity.

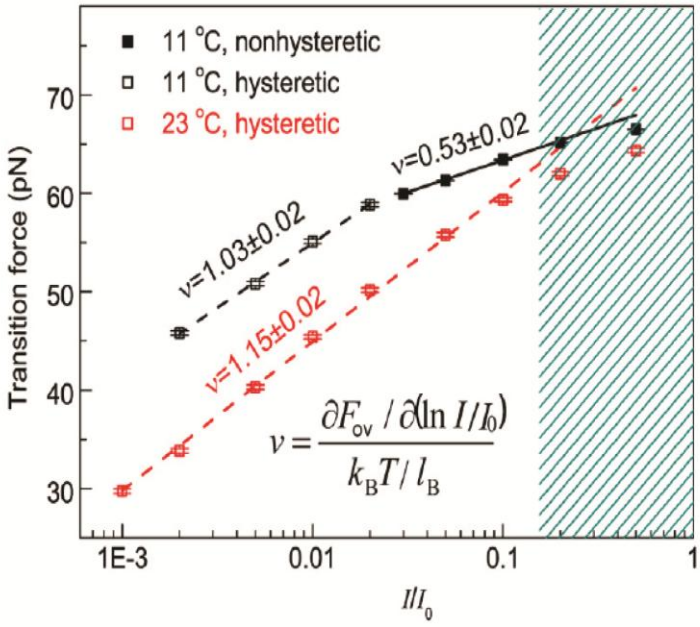


Figure 1.2.8 Salt-dependent transition force measurement indicates two different transitions. This figure is extracted from Zhang *et al.*(12).

Furthermore, B-to-2ss transition was later identified using a topologically closed DNA showing hysteresis during relaxation (13), therefore, the non-hysteretic was finally determined to be the B-to-S transition. The entropy and

enthalpy changes during B-to-2ss transition were found to have similar values as B-to-ss transition due to their DNA melting nature. With this new work, all three proposed structures during overstretching transition have been identified and their respective thermo-mechanical properties have been fully characterized. This completes the picture about the structure of DNA under tension and proves the existence of the S-DNA, providing a conclusion to the 17 years old debate. The temperature dependences of these three transitions are shown in Figure 1.2.9, and the entropy and enthalpy changes per base pair associated with all three transitions are shown in Table 1.2.4.

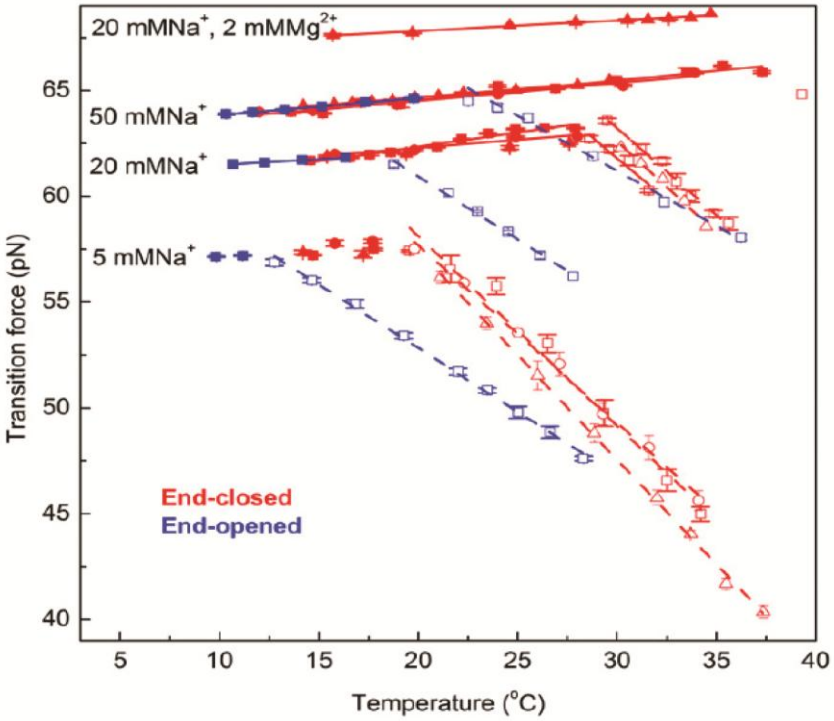


Figure 1.2.9 Temperature dependence of transition force for three distinct transitions. This figure is adopted from Zhang *et al.*(13).

Although all the three possible tension induced DNA overstretching transitions were identified, and the energy data during transitions from the B-DNA were measured, some concerns still remain regarding their relative stability due to lacking of direct observation of inter-conversion from one overstretched

structure to another. Some researchers proposed that the S-DNA is not an energetically stable structure compared to melted DNA; rather, it is likely a meta-stable structure with a long life time under tension (61,62). Therefore, it would be important to investigate the structural transition between these three overstretched DNA to see whether under certain condition, the S-DNA would be more stable than the other two structures.

Table 1.2.4 Comparison of entropy and enthalpy changes during different DNA overstretching transitions (13). The value of structural coefficient v is from our latest unpublished data (highlighted in red), v of thermal melting measurement is converted from SantaLucia *et al.* (20) for comparison (more details please refer to section 3.2.1).

	End-opend DNA				End-closed DNA				Thermal melting 50% GC
	B-to-S		B-to-ss peeling		B-to-S		B-to-2ss internal melting		
NaCl (mM)	50	20	50	20	50	20	20	5	5 - 50
ΔS cal/(K•mol)	-2.4 ± 0.2	-2.8 ± 1.7	16.4 ± 2.0	16.0 ± 1.8	-2.3 ± 0.2	-2.4 ± 1.2	16.6 ± 0.8	18.3 ± 1.5	~ 24
ΔH kcal/mol	1.1 ± 0.1	0.9 ± 0.5	6.3 ± 0.7	6.0 ± 0.7	1.1 ± 0.1	1.0 ± 0.4	5.8 ± 0.4	6.1 ± 0.8	8.2
v	0.47 ± 0.02		1.09 ± 0.15		0.48 ± 0.02		1.15 ± 0.16		0.66

Currently, our lab is working on such experiments, and I am involved by providing theoretical predictions for the experiments and explanations for experimental observations. All these theoretical works will be described in chapter 3, including transfer matrix calculation for equilibrium distributions of the respective DNA structural states and kinetics Monte Carlo (KMC) simulation of the possible DNA structural transitions. Besides, in an effort to provide insights to possible structural basis to understand the mysterious S-DNA, I have performed a novel full-atom steered molecular dynamics (SMD) simulation to produce an overstretched DNA by a quasi-equilibrium stretching approach.

1.3. Literature review on bacterial nucleoid-associated proteins (NAPs)

1.3.1 Introduction: NAPs in bacteria

Several thousand types of proteins are involved in the functioning of a bacterial cell (63). Among these bacterial proteins, a group of proteins, located at the bacterial nucleoid, known as nucleoid-associated proteins (NAPs), distinguished themselves by their unique properties. They are characterized by their DNA binding ability, are abundant in cell and are typically small in sizes (< 20 kDa). Through their DNA binding properties, NAPs play two important roles in bacterial cells: 1) they help to package chromosomal DNA into a well-organized and highly-compacted structures that allow them to fit into the small-sized bacterial cell; 2) they are involved in regulating expression of numerous genes at a global level. Henceforth, their presence is important for bacterial cell vitality and function.

Escherichia coli (*E. coli)* NAPs have been widely studied due to *E. coli*. being one of the most well-studied bacteria and a common pathogen related to human health. More than 300 protein species are associated with *E. coli* nucleoid (64,65), of which around 10 are identified as major NAPs due to their abundance and importance to cell survival (66). These 10 NAPs are HU (heat-unstable nucleoid protein), H-NS (histone-like nucleoid structuring protein), IHF (integration host factor), Lrp (leucine-responsive regulatory protein), Dps (DNA-binding protein from starved cells), Fis (factor for inversion stimulation), StpA (suppressor of td mutant phenotype A), CbpA (curved DNA-binding protein A), CbpB (Curved DNA-binding protein B) and DnaA (DNA-binding protein A). Most studies have been done on these 10 NAPs to understand the individual NAP structure, biochemical properties and cellular functions.

The functioning of NAPs depends on their DNA occupancy which is regulated by their competition for limited DNA binding sites. It has been found

that the individual population of different NAPs is not constant during cell growth and their population varies greatly depending on the cell growth phase (67). For example, at early stationary phase, the most abundant NAP is Dps while at the exponential phase, Fis becomes most abundant (67). It is therefore reasonable that competition of the limited DNA-binding sites by NAPs is regulated by their individual population which in turn depends on the growth phase (66,67). In addition the regulation of individual NAP population is often controlled by themselves or by other NAPs, as shown in the case of *E. coli* H-NS and StpA where both are able to silence their own gene expression and vice versa (68).

NAPs organize nucleoid chromosomal DNA through their specific or non-specific DNA interactions. These NAPs have very high DNA binding affinities with dissociation constants ranging from 25 to 250 nM (66). However, the interaction between NAPs and DNA is not trivial and is often complex and dynamic, allowing efficient packaging of the nucleoid DNA, fast response to environmental stimuli as well as providing complexity to the nucleoid architecture. In order to achieve dynamic response in regulating their *in vivo* functions, NAPs are endowed with multiple DNA-binding modes: HU can bend or stiffen DNA depending on HU concentration (69,70); H-NS is able to bridge DNA (71-74) and stiffen DNA depending on magnesium concentration (71,75); Lrp is capable of bridging DNA in its dimeric form (76) and wrapping DNA in its octameric form (77,78); StpA can bridge DNA in the presence of 10mM magnesium and stiffen DNA in the absence or at low concentration of magnesium (79); Fis is versatile as it can wrap, bridge and bend DNA (80-82). A cartoon illustration of different DNA-protein binding modes is summarized in Figure 1.3.1.

The multiple DNA-binding modes of NAPs do not only help to organize the nucleoid, but it also provides NAPs the unique abilities to perform gene regulation function. Many studies have shown that HU influences the expression of a wide range of genes involved in central metabolism and respiration (83-85); Lrp affects transcription of around 10% of genes in *E. coli*, which are involved in nutrient uptake, amino acid metabolism and phase-variable expression of pili (86-

89); Fis has a major impact on DNA transcription, replication and recombination (90-92); H-NS is a global gene silencer which can repress many genes, especially those that are laterally-acquired (93-96). Although these previous studies have advanced our understanding of the role NAPs play in gene regulation, how they mediate gene regulation is still not well understood. It is believed that the NAPs DNA-binding modes provide a mechanism in understanding how they perform their gene regulatory functions. This is supported by previous studies that showed H-NS performs its gene silencing through direct interaction with DNA.

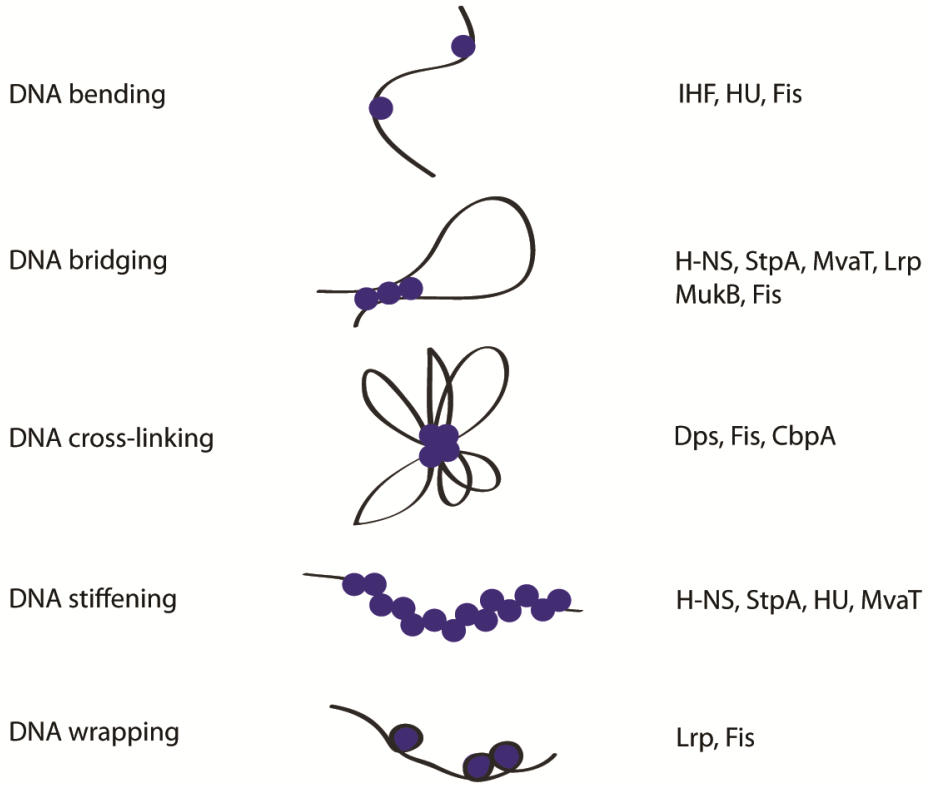


Figure 1.3.1 Cartoon illustration of different DNA-protein binding modes.

1.3.2 **H-NS family proteins in gram-positive bacteria**

As one of the earliest discovered NAPs (97), H-NS has been extensively studied for more than 35 years for both its DNA binding properties and its

influence on gene transcription. H-NS is a 15.6 kDa protein with 137 amino acids, which contains two functionally different domains, C-terminal DNA binding domain and N-terminal oligomerization domain, connected by a flexible linker (98-101) as illustrated in Figure 1.3.2. Studies have revealed that although larger oligomers exist, the *in vivo* active form is a dimer (102,103). Two possible dimerization forms of H-NS have been unveiled by NMR analysis (100,104) showing either parallel dimerization or anti-parallel dimerization exists depending on the orientation of the N-terminal domains as shown in Figure 1.3.2B&C.

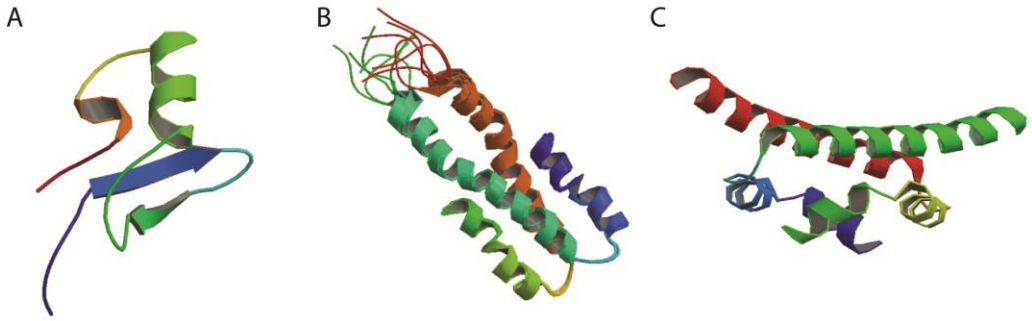


Figure 1.3.2 Solution structure of H-NS C-terminal binding domain (A) and N-terminal oligomerization domain (B) & (C). (A) The C-terminal binding domain H-NS_{91-137} was visualized by ribbon diagram, adopting from Protein Data Bank with pdb ID 1HNR, the structure is obtained from Shindo *et al.* (101). (B) Two N-terminal binding domains H-NS_{1-57} are visualized by ribbon diagram showing a parallel homodimer, adopting from Protein Data Bank with pdb ID 1LR1, the structure is obtained from Esposito *et al.* (100). (C) Two N-terminal binding domains H-NS_{1-46} are visualized by ribbon diagram showing a antiparallel homodimer, adopting from Protein Data Bank with pdb ID 1NI8, the structure is obtained from Bloch *et al.* (104).

It has been found that H-NS plays critical roles as a chromosomal DNA organizer and a global gene silencer (96). It can repressively influence the expression of up to 5% of the genes involved in central physiological process in *E. coli* (105,106). For example, H-NS silences genes that are responding to environmental changes (106,107) and also silences laterally-acquired foreign genes with AT-rich sequences (93).

H-NS family proteins, which are often defined by their capabilities to complement H-NS deficient mutants in *E. coli* (108), are widely conserved in gram-negative bacteria, such as StpA in *E. coli* (109), MvaT in *P. aeruginosa* (110), BpH3 in *Bordetella pertussis* (111) and VicH in *Vibrio cholera* (112). Although these proteins are often dissimilar to each other at sequence level, they usually exhibit remarkable structural and functional similarities with H-NS. For example, StpA, the H-NS paralogue in *E. coli*, shares 58% similarity with H-NS at amino acid level (109). While MvaT is only 18% similar to H-NS, it has similar domain organization and structure like H-NS (110). In general, these H-NS-like proteins often consist of a C-terminus DNA-binding domain and an N-terminus domain that mediates protein-protein interaction (99,100,110,113). As H-NS family proteins usually exist as dimers or higher-ordered oligomers, depending on solution condition and protein concentration, their oligomeric states are believed to be important for their functions (98,110,114).

1.3.3 Current DNA-protein binding modes in gene-silencing mechanism

NAPs organize DNA into various conformations and perform regulatory functions based on their DNA-binding properties. The distinct DNA-binding modes of individual NAP allow them to perform their specific functions.

H-NS, a global gene silencer, is believed to achieve gene-silencing function by its unique H-NS-DNA binding modes. As mentioned before, H-NS is able to bridge DNA to form DNA hairpins and loops at high magnesium conditions (> 5 mM), while it also stiffens DNA by forming rigid nucleoprotein filament at low magnesium conditions (0-2 mM) as shown in Figure 1.3.3 (71,72,75). Henceforth, two gene-silencing mechanisms have been proposed based on these two distinct H-NS-DNA binding modes: 1) for the H-NS-DNA bridging mode, it was proposed that RNA polymerase is trapped within a looped domain formed by

DNA-H-NS-DNA bridges (105,115,116) , causing RNAP to fail to continue elongation downstream as shown in Figure 1.3.4A. This trapping mechanism is supported by two experiments showing the co-localization of the H-NS and RNA polymerase at the bacterial promoters (117,118). A few years later, two more models of gene-silencing mechanisms were suggested (75) based on its H-NS-DNA stiffening mode. In one mechanism, after RNA polymerase binding to DNA, the DNA downstream were covered by H-NS rigid nucleoprotein filament which blocked RNA polymerase elongation as shown in Figure 1.3.4B. While in the other, the promoter regions were covered by H-NS filament, which blocked the access of DNA causing RNA polymerase failure to bind to the site thus leading to gene silencing as shown in Figure 1.3.4C.

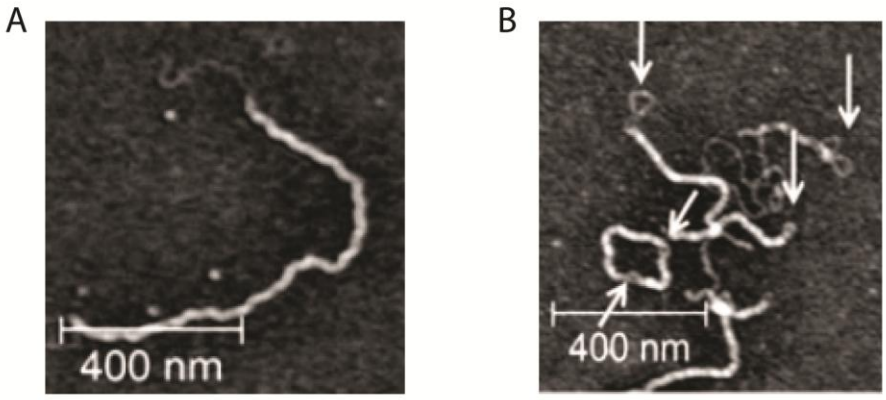


Figure 1.3.3 AFM imaging of H-NS-DNA complexes. (A) DNA stiffening caused by H-NS. 600 nM H-NS and 0.14 nM DNA were incubated for 4 hours in 50 mM KCl, pH 7.4.The brighter region indicated the DNA was coated with H-NS while the darker region indicated the naked DNA. (B) DNA bridging caused by H-NS. 600 nM H-NS and 0.14 nM DNA were incubated for 40 min in 50 mM KCl, 10 mM $MgCl_2$, pH 7.4. Large hairpins and loops were observed. These images are extracted from Liu *et al.* (71).

By now, the stiffening mode has been paid more attention as it is present in more physiological condition as compared to the bridging mode. Moreover, the H-NS rigid nucleoprotein filament was shown to be sensitive to temperature and pH changes while the bridging was not affected (71), which implies that the stiffening mode is closely related to H-NS gene-silencing since H-NS especially silences genes responding to environmental changes. This is further demonstrated

by a function-lost H-NS mutants study, where the ability of DNA bridging is still preserved by the function-lost H-NS mutants, while the mutants failed to stiffen DNA due to a lost in ability to form the rigid nucleoprotein filament (119). This observation indicates that the filament formation may serve as a structural basis for H-NS gene silencing. Furthermore, an H-NS and SsrB antagonization study also showed that the anti-silencing protein SsrB can only relieve the H-NS gene-silencing by displacing H-NS proteins on DNA under the condition when it forms H-NS nucleoprotein filament (120).

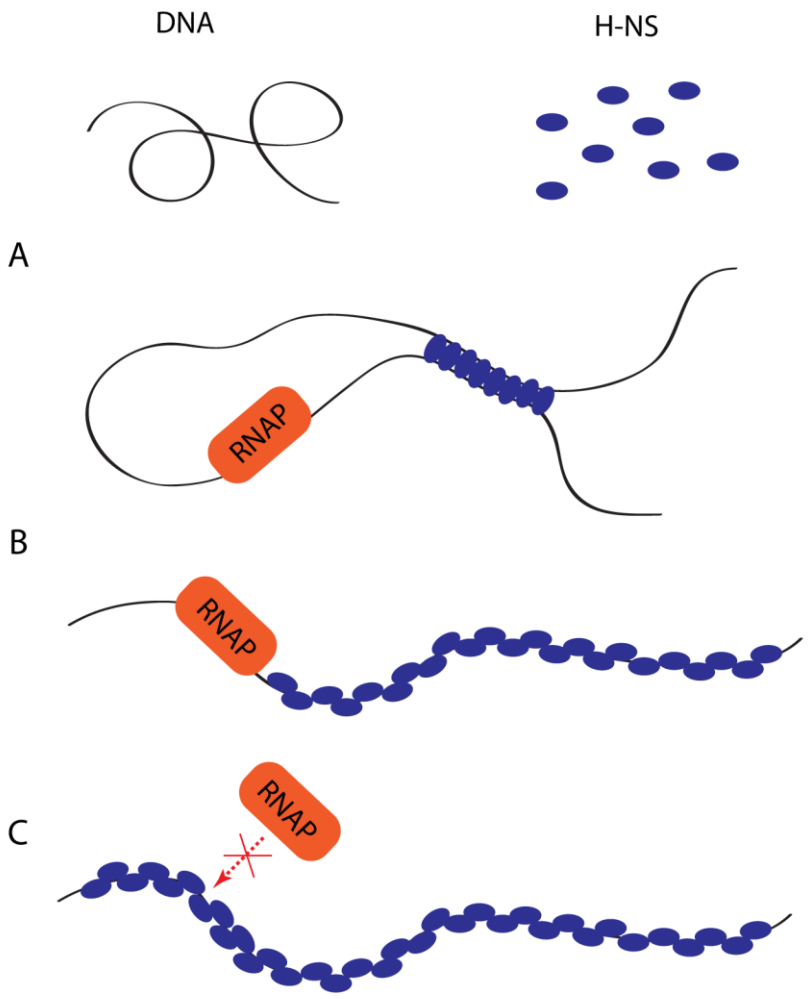


Figure 1.3.4 Two distinct gene-silencing mechanisms by two differernt H-NS-DNA binding modes are proposed. (A) Trapping mechanism by DNA-H-NS-DNA bridges. (B) DNA-H-NS nucleoprotein filament blocks the way of transcription by RNA polymerase. (C) DNA-H-NS nucleoprotein filament blocks DNA access by RNA polymerase.

Moreover, the formation of similar rigid nucleoprotein filament was also observed in H-NS family proteins in gram-negative bacteria such as *E. coli.* StpA and *P. aeruginosa* MvaT in physiological conditions (79,114), which implies a potentially universal gene-silencing mechanism by the formation of the rigid nucleoprotein filament.

1.3.4 Lsr2 protein in *Mycobacterium tuberculosis*

Although H-NS family proteins have been found widespread among gram-negative bacteria, such proteins were much less reported in gram-positive bacteria. Up to date, the only proposed H-NS family protein in gram-positive bacteria is Lsr2 in *Mycobacterium tuberculosis* (121).

Lsr2 is a small (~12 kDa), basic, abundant protein present in all mycobacterial genomes, including *Mycobacterium tuberculosis*, one of the most contagious and deadly human-directed pathogens. Previous studies have shown that similar to H-NS in gram-negative bacteria, Lsr2 is a DNA bridging protein (122), plays a role in chromosomal DNA organization (122) and performs regulatory functions as a gene-silencer preferentially binding to AT-rich sequences (123,124). It also influences multiple cellular processes such as cell wall biosynthesis and antibiotic resistance (125-127). Similar to the structure of H-NS, Lsr2 preserves a two-domain structure, including C-terminal DNA-binding domain and N-terminal dimerization domain (124,128). In addition, it has been shown that H-NS and Lsr2 exhibit similar binding specificity and display remarkable three-dimensional resemblance between their C-terminal binding domains (Fig. 1.3.5) (124). More surprisingly, Lsr2 is capable of complementing the phenotypes related to H-NS mutation in *E. Coli*, just like other H-NS family proteins in gram-negative bacteria (121).

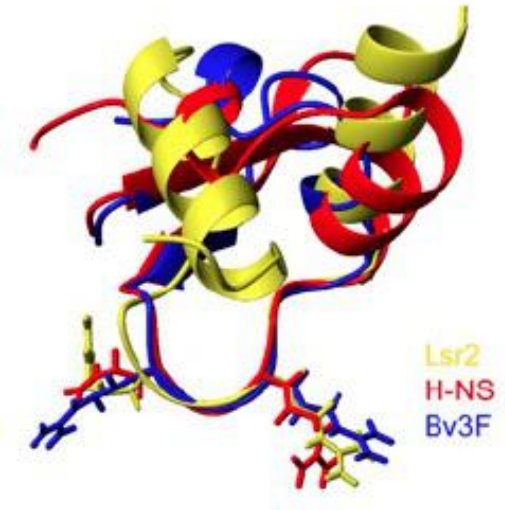


Figure 1.3.5 Lsr2 C-terminal domain shows remarkable three-dimensional resemblance with that of H-NS protein. This picture is adopted from Gordon *et al.* (124).

As discussed in the previous section, the NAP DNA-binding mode is the mechanical basis of how they organize DNA into various conformations and perform regulatory functions. An example has been shown in the case of the two distinct DNA-binding modes of H-NS. Therefore, to understand how Lsr2 mediates its various biological functions, it is important to know the Lsr2-DNA binding mode and its dependency on environmental factors. Although Lsr2 was shown to bridge DNA using AFM imaging (Fig. 1.3.6) (122), a more thorough investigation using a combination of single-molecule manipulation and imaging techniques is necessary to obtain a comprehensive picture of Lsr2-DNA binding mode(s). The power of such combination has been demonstrated in the cases of H-NS (71), StpA (79) and MvaT (114) as compared with previous AFM imaging studies that only revealed their DNA-bridging properties (23,32).

Moreover, it will be interesting to investigate if the gram-positive H-NS family protein Lsr2 is also able to form nucleoprotein filament, a novel feature shared among H-NS family proteins in gram-negative bacteria. If this is demonstrated, it would be the first evidence to show that the nucleoprotein filament formation capability is conserved for H-NS family proteins across different phylum groups.

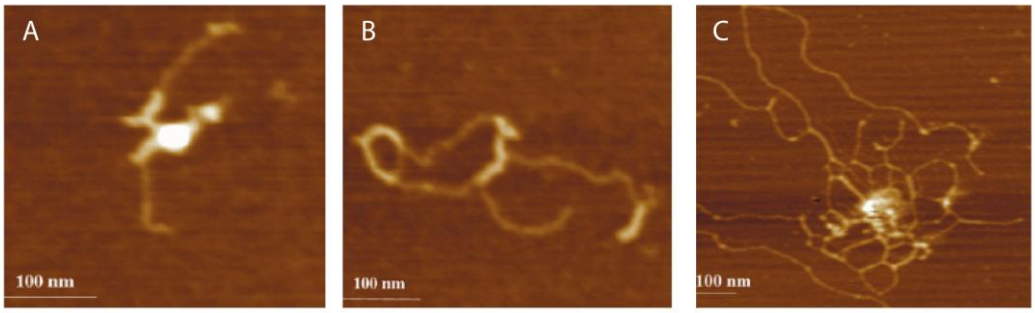


Figure 1.3.6 AFM imaging of Lsr2-DNA complex. (A)-(B), Lsr2-DNA complexes at low protein /DNA ratios (1 dimer per 260 bp). (C) Lsr2-DNA complexes at high protein /DNA ratios (1 dimer per 65 bp). These images are extracted from Chen et al. (122).

1.4. Single-molecule manipulation technologies and theoretical models

1.4.1 Introduction

The fast development in modern technologies makes it possible to observe and manipulate bio-particles in nanometer scale, which in the mean time have boosted research in the field of single-molecule biophysics. The most up-to-date techniques utilized are single-molecule force microscopy, which has become a useful tool to investigate the forces and motions related to bio-molecules and enzymatic activity. Among diverse techniques, three kinds of single-molecule force microscopy: optical tweezers, magnetic tweezers and atomic force microscopy, distinguish themselves as the most popular ones by their individual advantages.

In the following three sections, individual single-molecule force spectroscopy will be introduced in detail. The last section will be devoted to some theories related to the analysis of the data obtained from single molecule experiments.

1.4.2 Optical tweezers

The working principle of optical tweezers is that a dielectric particle experiences three-dimensional restoring force from a laser beam, and is trapped near the focus of the laser beam.

The dielectric particles are usually micrometer-sized microspheres, or even living cells such as bacteria and yeast, which can be directly manipulated by the optical traps. The laser beam is tightly focused with a microscope objective lens of high numerical aperture to generate a steep electric field gradient at the focus. The dielectric particle in the vicinity of the focus is an induced dipole and is

subjected to restoring force directed towards the focus due to the interaction with the electric field gradient. In the meantime, the dielectric particle also experiences a pushing force along the direction of light propagation due to photon scattering. To achieve a stable trapping, the restoring force must overcome the pushing force, which wants to displace the particle from the focus. Therefore, the stable trapping position usually locates at a region slightly past the laser focus. The force is linearly proportional to the displacement for small displacement (~ 150 nm) near the focus, where the optical tweezers can be treated as a spring, whose spring constant depends on the steepness of the electric gradient (129). The tighter the laser beam is focused, the steeper the gradient is, and the larger the restoring force is generated. For single-beam optical tweezers, the molecule of interested is often tethered between a surface and a laser trap (upper panel in Fig. 1.4.1); while for dual-beam optical tweezers, the molecule is tethered between two laser traps (lower panel in Fig. 1.4.1).

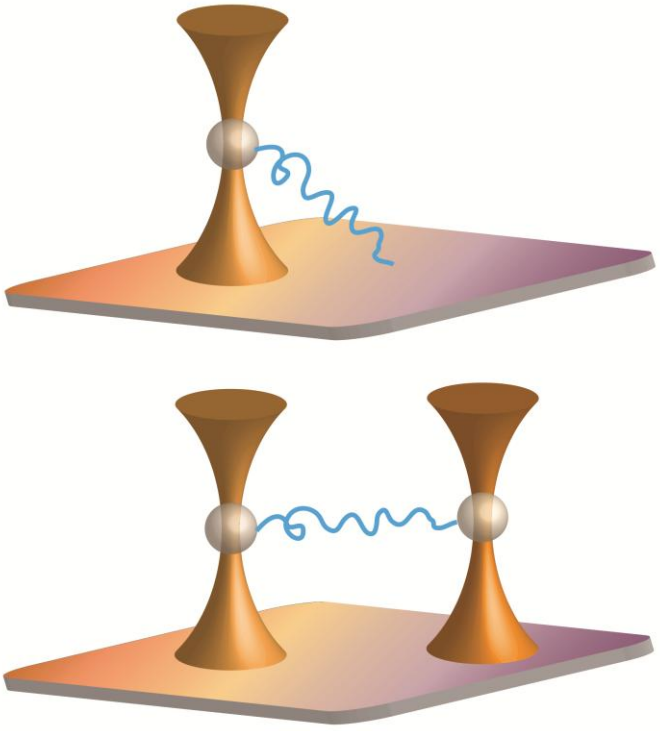


Figure 1.4.1 Sketch of general setup of optical tweezers. Upper panel shows single-beam trap, lower panel shows dual-beam trap.

Optical tweezers can achieve spatial resolution of sub-nanometer accuracy and temporal resolution of sub-millisecond accuracy, while exerting forces up to 100 pN on nanometer-sized to micrometer-sized particles. These merits make it a versatile tool suited for measurement of force and motion at the molecular level. However, there are also some limitations in the application of optical tweezers. First, as force is due to the interaction between the dielectric particles and the electric field gradient, any dielectric impurities in the buffer will be attracted to the focus. Therefore, it demands sample with high purity and exempts optical tweezers from long-time experiments. Second, the sample will be heated during the experiments due to the laser beam, which is a big problem for experiment sensitive to temperature. Third, optical damage caused by the laser trapping may also happen, although the detailed mechanism is not well understood yet.

1.4.3 **Magnetic tweezers**

The working principle of magnetic tweezers is that superparamagnetic beads placed in a magnetic field are subjected to force and torque generated by the magnetic field, and then transmit the force or torque to the molecule of interest.

The most used magnetic beads are commercially superparamagnetic beads, which are magnetized in the presence of an external magnetic field and become nonmagnetic upon removal of external magnetic field. The source of the external magnetic field is usually a pair of permanent magnets, which is the most convenient way to produce a magnetic field. The force $\vec{F}$ exerted on the magnetic beads is proportional to the gradient of the field $\vec{F} = (\vec{\mu} \cdot \nabla)\, \vec{B}$, where $\vec{\mu}$ denotes the magnetic moment of the superparamagnetic beads in the magnetic field $\vec{B}$. The torque placed on the bead $\vec{\tau} = \vec{\mu} \times \vec{B}$ will rotate the bead until its magnetic moment aligns with the external field, which allows the rotation of the bead that follows the rotation of the external field. The molecule of interested is often tethered between a surface and a magnetic bead for observation and manipulation.

A microscope objective used to catch an image of the movements of the magnetic bead is usually put underneath the whole setup. A simple illustrative sketch of the general setup of the magnetic tweezers is shown in Figure 1.4.2.

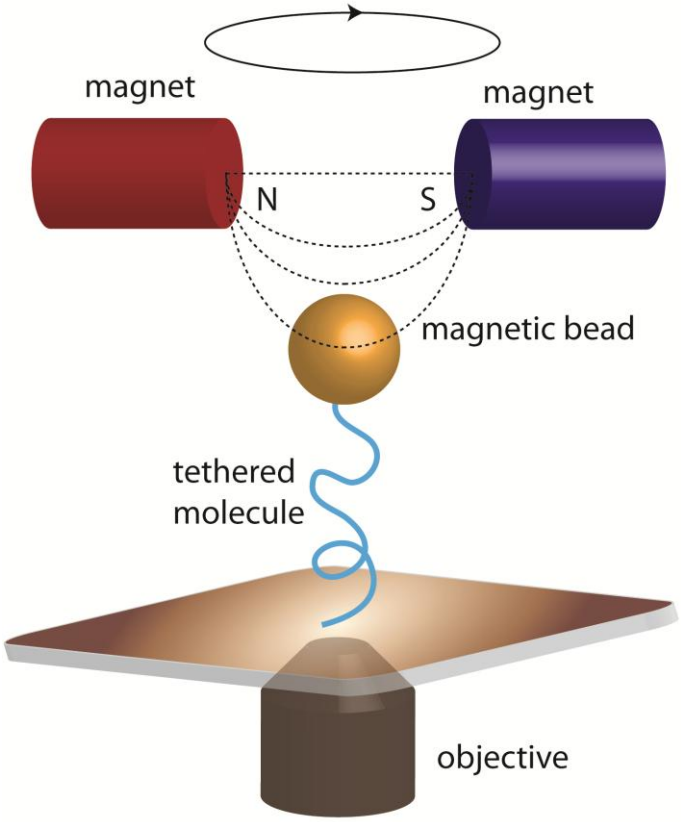


Figure 1.4.2 Sketch of general setup of magnetic tweezers.

Magnetic tweezers is able to impose force in excess of 1 nN (129) and simultaneously measure bead displacement with nanometer spatial resolution and millisecond temporal resolution. Compared to Optical tweezers, magnetic tweezers has its own strong points. First, it gives an intrinsic constant force. As the motion of the molecule is much smaller compared to the length scale of the changes of the gradient of magnetic field, the force imposed on the molecule can be considered constant for a given magnet position, which makes the magnetic tweezers a perfect force clamp. Second, magnetic tweezers is free from sample

heating and photodamage problems. Third, in addition to imposing force, it can also conveniently impose torque to samples, which is much harder for optical tweezers. This is very useful when studying topics related to DNA supercoiling. Fourth, magnetic manipulation is very selective for magnetic beads and insensitive to sample purity, which make it quite suitable for measuring in complex environment and long-time experiments. However, there are also some drawbacks for magnetic tweezers. As the image and data collection are based on CCD camera, the bandwidth and sensitivity are also limited by video-based detection, which prevents the direct detection and measurement of very fast or small displacement. Furthermore, although it can easily apply torque, it is overall less manipulable compared to optical tweezers.

1.4.4 **Atomic force microscopy**

Atomic force microscopy (AFM) is a dual-functional technique, which is not only capable of exerting force on sample as other single-molecule force spectroscopy, but also a imaging tool that can map the topography of a surface at sub-nanometer resolution.

The general working principle of AFM is that a thin sharp tip is mounted at the extremity of a flexible cantilever to probe multiple properties of the sample, including the mechanical properties and the topographic features. The lateral or vertical displacement of the sample with respect to the cantilever tip is controlled by a piezoelectric actuators either attached to the sample or the cantilever. The interaction between the surface and the cantilever tip generates force resulting in the deflection of the flexible cantilever, which is monitored by a laser beam. Further, the deflection of the laser beam is captured by a photodiode to achieve precise measurement of the deflection of the cantilever. Within a reasonable force range, the cantilever behaves like a linear spring obeying Hooke's law and exerting force $F = -kx$ depending on the stiffness k of the cantilever. A simple illustrative sketch of the general setup of AFM is shown in Figure 1.4.3.

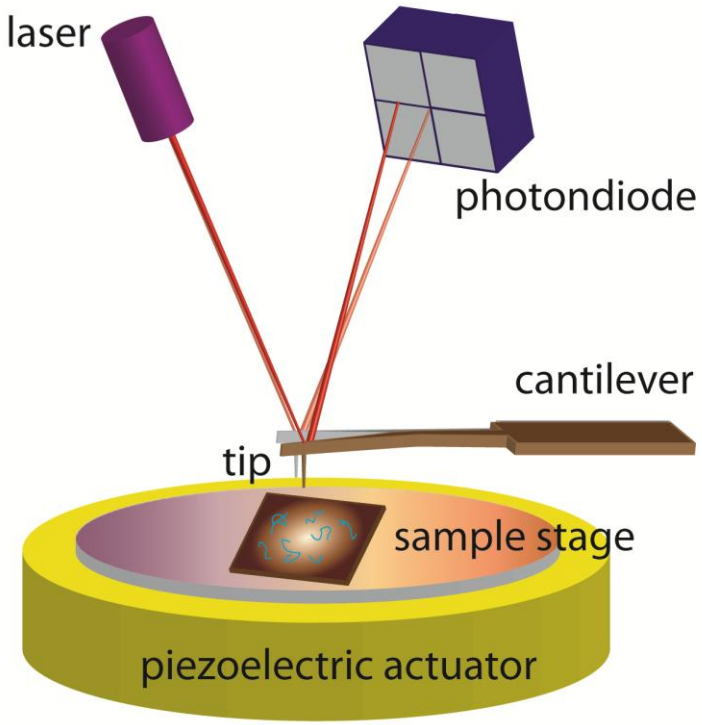


Figure 1.4.3 Sketch of general setup of AFM.

Although sharing similar working principle, the working ways for these two working modes are quite distinct from each other. As a molecular force probe, the sample of interested is tethered between the surface and the cantilever tip. The cantilever moves perpendicular to the sample surface, changing the force exerted on the sample. While it is used in the imaging mode, no tethering is made between the sample surface and the cantilever tip. The cantilever tip moves in parallel with the surface to scan the sample surface line by line. Different surface topography generates different force to the cantilever as it approaches to the surface; the signal of the distribution of the force then can be converted to the morphological information of the surface by the control software.

AFM-based force spectroscopy is featured by its high force (~ 10 nN) and high speed pulling, which has been proven to be a powerful tool to study the rupture of the unfolding of proteins (130) and nucleic acids (131). Further combining the imaging mode with the force probe mode will make it a ultimate

new technique among these three single-molecule force spectroscopy (132). However, as the high stiffness of the cantilever boosts the upper limit of the bandwidth, it also raises the lower limit (~ 10 pN), which makes it very difficult to study many biological processes and molecular structure transitions happen only under small forces such as protein refolding. Another challenging issue is to distinguish the signal of a molecule of interest from that of non-specific interactions, as well as the artificial signals due to background noise.

1.4.5 **Effects of DNA-distorting protein on DNA force response**

The most important information obtained from the single-molecule experiment is the force-extension curve of the molecule of interest, which can characterize the mechanical properties of the molecule such as DNA or a DNA-protein complex.

As discussed in section 1.2.3, below a force of 20 pN, the force-extension curve of naked B-DNA can be well characterized by the WLC mode. DNA-binding proteins, such as NAPs discussed in section 1.3, are able to non-specifically bind to DNA and distort DNA upon binding; therefore, they change the mechanical properties of DNA.

A theoretical model using a discrete wormlike-chain model and the transfer matrix method to investigate the effect of DNA-distorting proteins on DNA force response has been established by Yan *et al.* in 2003 (133). According to this model, different DNA-distorting proteins change DNA elasticity in different ways; based on which, these proteins can be roughly divided into three main groups: DNA-bending proteins, DNA-stiffening proteins and DNA-intercalating proteins. A DNA-bending protein causes a local bending upon association with DNA; a DNA-stiffening protein stiffens DNA when binding to DNA; a DNA-intercalating protein stretches DNA without introducing kinks when interacting with DNA. Their effects on the force response of DNA are shown in Fig 1.4.4.

For DNA-bending proteins (Fig. 1.4.4 A&B), the force-extension curve gradually shifts to higher force as more and more proteins bind to DNA (from bare to saturated). This large-force shifting behavior has been observed in the single-molecule measurement of the interaction between DNA and IHF (134,135), a typical DNA-bending protein in *E. coli*.

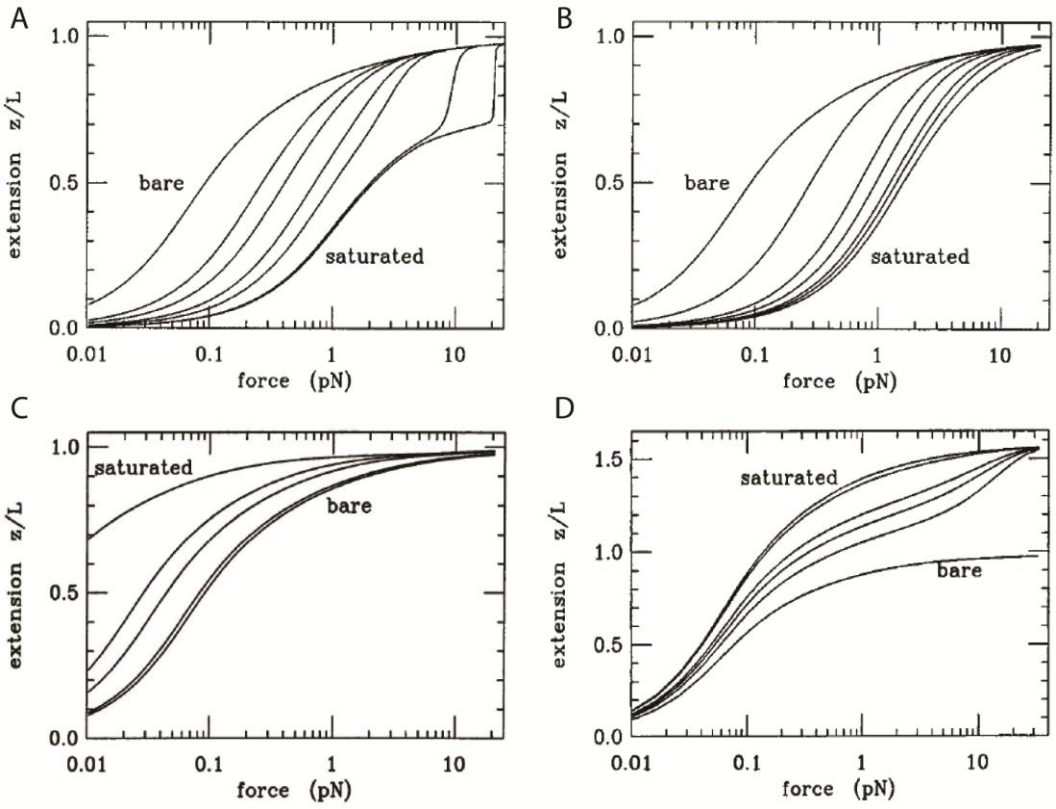


Figure 1.4.4 Effects of DNA-distorting proteins on DNA force response. (A) Force-extension curves of DNA-bending proteins for stiff protein-DNA complex, where unbinding happened under large force. (B) Force-extension curves of DNA-bending proteins for flexible protein-DNA complex. (C) force-extension curves of DNA-stiffening proteins. (D) force-extension curves of DNA-intercalating ligands. This figure is revised from Yan *et al.* (133).

For DNA-stiffening proteins (Fig. 1.4.4 C), the force-extension curve shows an increasing shift to lower force until binding saturation. The *E. coli.* protein H-NS (75) and its paralog StpA (79) are shown to follow this trend under certain experimental conditions.

For DNA-intercalating proteins (Fig. 1.4.4 D), not only the force-extension curve shifts to lower force, but also the contour length of the DNA-protein complex gets longer. This intercalating event was usually observed for drugs and proteins, inserting a ligand into the DNA backbone, such as YOYO-1(136).

Therefore, based on distinct changes of the force-extension of the DNA-complex, one can distinguish different effects by different DNA-distorting proteins.

1.5. Objective of the study

The work for this thesis is split into the following two parts.

The transition between different DNA forms is a fundamental and important process involved in essential biological processes such as DNA replication and transcription. In the first part of the thesis, the force-induced DNA structural transitions were thoroughly investigated by the transfer matrix method and kinetic Monte Carlo simulations. The simulation results are in good agreement with the experimental observations, providing a promising theoretical framework for DNA structure analysis. Furthermore, the B-to-S transition was simulated by a novel steered molecular dynamics simulation, which shed light on revealing the structure of the mysterious S-DNA.

In the second part of this thesis, the mechanism of DNA organization by *Mycobacterium tuberculosis* protein Lsr2 was studied using a combination of transverse magnetic tweezers and atomic force microscopy (AFM) imaging. The results showed that Lsr2 cooperatively binds to DNA forming a rigid Lsr2 nucleoprotein complex that restricts DNA accessibility and mediate further DNA condensation. In addition, the responses of Lsr2 nucleoprotein complex to environmental factors (salt concentration, temperature and pH) were studied over physiological ranges. These results provide mechanistic insights into how Lsr2 may mediate its gene silencing, genomic DNA protection and organization functions *in vivo*.

1.6. Organization of the thesis

Chapter 1 provided the biological and physical background of the research topics. Previous works related to DNA micromechanics and bacterial NAPs were extensively reviewed and the significances of the studies were also discussed. Then an overview of recently development in single molecule techniques was addressed, where the principles of three major techniques were discussed in detail. Further, a theoretical model used in analyzing the data obtained from single molecule experiments was also introduced. In the end of this chapter, the objectives of this thesis were presented.

Chapter 2 introduces two single-molecule techniques that have been used in this study: magnetic tweezers and AFM. The specification regarding the instruments used in the study was presented. Moreover, the experimental procedures and the sample preparation are also addressed. At last, the principle and the algorithm of the theoretical methods used in this thesis, including transfer matrix, kinetic Monte Carlo and steered molecular dynamics simulation, are also introduced.

Chapter 3 shows the results of the theoretical study of the force-induced DNA structural transition. These results are compared with experimental results, showing good agreement. In the end of this chapter, the potential S-DNA structure is probed by SMD simulation.

Chapter 4 reports the results of the study on the Lsr2-DNA interaction. They indicate that the capability of forming nucleoprotein filament is shared among H-NS like proteins in both gram-negative and gram-positive bacteria, and the nucleoprotein filament may be the structural basis for H-NS gene silencing.

In Chapter 5, we summarize the results of these two studies and draw some conclusions based on these findings.

CHAPTER 2 Methods and Material

2.1 Single-molecule manipulation by magnetic tweezers

2.1.1 Instrument introduction

The magnetic tweezers utilized in this work is our homemade transverse magnetic tweezers, whose basic setup is sketched in Fig. 2.1.1.

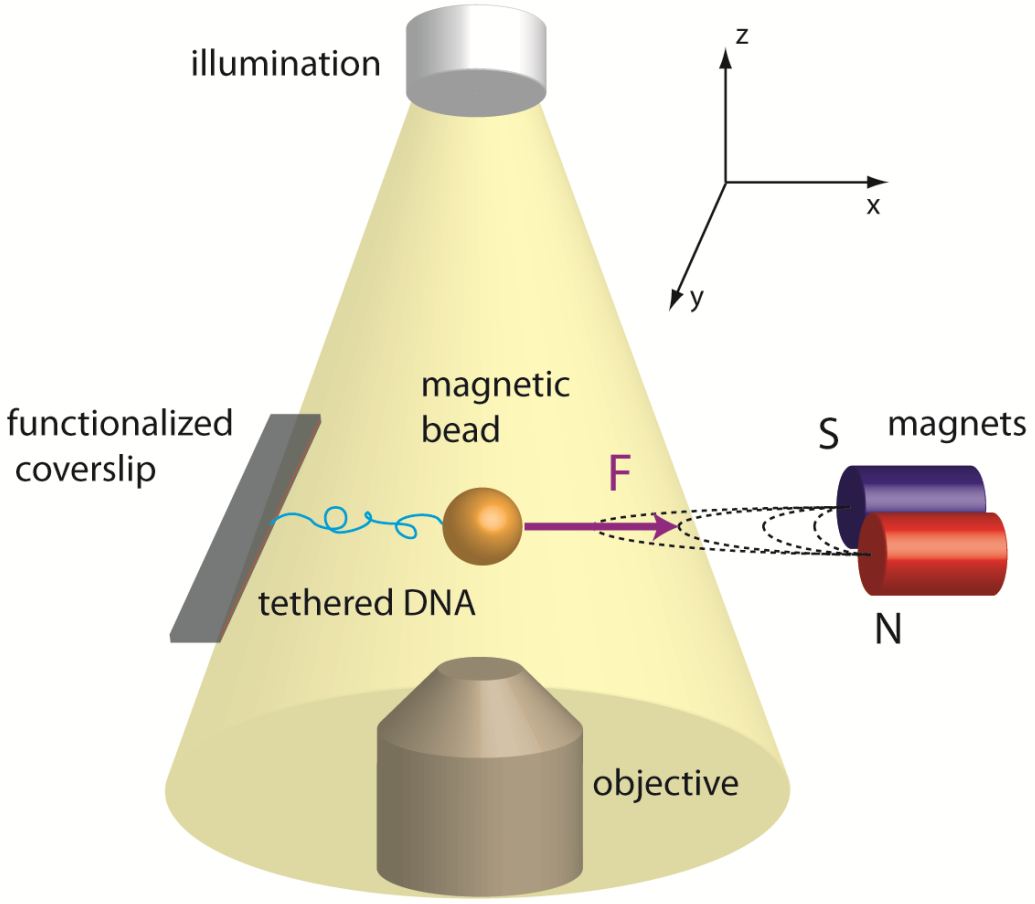


Figure 2.1.1 Sketch of our homemade transverse magnetic tweezers setup.

The bio-molecule of interest (e.g. DNA) is tethered between a functionalized coverslip edge and a paramagnetic bead. A pair of permanent magnets are placed in the same plane of the coverslip (x-y plane), generating a magnetic field gradient (x-direction) perpendicular to the coverslip edge (y-direction); therefore, the magnetic bead experiences a force in the direction towards the magnets. The magnitude of the force imposed on the sample is controlled by changing the distance between the coverslip edge and the magnets. Illumination comes from the top (z-direction) and a microscope objective underneath captures images and movements of the magnetic bead on its focal plane. The images obtained by the objective are then transferred to a CCD camera, and further input to the computer for image analysis. The images of this set of instrument are shown in Figure 2.1.2.

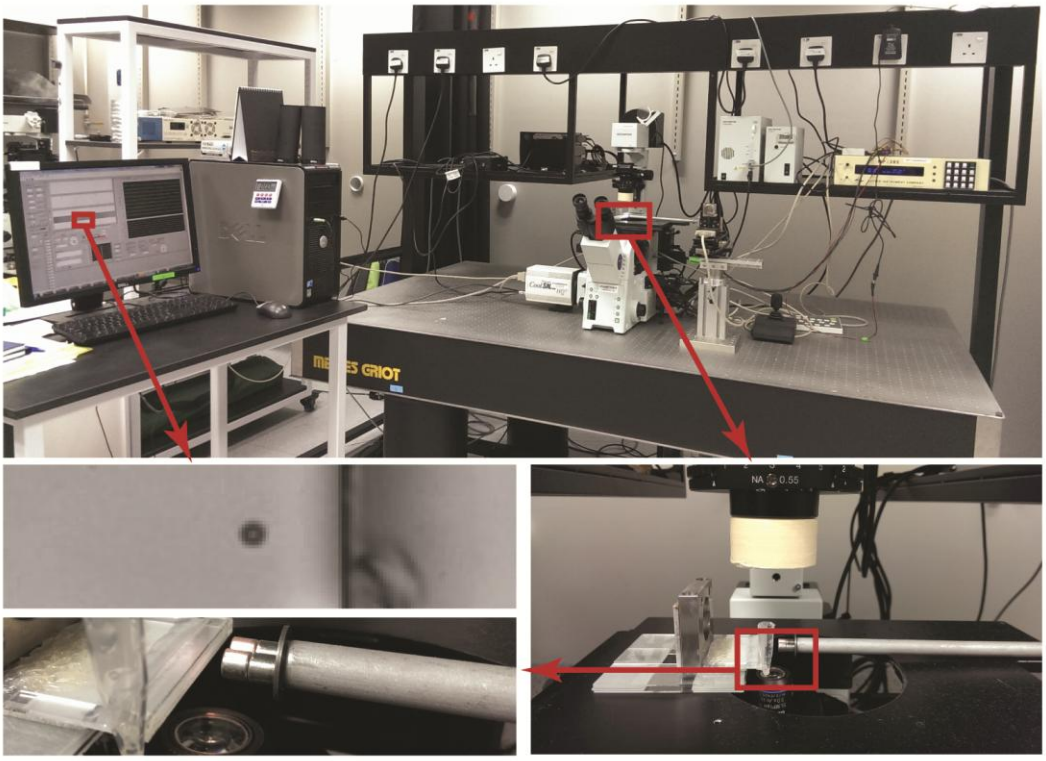


Figure 2.1.2 Images of homemade transverse tweezers. The upper panel is the overview of the whole setup of the instrument; the lower panels are the zoom-in pictures of the details.

2.1.2 **Force calibration of magnetic tweezers**

The force calibration of the magnetic tweezers is generally based on the pendulum model as illustrated in Figure 2.1.3.

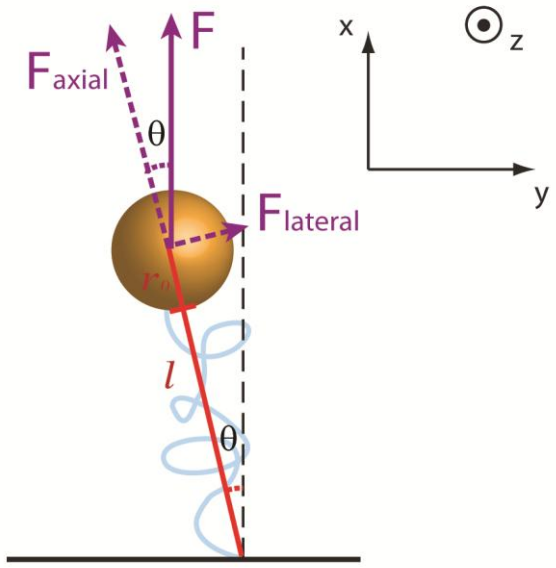


Figure 2.1.3 Pendulum model used in force calibration of magnetic tweezers.

In our setup, the motion of the bead in the z direction is confined by the magnets, so the bead only moves in the x-y plane. Force is along the x-direction, and the movement of the bead can be treated as harmonic oscillation. Therefore, the lateral stiffness is

$$\alpha = \frac{F_{lateral}}{(l + r_0)\, tan\theta} = \frac{F \sin\theta}{(l + r_0)\, tan\theta} \approx \frac{F}{l + r_0}$$

where l is the end-to-end distance of the polymer, r_0 denotes the radius of the magnetic bead and the approximation is due to that θ is always very small. According to equipartition relation, we have:

$$\frac{1}{2}\alpha\delta_y^2 = \frac{1}{2}k_B T$$

where δ_y^2 denotes the fluctuation of the magnetic bead in y-direction. Substituting F into the equation and rearranging it, we get

$$F = \frac{k_B T}{\delta_y^2}(l + r_0) \approx F = \frac{k_B T}{\delta_y^2} < x >$$

where $< x >$ denotes the average length of the tether along the force direction (x-direction). Therefore, with δ_y^2 and $< x >$, we can easily calculated the force imposed on DNA.

Another issue of interested in force calibration is how long should it take to accurately measure δ_y^2. As we focus on the motion in y-direction, the space-time distribution of the Brownian motion in this direction follows the one-dimensional Gaussian distribution,

$$\rho(y, \tau) = \frac{1}{2\sqrt{\pi D \tau}} e^{\frac{-y^2}{4D\tau}}$$

where D is the diffusion coefficient, τ is the diffusion time, which sets the time limit for it to reach Gaussian distribution. Therefore, in order to get the accurate δ_y^2, we have to make sure the measuring time $t \gg \tau$ is satisfied during the measurement.

According to the properties of the Gaussian distribution, we have

$$\delta_y^2 = 2D\tau$$

On the other hand, according to Einstein relation, we have

$$\xi D = k_B T$$

where $\xi = 6\pi\eta r$ is the viscous friction coefficient which depends on the shape and size of the bead, and η is the viscosity of the fluid and r is the radius of the bead. Taking the above two equations together to get rid of D and substituting ξ inside, we obtain

$$\tau = \frac{3\pi\eta r}{k_B T}\delta_y^2$$

Since we have $F = \frac{k_B T}{\delta_y^2} < x >$, we further get

$$\tau = \frac{3\pi\eta r}{F} < x >$$

From above equation we can see that the diffusion time scale at low forces is longer than that at high forces, which means that the measuring time under low forces needs to be much longer than that under high forces.

For example, in our experiment, we used a magnetic bead of 1.4 μm radius to pull a λ-DNA (48,502 bp, with contour length of 16.9 μm) with typical pulling force range between 0.1 to 20 pN. By calculating based on the above formula, the diffusion time τ is in the range of 1.19 ~ 0.01 seconds. Therefore, the time to make the measurement is usually above 30 seconds.

2.1.3 **Channel fabrication**

The channels used in the transverse magnetic tweezers are also our homemade channels as sketched in the upper panel in Figure 2.1.4.

As can be seen, three pieces of #1.5 glass slides (grey) are stuck together to form a notch in between. One #0 coverslip (black), whose edge (red) is functinalized with streptavidin, is inserted between the lower and the middle glass slides. A fraction of a piece of #0 cover glass (yellow) is precisely cut and attached to the upper and lower edges of the notch to seal the notch to form a channel. Then two glass capillary tubings (green) are inserted into the two holes located at both sides to allow buffer flowing in the channel. One plastic tube (orange) is attached to one of the capillary tubings; therefore, can hold amount of working buffer. The other capillary tubing can be attached to a syringe pump,

which can control the flow speed in the channel according to requirement. The images of the channel are shown in the lower panels in Figure 2.1.4.

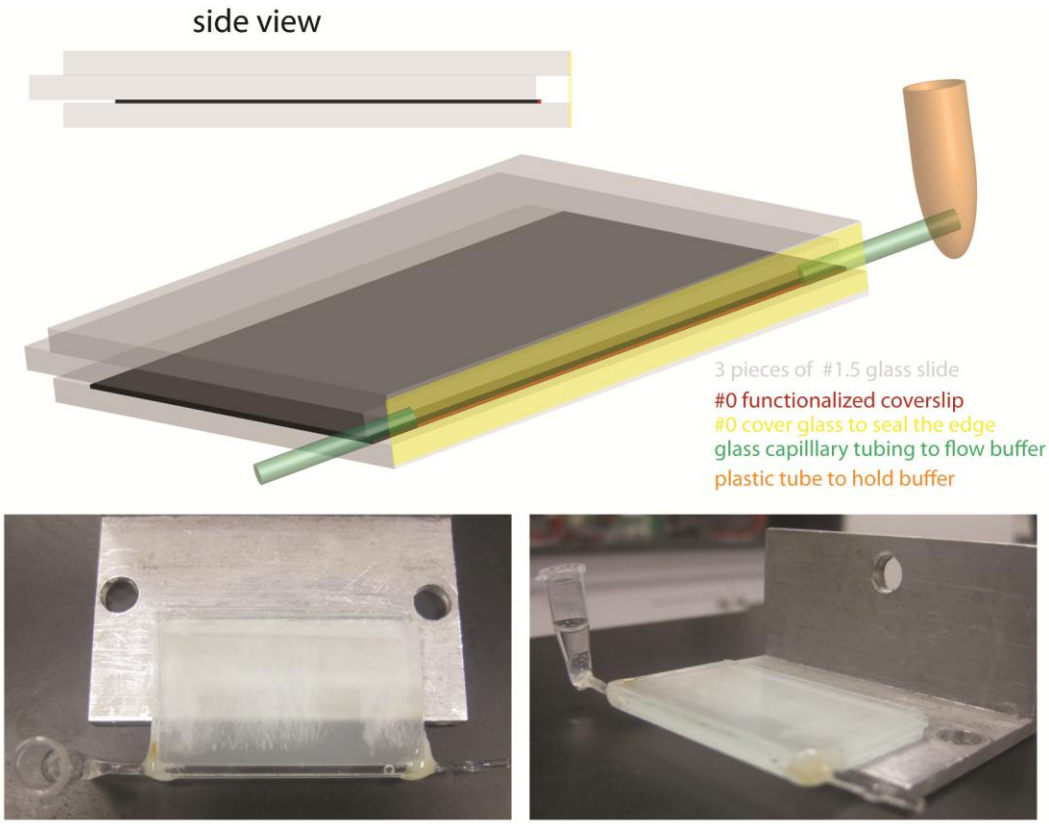


Figure 2.1.4 Illustration of homemade transverse channel. Upper panel is the cartoon sketch of the channel; lower panels are the image of the channel with top view (left) and side view (right).

2.1.4 **Protocol for functionalization of coverslip**

The protocol to functionalize the coverslip is as follows.

First of all, the edge of the coverslip has to be polished as to reduce the shadow of the edge of the image. The polished coverslip is firstly washed with detergent and immersed in methanol or acetone for 15 minutes sonication. This procedure is aimed to remove all the dusts and nasties on the coverslip surface. The cleaned coverslip then is boiled with solution containing hydrogen peroxide

and sulfuric acid at the ratio of H_2SO_4 : H_2O_2: H_2O = 1:1:4, for at least one hour at 110 ˚C. This process can further clean the surface and most importantly, hydroxylate the glass surface, preparing for the next step. The sulfuric acid treated coverslip then is thoroughly washed with deionized water, after which the edge is incubated with methanol solution containing 0.1% (3-Aminopropyl) triethoxysilane (APTES), for one hour at room temperature. The hydroxyl group on the surface is assumed to interact with APTES to form covalent bonds so that the surface is coated with ATPES. The coverslip has to be sonicated with pure methanol at least three times, each of which at least five minutes, to remove the excess corsslinked APTES polymers forming at the surface. Following this, the coverslip edge is incubated with 1X phosphate-buffered saline (PBS) buffer containing 5% glutaraldehyde, in 4 ˚C for at least 4 hours. Then glutaraldehyde is coated on the surface by covalently bonding to APTES. After this treatment, the surface is ready to coat with streptavidin or other antibody for tethering DNA. In our case, the coverslip edge is incubated with 1X PBS buffer with streptavidin concentration of 0.02 mg/ml, in 4 ˚C for at least 8 hours or overnight, and then incubated with 2% bovine serum albumin (BSA) buffer for long usage.

2.2 Atomic force microscopy imaging of DNA and DNA complex

2.2.1 Instrument introduction

The AFM used in this thesis is based on its imaging mode. The working principle of AFM imaging technique has already been discussed in section 1.4.4. Our AFM instrument is a commercially available one (5500 AFM, Agilent Technologies), whose basic setup is illustrated in Figure 2.2.1. Different from the setup shown in Figure 1.4.3, the piezoelectric actuator is attached to the cantilever rather than the sample stage for our instrument. A mica surface coated with glutaraldehyde is used as the background surface, the DNA molecule or DNA-protein complex are deposited on the gulutaraldehyde surface for scanning. The images of the instrument are shown in Figure 2.2.2.

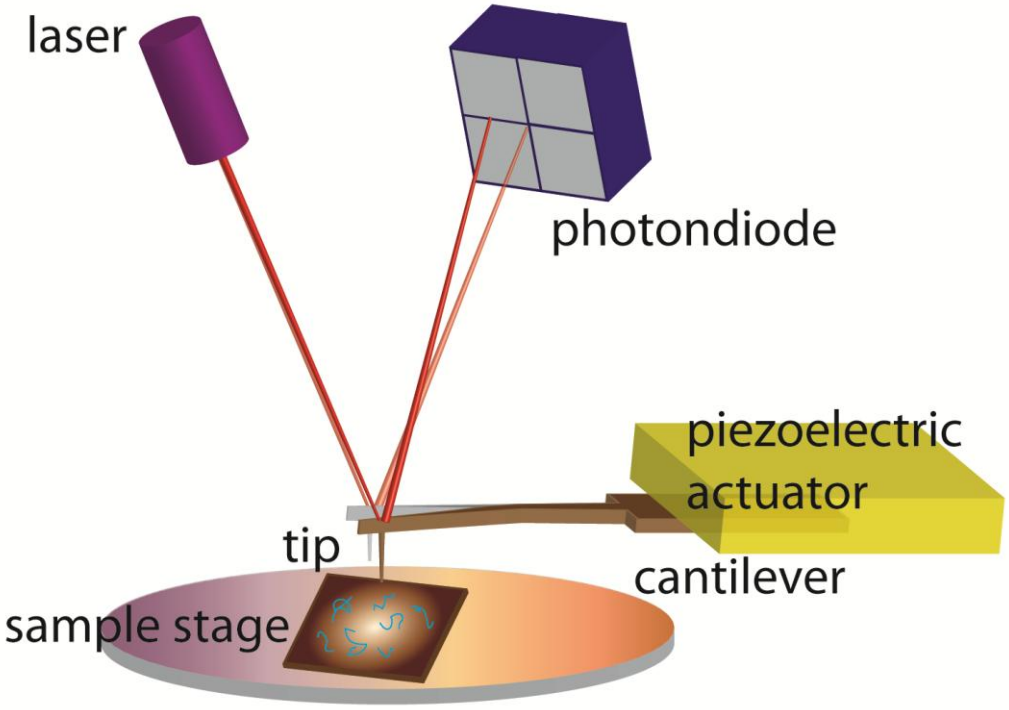


Figure 2.2.1 Cartoon sketch of the AFM setup used in the lab.

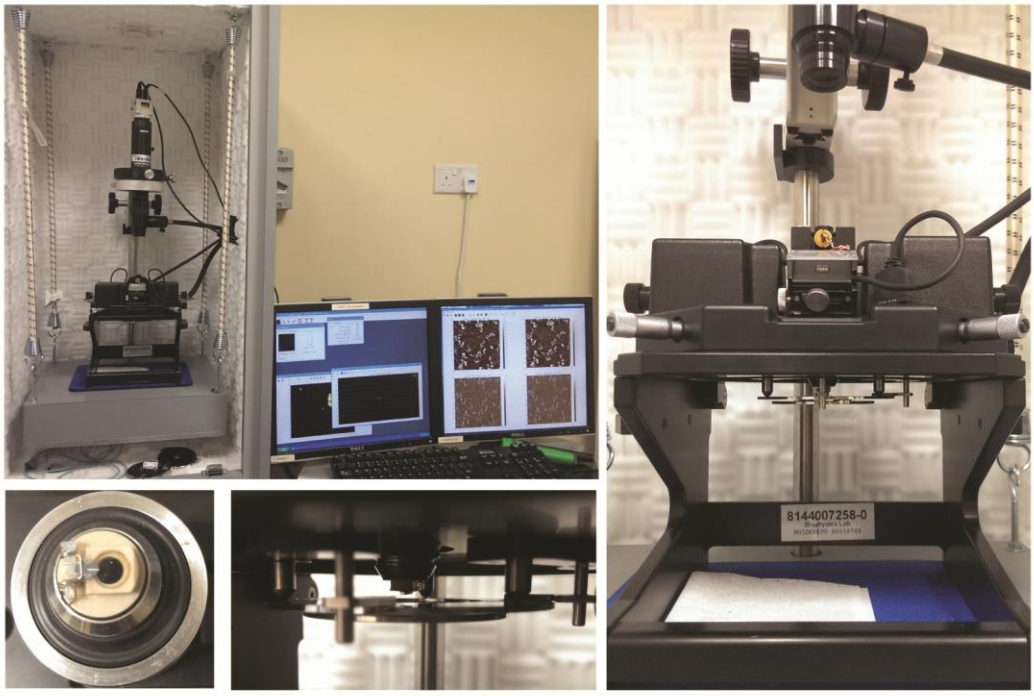


Figure 2.2.2 Images of the AFM instrument. Upper left panel shown the overview of the instrument; lower left panel is the AFM cantilever tips attached with piezoelectric actuator; lower middle panel shows the zoom in picture of working status of the cantilever and the sample stage; right panel is the zoom in picture of the AFM core parts.

2.2.2 **Functionalization of glutaraldehyde modified mica surface**

Although fresh mica surface is usually chosen to be the background surface for DNA imaging, the mica surface coated with glutaraldehyde is used in our experiment due to many reasons. As both DNA and the surface of fresh mica is negatively charged, the attachment between the DNA and the fresh mica surface has to be aided by divalent cation (e.g., Mg^{2+}). Therefore, the imaging buffer has to contain amount of divalent cation (usually 10 mM Mg^{2+}), which will interfere the interactions between DNA and protein, as shown in the case of DNA-HNS interaction (71), where two distinct DNA-binding modes are found in the absence or presence of divalent salt. While the glutaraldehyde surface is neutrally charged, the binding between the surface and the sample is based on the formation of the

covalent bond between the sample and the surface; therefore preventing the interference from the divalent cation. Furthermore, as DNA-protein complex is immobilized by the surface, they cannot diffuse into the solution when incubated on the mica, hence reduced the non-specific aggregation of the protein or DNA-protein complex, which further improves the clarity of the imaging.

The preparation of the glutaraldehyde coated mica surface is well established (71,137,138), and will be briefly discussed below. First, with one side of the mica stuck to the sample stage, a clean flat mica surface has to be prepared on the other side, which is usually done by quickly peeling off the outer-layers with a double-sided adhesive tape. The area of the mica surface is usually around 1cm × 1cm. Second, the surface is coated with a layer of ATPES by a 10 minutes incubation of 1% ATPES diluted with deionized water. Then the surface has to be rinsed with deionized water and dried with nitrogen gas. Then the ATPES treated mica is put in the oven for 20 minutes at 100 ˚C, which is for better immobilizing the ATPES molecule on the surface. Third, a volume of 50 μl of 1% glutaraldehyde diluted with 1XPBS buffer is incubated on the surface for 15 minutes and subsequently rinsed with deionized water and dried with nitrogen gas. Therefore, the glutaraldehyde modified mica surface is ready to use.

2.3 Application of transfer matrix, kinetic Monte Carlo and steered molecular dynamics simulation in theoretical studies

There are three methods employed in the theoretical studies of the force-induced DNA structural transitions: transfer matrix, kinetic Monte Carlo (KMC) and steered molecular dynamics (SMD) simulation. This section is mainly dedicated to a general introduction of the principles and the algorithms of these methods. More detailed implementation of these methods will be thoroughly described in section 3.2.

The transfer matrix method is usually used to solve problems in statistical mechanics. The basic idea is to write the partition function in the form

$$Z = p_1 \left(\prod_{i=1}^{N-1} T_i \right) p_N$$

where p_1 and p_N are usually k-dimensional vectors and the matrix T_i is $k \times k$ dimensional matrix called transfer matrix . The key feature of the system suitable for transfer matrix method is that the whole system can be divided into a sequence of subsystems that only interacts with adjacent subsystems, such as a lattice Ising model.

The DNA molecule is an ideal model for transfer matrix calculations as in the NN model only the stacking energy between adjacent base pairs are considered. The first application of transfer matrix method to DNA molecule was done by Yan *et al.* in 2003 (133), which correctly predicts the force response of the DNA polymer. Later in 2005, a more thorough study of applying transfer matrix to studying the DNA properties was published, where analytical solutions were provided (139).

The KMC method is a Monte Carlo method based simulation, which is able to compute the time evolution of a series of processes with given reaction rates. The flow chart of the KMC algorithm is illustrated in Figure 2.3.1.

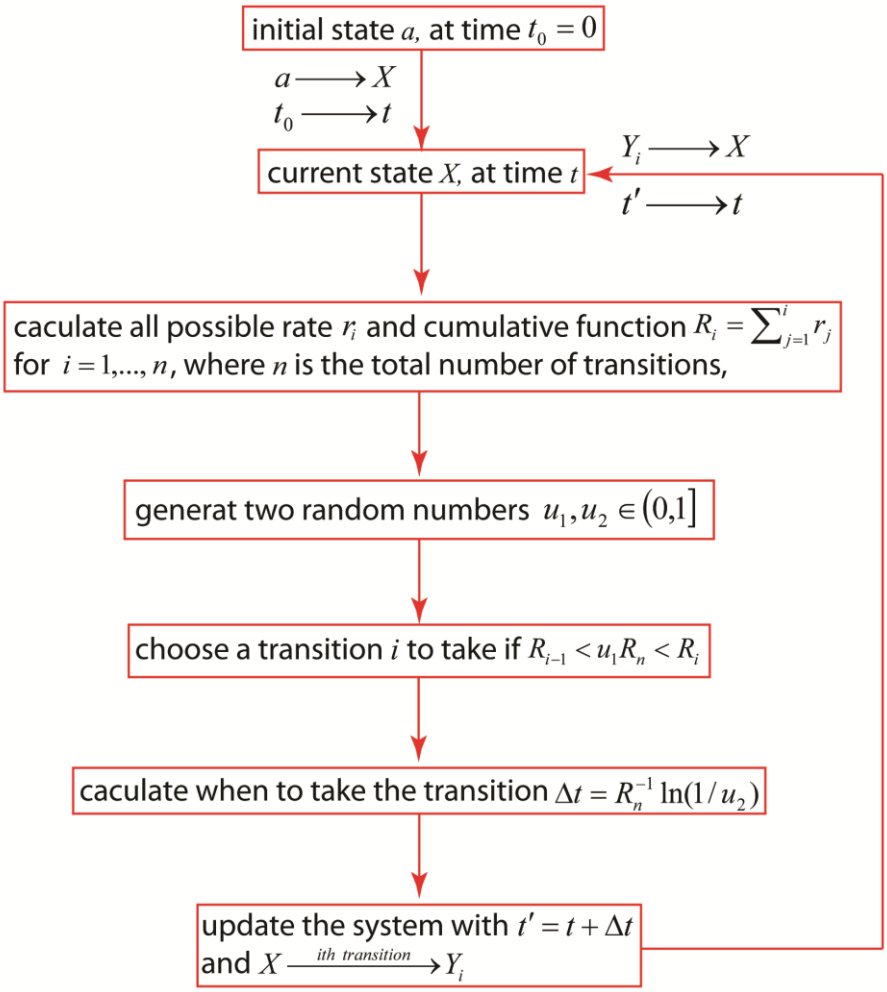


Figure 2.3.1 Flow chart of the KMC algorithm.

The way for a system with an initial state a at time t_0 to evolve is as follows. First, based on the current state X at time t, calculate all the possible transition rates r_i and all the cumulative functions $R_i = \sum_{j=1}^{i} r_j$ for i=1,2,...n, where n is the total number of transitions. Second, generate two random numbers $u_1, u_2 \in (0,1]$. Third, choose a possible state Y_i to evolve by a selection criteria $R_{i-1} < u_1 R_n < R_i$. Fourth, calculate the time to take the transition $\Delta t = R_n^{-1} \ln(1/u_2)$. Fifth, update the system with the current state Y_i at time $t' = t + \Delta t$. Then, repeat the steps all over again to let the system evolve with time as shown in Figure2.3.2.

The key feature of KMC algorithm is that if the reaction rate associated with the each process is correctly given, the distribution of the reaction rate follows Poisson distribution and different processes are independent, it can correctly predict the correct time scale for the evolution of the system. One thing to be noticed is that KMC method itself cannot predict the reaction rate.

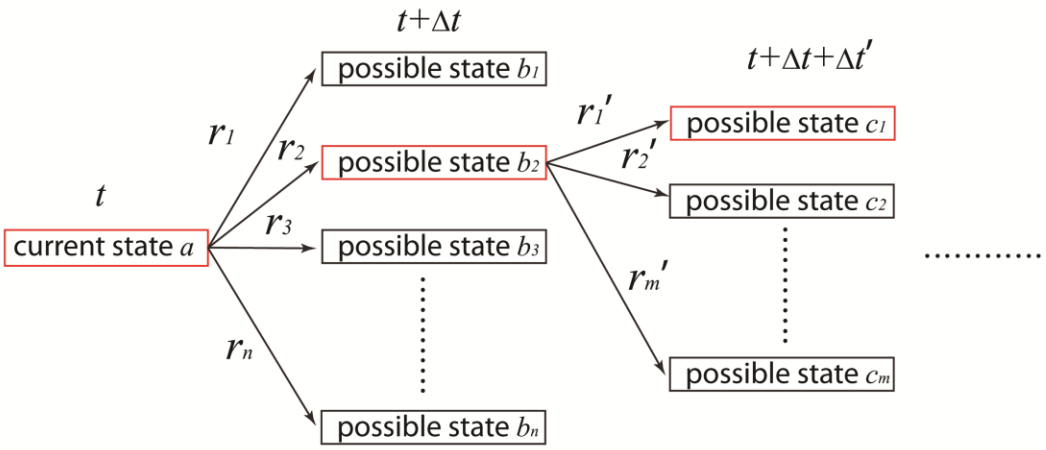


Figure 2.3.2 Time evolution of the system by KMC method.

The first paper describing the basic features of the KMC method was published by Young *et al.* in 1966 (140). Later in 1975, Bortz *et al.* independently developed KMC method for simulating the Ising model (141). Then in 1997, Gillespie *et al.* published a paper to apply this algorithm to chemical reactions (142). Nowadays, the KMC simulation is widely used in physics and chemistry to simulate non-equilibrium processes.

The molecular dynamics (MD) simulation is used to simulate the physical movements of atoms and molecules, based on numerically solving the Newton's equation of motion of all atoms in the system. A highly simplified flow chart of MD simulation algorithm is illustrated in Figure 2.3.3.

Given the initial positions and velocities of all atoms in the system, the potential energies of all the atoms in the system are obtained based on the

interaction between atoms. Then the force experienced by each atom can be calculated by $\vec{F}_i = -\nabla V(\vec{r}_i)$; therefore the acceleration on each atom is $\vec{a}_i = \vec{F}_i/m_i$. Subsequently, choose a very short time Δt, and make movements of all atoms according to equations of motion. At last, update the system with new position and velocity of each atom. Then the time evolvement of the system is achieved by iteratively solving the Newton's equation according to the potential energy of the system. Usually, the potential energy of the system is defined by a force field, which provides parameters of mathematical functions describing the potential energy for every type of atoms. Different force fields are build for different purposes. There are several classical force fields especially designed for proteins and DNA, such as AMBER and CHARMM.

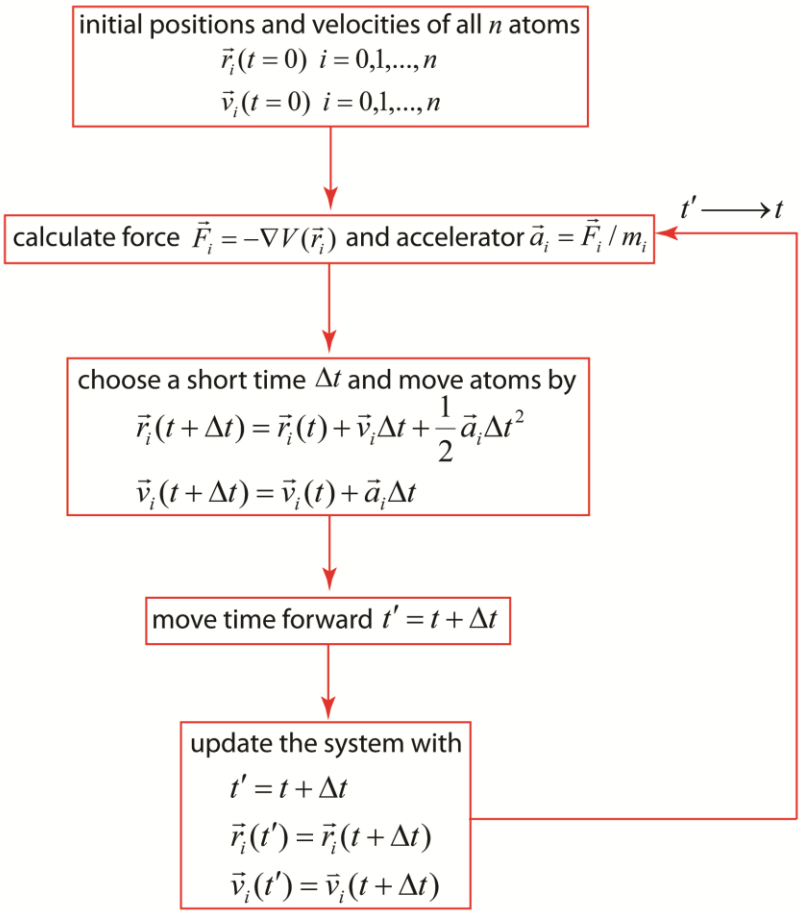


Figure 2.3.3 Simplified flow chart of MD simulation algorithm.

The algorithm of SMD simulation is essentially the same as that of MD simulation, where the main difference is that an external force is applied to the molecule. This is very useful for studying the mechanical unfolding of proteins or force-induced DNA structural transitions. SMD can provide a useful explanation and structural insight for single-molecule experiments, such as conformational changes of the stretched molecule. Typical steps for SMD simulation are as follows. First, identify a hypothesis for the conformational transition. Second, apply force to induce the transition. At last, analyze results to test the hypothesis.

CHAPTER 3 Theoretical studies of DNA structural transitions

3.1 Introduction

DNA undergoes an overstretching transition at a force around 65 pN, where its extension suddenly increases to ~ 1.7 times its contour length. The debate over this transition has been well elucidated recently revealing that three distinctive transitions are involved and subjected to different experimental conditions and DNA topologies. S-DNA, which appears under force larger than 65 pN, is a new form of DNA, whose properties are not well understood and characterized yet.

In this chapter I focused on the force-induced DNA structural transitions between different DNA structures under tension, namely B-DNA, 1ssDNA, 2ssDNA and S-DNA.

Firstly, the stabilities of different DNA structures were investigated by directly plotting out the phase diagram. Moreover, the transfer matrix method was used to calculate the equilibrium distribution of the B-DNA, 2ssDNA and S-DNA for a 7535 bp topologically closed DNA under different forces in different buffer condition. It successfully predicted the inter-conversion between S-DNA and 2ssDNA under high force in different salt concentrations. In addition, a kinetic Monte Carlo simulation was used in studying the kinetics of these transitions. It successfully reproduced the B-to-ss, B-to-S and S-to-ss transitions observed in experiments, but failed to describe the B-to-2ss and S-to-2ss transitions.

At last, the structure of S-DNA is probed by steered molecular dynamics using GROMACS. Although previous studies of MD simulation of the DNA overstretching transition have done similar work, the detailed structure of S-DNA is still a mystery unsolved due to many reasons. In our simulation, by a quasi-

equilibrium stretching process of an alternating GC sequence, a highly possible candidate for the mysterious S-DNA is obtained.

3.2 Methods

3.2.1 Energy analysis

Generally, there are two kinds of free energy involved in our systems, namely the base pair stacking energy present in dsDNA (B-DNA, S-DNA) and the force-induced conformational energy present in all DNA structures (B-DNA, S-DNA, 1ssDNA and 2ssDNA).

The base pair stacking energy for B-DNA g_B has already been discussed in section 1.2.2, which is dependent on sequence, temperature and salt concentration. Further, two sets of the B-DNA stacking energy have been provided by different research groups (20,46) as shown in Table 1.2.1 and Table 1.2.2 respectively. In our study, we used the value provided by SantaLucia (20) as shown in Table 1.2.1.

However, the stacking energy g_S for S-DNA is unknown. The derivation of g_S is a little bit tricky. The force-induced peeling transition from B-DNA to 1ssDNA can take two different pathways: i) directly peeling from B-DNA; ii) overstretching to S-DNA and then peeling from S-DNA (Fig. 3.2.1).

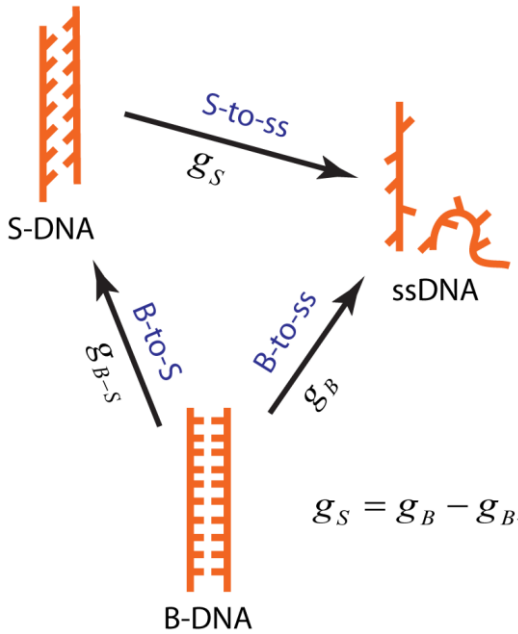


Figure 3.2.1 Two different pathways from B-DNA to ssDNA

Two different pathways do not alter the energy difference between the B-DNA and the 1ssDNA, therefore:

$$g_S = g_B - g_{B-S}$$

As we already know the B-DNA base pair stacking energy g_B, as long as the changes of stacking energies from B-DNA to S-DNA g_{B-S} is known, the S-DNA stacking energy can be obtained. The enthalpy and entropy changes during B-to-S transition have already been measured by single-DNA stretching experiment in different salt concentration for different DNAs as summarized in Table 3.2.1.

Table 3.2.1 Enthalpy and entropy values for the B-to-S transition in different salt concentration for different DNAs. These values are adopted from Zhang *et al.* (12,13).

NaCl [mM]	20	50	150	500	1000
ΔS [cal/(K•mol)]	2.8±1.7	2.4±0.2	3.8±0.6	3.0±0.7	3.6
ΔH [kcal/mol]	-0.9±0.5	-1.1±0.1	-0.9±0.2	-1.1±0.2	-1.036

Therefore, g_{B-S} can be easily obtained by the following equation:

$$g_{B-S} = \Delta H - T\Delta S$$

One thing to be mentioned here is that B-to-S transition is highly cooperative and it is shown to be not significantly dependent on DNA sequence, therefore, the stacking energy change of B-to-S transition does not have sequence dependence.

In our study, we used the enthalpy and entropy values in 1 M NaCl ($I_0 = 1\ \mathrm{M}$) as the standard value for B-to-S transition and we can calculate g_{B-S} in different salt concentrations by the following equation:

$$g_{B-S}(I) = g_{B-S}(I_0) + m \times \ln\left(I/I_0\right)$$

where $m = 0.086\ \mathrm{kcal/mol}$ was measured by experiment (12).

Therefore, with g_{B-S} known, we can easily calculate g_S for different DNA sequence in different temperature and salt concentrations. Note that although

g_{B-S} is not dependent on sequence, g_S is sequence-dependent because g_B is sequence-dependent.

The force response of B-DNA (x_B), S-DNA (x_S) and ssDNA (x_{ss}) has already been discussed in section 1.2.3 and 1.2.4. Therefore the conformational free energy can be easily calculated by

$$w_{B/S/ss}(f) = -\int_0^f x_{B/S/ss}(f')\,df'$$

for B-DNA, S-DNA and 1ssDNA respectively. For two parallel ssDNA under tension (2ssDNA), each of the strands holds only half of the force, and the conformational free energy will be termed

$$w_{2ss}(f) = 2w_{ss}(\frac{f}{2}) = -2\int_0^{\frac{f}{2}} x_{ss}(f')\,df'$$

Then we can summarize the total free energy for different DNA structures:

$$G_B = w_B - g_B$$

$$G_S = w_S - g_S$$

$$G_{ss} = w_{ss}$$

$$G_{2ss} = w_{2ss}$$

If we choose the B-DNA conformational free energy as the reference energy, we then can rewrite the energies to be

$$G_B = w_B$$

$$G_S = w_S + g_{B-S}$$

$$G_{ss} = w_{ss} + g_B$$

$$G_{2ss} = w_{2ss} + g_B$$

During a DNA structural transition, the free energy change is the sum of the stacking energy difference (Δg) and force-dependent conformational energy difference (Δw):

$$\Delta G = \Delta g + \Delta w$$

where

$$\Delta w = -f\Delta x$$

$$\Delta g = \Delta H - T\Delta S + \Delta g_{salt}$$

where Δx is the DNA extension change per base pair during transition, ΔH is the enthalpy change during the transition, ΔS is the entropy change during the transition, Δg_{salt} is salt dependent free energy change, shown to be $\Delta g_{salt}(I) = vk_BT\Delta x \ln(I/I_0)/l_B$ by Rouzina *et al.* (60), where $l_B = 0.71$ nm is the Bjerrum length and v is a structural coefficient describing the salt dependence of the transition. Therefore the free energy ΔG could be further broken down to

$$\Delta G = \Delta H - T\Delta S + vk_BT\Delta x \ln(I/I_0)/l_B - f\Delta x$$

At the transition force $f = f_{\mathrm{ov}}$, $\Delta G = 0$, we have

$$\Delta H - T\Delta S + vk_BT\Delta x \ln(I/I_0)/l_B = f_{\mathrm{ov}}\Delta x$$

The transition force f_{ov} will be changed in different temperature and salt conditions. Therefore, taking derivative of $\ln(I/I_0)$ at both sides, we get

$$v = \left(\frac{\partial f_{\mathrm{ov}}}{\partial \ln(I/I_0)}\right)\frac{l_B}{k_BT}$$

This relation is usually used to obtain v for a transition in single-DNA stretching experiment, in which the overstretching force f_{ov} is measured in different salt concentration at a fixed temperature (12,13).

In the thermal melting experiment in SantaLucia *et al.* (20), the salt dependent free energy is

$$\Delta g_{salt}(I) = m \times ln\,(I/I_0) = v k_B T \Delta x_{B-ss} ln\,(I/I_0)/l_B$$

where $m = 0.114$ kcal/mol, as mentioned in section 1.2.2, $\Delta x_{B-ss} \approx 0.2$ nm. Therefore, the structural coefficient for thermal melting transition $v = m\frac{l_B}{k_B T \Delta x_{B-ss}} = 0.66$, as indicated in Tab. 1.2.4.

Further, taking derivative of *T* at both sides, we get

$$\Delta S = -\left(\frac{\partial f_{\mathrm{ov}}}{\partial T}\right)\Delta x + v k_B \Delta x \ln\,(I/I_0)/l_B$$

where $v k_B \Delta x \ln\,(I/I_0)/l_B$ is a constant for a given salt concentration. Similarly, the entropy change can be obtained by measuring the overstretching force at different temperature at a fixed salt concentration. This is how entropy change of a transition is measured by single-DNA stretching experiment.

3.2.2 **Transfer matrix calculation**

The transfer matrix method was usually used for calculation of physical parameters in equilibrium state. As mentioned in section 2.3, the key feature of transfer matrix is to divide the whole system into a sequence of subsystems only interacting with neighbors. This is just the case of DNA, whose base pair stacking energy is only related to adjacent base pairs. Therefore, we use this method to calculate the equilibrium distribution of different DNA states under different force and buffer conditions. Here, we don't consider the 1ssDNA from the peeling transition, because it is a kinetic process, which we will deal with in the next section using a different method. Therefore, a topologically closed DNA (two strands of the double helix were looped together) can have three different states if stretched: stable B-DNA under small force, S-DNA by B-to-S transition or 2ssDNA by internal melting of the double helix as shown in Figure. 3.2.2. When equilibrium is reached, the percentage of each state is fixed.

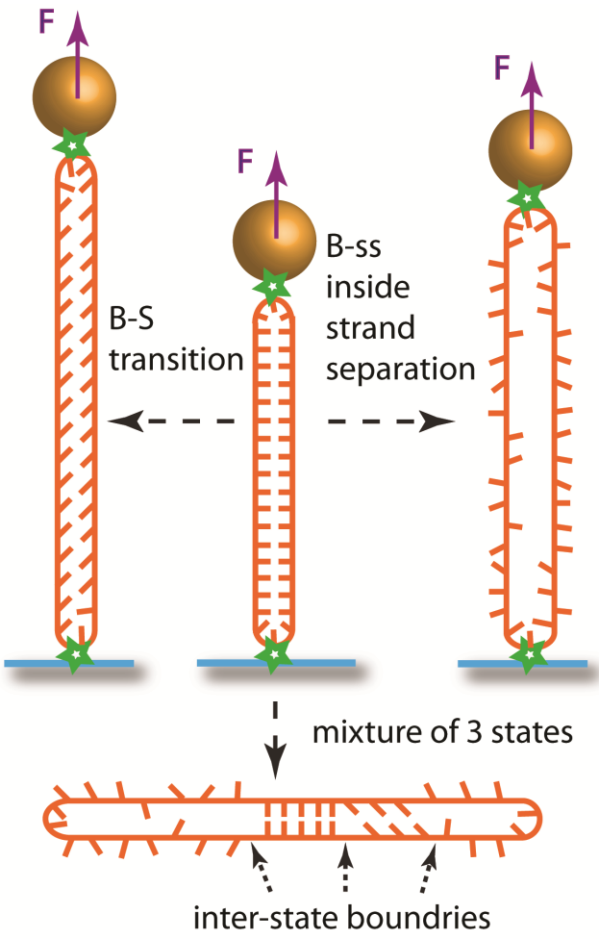


Figure 3.2.2 Cartoon illustration of possible states for a nickless end-closed DNA under force.

In our model, each base pair can adopt one of these three states B-DNA ($b_i = 0$), S-DNA ($b_i = 1$) or 2ssDNA ($b_i = 2$). So the free energy describing an N base pair DNA molecule under force is

$$\begin{aligned}\Delta G = \sum_{i=1}^{N}\{&w_B(f)\delta_{b_i,0} + w_S(f)\delta_{b_i,1} + w_{2ss}(f)\delta_{b_i,2} + g_B(i)\delta_{b_i,2}\delta_{b_{i+1},2} \\ &+ g_{B-S}\delta_{b_i,1}\delta_{b_{i+1},1} + C_{B-2ss}\left(\delta_{b_i,0}\delta_{b_{i+1},2} + \delta_{b_i,2}\delta_{b_{i+1},0}\right) \\ &+ C_{B-S}\left(\delta_{b_i,0}\delta_{b_{i+1},1} + \delta_{b_i,1}\delta_{b_{i+1},0}\right) \\ &+ C_{S-2ss}\left(\delta_{b_i,1}\delta_{b_{i+1},2} + \delta_{b_i,2}\delta_{b_{i+1},1}\right)\}\end{aligned}$$

where the first three terms describe the force dependent conformational free energy of B-DNA, S-DNA and 2ssDNA respectively; $g_B(i)$ describes the

sequence dependent base pair stacking energy for B-DNA; g_{B-S} describes the energy difference between B-DNA and S-DNA; C_{B-2ss}, C_{B-S} and C_{S-2ss} set the boundary energies (the energy to create a boundary) of B-2ss boundary, B-S boundary and S-2ss boundary as shown in the lower part of Figure 3.2.3. The boundary energies originate from the stacking energies of base pairs, and we chose $C_{B-2ss} = 4k_BT$, $C_{B-S} = 7k_BT$ and $C_{S-2ss} = 5k_BT$ for simplicity in our model. The reason why we chose these values will be discussed in section 3.3.3.

Therefore, the partition function is termed to be

$$Z = \sum_{\{b_i\}} \exp\left(-\frac{\Delta G}{k_BT}\right)$$

$$= \sum_{b_1=0,1,2} \cdots \sum_{b_N=0,1,2} \prod_{i=1}^{N-1} \exp\{-[w_B(f)\delta_{b_i,0} + w_S(f)\delta_{b_i,1}$$
$$+ w_{2ss}(f)\delta_{b_i,2} + g_B(i)\delta_{b_i,2}\delta_{b_{i+1},2} + g_{B-S}(i)\delta_{b_i,1}\delta_{b_{i+1},1}$$
$$+ C_{B-ss}(\delta_{b_i,0}\delta_{b_{i+1},2} + \delta_{b_i,2}\delta_{b_{i+1},0})$$
$$+ C_{B-S}(\delta_{b_i,0}\delta_{b_{i+1},1} + \delta_{b_i,1}\delta_{b_{i+1},0})$$
$$+ C_{S-ss}(\delta_{b_i,1}\delta_{b_{i+1},2} + \delta_{b_i,2}\delta_{b_{i+1},1})]/k_BT\}$$

If we define

$$T_i = \begin{bmatrix} e^{-(w_B(f)/k_BT} & e^{-(w_B(f)+C_{B-S})/k_BT} & e^{-(w_B(f)+C_{B-ss})/k_BT} \\ e^{-(w_S(f)+C_{B-S})/k_BT} & e^{-(w_S(f)+g_{B-S})/k_BT} & e^{-(w_S(f)+C_{S-ss})/k_BT} \\ e^{-(w_{2ss}(f)+C_{B-ss})/k_BT} & e^{-(w_{2ss}(f)+C_{S-ss})/k_BT} & e^{-(w_{2ss}(f)+g_B(i))/k_BT} \end{bmatrix}$$

we can rewrite the partition function to be

$$Z = p_1 \left(\prod_{i=1}^{N-1} T_i\right) p_N$$

where $p_1 = (1 \quad 1 \quad 1)$ and $p_N = (1 \quad 1 \quad 1)^T$. As our system doesn't have boundary conditions, Z here is the total partition function, which sums over all possible configurations. Then the partition function can be calculated numerically for any given DNA sequence. The percentage of B-DNA is

$$P_B = \frac{n_B}{N} = \frac{\sum_{i=1}^{N} \delta_{b_i,0}}{N} = \frac{\frac{\partial \ln Z}{\partial(-w_B(f)/k_BT)}}{N} = -\frac{k_BT}{N}\frac{\partial \ln Z}{\partial w_B(f)}$$

Similarly, the percentages of S-DNA and ssDNA are

$$P_S = -\frac{k_BT}{N}\frac{\partial \ln Z}{\partial w_S(f)}$$

$$P_{ss} = -\frac{k_BT}{N}\frac{\partial \ln Z}{\partial (w_{2ss}(f))}$$

To numerically calculate these values, we compute

$$P_B = -\frac{k_BT}{N}\frac{\ln Z(w_B(f)+\delta) - \ln Z\big(w_B(f)\big)}{\delta}$$

$$P_S = -\frac{k_BT}{N}\frac{\ln Z(w_S(f)+\delta) - \ln Z\big(w_S(f)\big)}{\delta}$$

$$P_{ss} = -\frac{k_BT}{N}\frac{\ln Z((w_{2ss}(f)+\delta) - \ln Z\big((w_{2ss}(f)\big)}{\delta}$$

where δ is a very small quantity, reflecting a very small variance of the energy.

3.2.3 **Kinetic Monte Carlo simulation**

The KMC simulation in this thesis was mainly used for studying the kinetics of the force-induced DNA structural transitions. Generally, there are two types of transition: one is the melting process from B-DNA (B-to-ss and B-to-2ss transition) or S-DNA (S-to-ss and B-to-2ss transition), where a base pair of a dsDNA is destroyed, the other one is the highly cooperative B-to-S transition, where the base pairs are still maintained. In the melting transition, there are also two types, one is the peeling transition from B-DNA or S-DNA for DNA with open ends, the other one is the internal melting transition occurred in B-DNA or S-DNA for topologically closed DNA. In order to compare the KMC simulation

with experimental results, the exact DNA sequences and constructs in single-molecular experiments were used respectively for all transitions.

As discussed in section 2.3, the key point of KMC simulation is to know the reaction rates for all possible reactions for a given state.

An N base pair DNA is indexed from 1 to N from the 5' end to the 3' end along one strand. For the peeling transition, the DNA is assumed to be free from nicks and peeling can only occur from the two ends generating two B-ss forks i, j ($0 \leq i \leq j \leq N$) as shown in Figure 3.2.3. Each fork can either move forward to open a B-DNA base pair or backward to close a B-DNA base pair. The state of the system can be defined by (i, j) and four possible transitions from the current state are (i, j+1), (i, j-1), (i+1, j), (i-1, j).

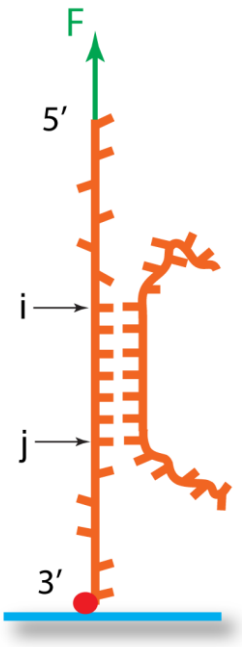


Figure 3.2.3 Theoretical models for peeling transition from dsDNA (B-DNA or S-DNA).

When a B-DNA base pair was destroyed leaving 1ssDNA under tension, the free energy change is

$$\Delta G_{B-ss}(i,f) = G_{ss}(f) - G_B(i,f) = w_{ss}(f) + g_B(i) - w_B(f)$$

Then the free energy associated with the 5' fork *i* movement is

$$\Delta G_{B-ss}^{5'}(i,f) = \sum_{n=1}^{i} \Delta G_{B-ss}(n,f)$$

and with the 3' fork *j* movement is

$$\Delta G_{B-ss}^{3'}(j,f) = \sum_{m=1}^{j} \Delta G_{B-ss}(m,f)$$

Taking together, the free energy associated with a state under force f is

$$\Delta G_{B-ss}(i,j,f) = \Delta G_{B-ss}^{5'}(i,f) + \Delta G_{B-ss}^{3'}(j,f)$$

With these energies, we then can define the reaction rates for these four possible reactions: $r_o(i,f)$ for transition from (*i*, *j*) to (*i*+1, *j*) with opening a base pair at *i* fork, $r_c(i,f)$ for transition from (*i*, *j*) to (*i*-1, *j*) with closing a base pair at *i* fork, $r_o(j,f)$ for transition from (*i*, *j*) to (*i*, *j*-1) with opening a base pair at *j* fork and $r_c(j,f)$ for transition from (*i*, *j*) to (*i*, *j*+1) with closing a base pair at *j* fork. Since the occurrence of the base pairs opening and closing obeys a Boltzmann distribution, the opening and closing rates at *i* fork or *j* fork satisfying

$$\frac{r_o(i,f)}{r_c(i,f)} = e^{-\Delta G_{B-ss}(i,f)/k_BT} \qquad \frac{r_o(j,f)}{r_c(j,f)} = e^{-\Delta G_{B-ss}(j,f)/k_BT}$$

In our case, we chose

$$r_o(i,f) = r_{B-ss} e^{(w_B(f)-g_B(i))/k_BT}$$

$$r_o(j,f) = r_{B-ss} e^{(w_B(f)-g_B(j))/k_BT}$$

$$r_c(i,f) = r_c(j,f) = r_{B-ss} e^{w_{ss}(f)/k_BT}$$

where r_{B-ss} is the attempting rate in the simulation for B-to-ss peeling transition, representing the microscopic rate for a base pair to move together or apart in the absence of tension. This is an adjustable parameter, whose value does not change the Boltzmann distribution of the states. Here we chose $r_{B-ss} = 5 \times 10^8$ s^{-1} so

that the time scale of the results is comparable with single-molecule experiment. Furthermore, two special cases should be treated individually. The first one is i=1 or j=576, in which $r_c(i=0,f)=r_c(j=576,f)=0$. The second one is that only one closed base pair left, therefore $r_o(i,f)=r_o(j,f)=r_{B-ss}e^{w_B(f)/k_BT}$. With these reaction rates, we then can follow the Gillespie's algorithm (142) to do the KMC simulation as discussed in section 2.3.

The model we used for S-to-ss peeling transition is exactly the same as one in the B-to-ss transition by substituting in the S-DNA stacking energy $g_S(i)$ and the S-DNA conformational free energy $w_S(f)$. Two forks *i*, *j* are also generated by peeling from the two ends and four possible states transitions are involved due to fork movements. Similarly, the free energy change in the S-to-ss transition, where an S-DNA base pair was destroyed, is obtained by

$$\Delta G_{S-ss}(i,f)=G_{ss}(f)-G_S(i,f)=w_{ss}(f)+g_S(i)-w_S(f)$$

And he free energy associated with a state under force f is

$$\begin{aligned}\Delta G_{S-ss}(i,j,f)&=\Delta G^{5'}_{S-ss}(i,f)+\Delta G^{3'}_{S-ss}(j,f)\\&=\sum_{n=1}^{i}\Delta G_{S-ss}(n,f)+\sum_{m=1}^{j}\Delta G_{S-ss}(m,f)\end{aligned}$$

Further, the four reaction rates associated with base pair opening and closing at *i* and *j* forks are chosen to be

$$r_o(i,f)=r_{S-ss}e^{(w_S(f)-g_S(i))/k_BT}$$

$$r_o(j,f)=r_{S-ss}e^{(w_S(f)-g_S(j))/k_BT}$$

$$r_c(i,f)=r_c(j,f)=r_{S-ss}e^{w_{ss}(f)/k_BT}$$

where the attempting rate for S-to-ss peeling transition $r_{S-ss}=5\times10^8$ s^{-1} is chosen to be the same as one in the B-to-ss peeling transition.

For the B-to-2ss internal melting transition, bubbles can be created in any location along the DNA sequence. We use n_i to denote the state of the ith base pair and it can either be in the closed state ($n_i = 0$) or open state ($n_i = 1$). Therefore, we define the state of a N bp DNA by an array of indices $\{n_i\}$, with i ranging from 1 to N. There are N possible state transitions according to the current state $\{n_i\}$. The opening rate or the closing rate of a base pair depends on the states of the adjacent base pairs.

For a base pair n_i, there are four different combinations depending on the states of its adjacent base pairs: $n_{i+1} = n_{i-1} = 0$ or $n_{i+1} = n_{i-1} = 1$ or $n_{i+1} = 0, n_{i-1} = 1$or $n_{i+1} = 1, n_{i-1} = 0$.

If in the case $n_{i+1} = n_{i-1} = 0$, where two adjacent base pairs are all closed, the initial formation of a bubble requires forming two B-2ss boundaries (i-1,i+1) as shown in Figure 3.2.4. Therefore, the free energy changes for creating a initial bubble inside is

$$\Delta G_{B-2ss}(i, f) = w_{2ss}(f) + 2C_{B-2ss} - w_B(f)$$

where $C_{B-2ss} = 4k_BT$ is the B-2ss boundaries energy as defined in section 3.2.2. The free energy that describes the system has already been discussed in section 3.2.2.

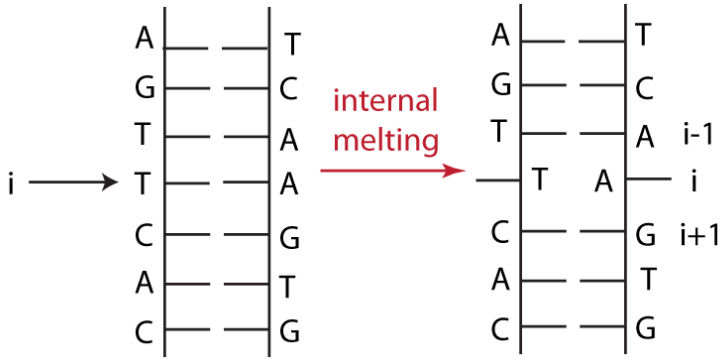


Figure 3.2.4 Initial formation of a bubble breaks two base pairs stacking energy and creates two boundaries i-1, i+1.

The opening rate and closing rate for this base pair is

$$r_o(f) = r_{B-2ss} e^{(w_B(f) - 2C_{B-2ss})/k_B T}$$

$$r_c(f) = r_{B-2ss} e^{w_{2ss}(f)/k_B T}$$

If in the case $n_{i+1} = 0, n_{i-1} = 1$ or $n_{i+1} = 1, n_{i-1} = 0$, it resembles the i or j fork movement in the peeling transition we discussed before with free energy changes of opening a base pair

$$\Delta G_{B-2ss}(i, f) = G_{2ss}(f) - G_B(i, f) = w_{2ss}(f) + g_B(i) - w_B(f)$$

And the opening rate and the closing rate is chosen to be

$$r_o(i, f) = r_{B-2ss} e^{(w_B(f) - g_B(i))/k_B T}$$

$$r_c(f) = r_{B-2ss} e^{w_{2ss}(f)/k_B T}$$

If in the case $n_{i+1} = n_{i-1} = 1$, there is no stacking energy involved, therefore the free energy changes of opening a base pair is

$$\Delta G_{B-2ss}(i, f) = G_{2ss}(f) - G_B(i, f) = w_{2ss}(f) - w_B(f)$$

And the opening rate and the closing rate is chosen to be

$$r_o(i, f) = r_{B-2ss} e^{w_B(f)/k_B T}$$

$$r_c(f) = r_{B-2ss} e^{w_{2ss}(f)/k_B T}$$

Similarly, we can use this model to calculate all the energies and reaction rates involved in the S-ro-2ss transition by substituting with the S-2ss boundary energy $C_{S-2ss} = 5k_B T$ as defined in section 3.2.2, S-DNA stacking energy $g_S(i)$ and the S-DNA conformational free energy $w_S(f)$.

Last, for B-to-S transition, there is no breaking of the base pair, but changes of the stacking energy. Similarly, the DNA state can be described by an array of indices $\{n_i\}$ ($i = 1,2 \ldots N$), denoting whether the ith base pair is in B-form ($n_i = 0$) or S-form ($n_i = 1$). For a N base pair DNA, there are N possible state

transitions according to current state $\{n_i\}$. We can use the same algorithm as one employed in the internal melting transition to simulate the state evolvement by substituting with the B-S boundary energy $C_{B-S} = 7k_BT$, the sequence-independent stacking energy difference between B-DNA and S-DNA g_{B-S} and the S-DNA conformational free energy $w_S(f)$.

3.2.4 **Steered molecular dynamics simulation**

The SMD simulation is performed to probe the structure of S-DNA by simulating the structural transition from B-DNA to S-DNA. This simulation is carried out in GROMACS, a versatile package to perform MD simulation and the force field adopted in this study is Amber99.

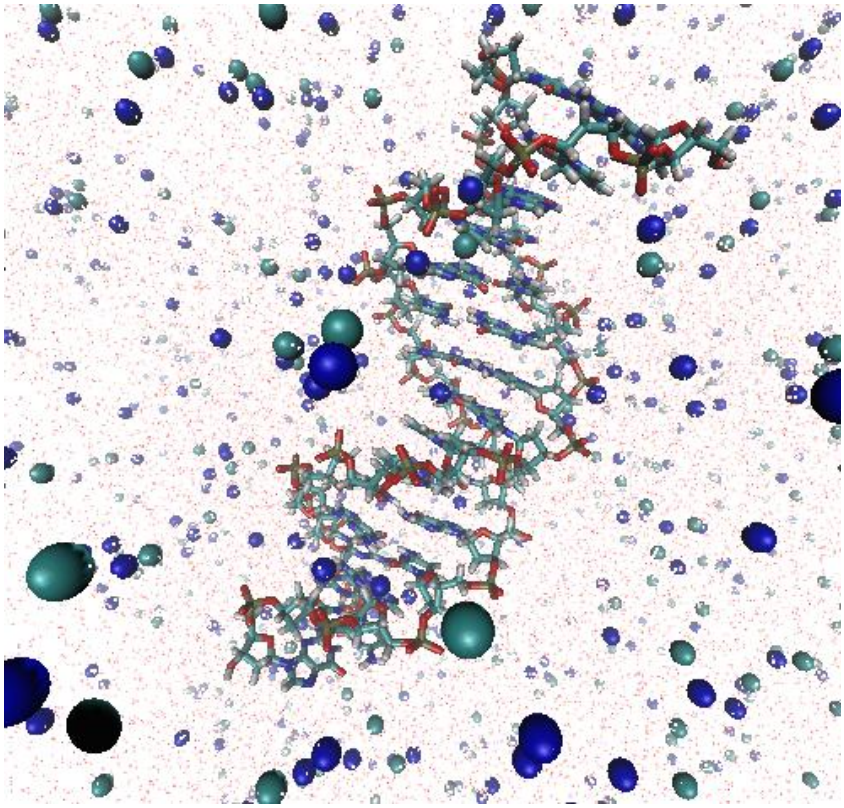


Figure 3.2.5 Modeling system with 15 bp DNA (alternating GC sequence) in 1 M Na^+. This image is generated by VMD program http://www.ks.uiuc.edu/Research/vmd/. The red dots are H_2O molecules, the blue beads are Na^+ and the cyan beads are Cl^-.

A 15 bp alternating GC sequence (gcgcgcgcgcgcgcg) was put in a water box with a dimension of $4\ \text{nm} \times 4\ \text{nm} \times 14\ \text{nm}$. 135 Na^+ and 107 Cl^- were added into the system to achieve 1 M Na^+ concentration as shown in Figure 3.2.5. The uneven numbers of Na^+ and Cl^- resulted from the fact that non-terminal phosphate group is negatively charged defined in the program; therefore the negative charges from 28 non-terminal phosphate groups were neutralized by excess Na^+. The system temperature was set to be 17 ˚C to stabilize DNA. Such choice of simulation parameters (alternating GC sequence with high salt concentration and low temperature) was meant to favor the formation of S-DNA.

Umbrella sampling was utilized to achieve the B-to-S transition, which overcomes a huge energy barrier between different DNA states. This could be done by imposing force with a spring, whose two ends were attached to the first and the last base pair of this DNA segment. By adjusting the original length of the spring, we can either stretch (original length longer than DNA) or compress the DNA (original length shorter than DNA).

In our system, two ends of the spring were fixed at the center of the mass of the first and last base pair respectively, along the central axis of DNA. In our case, in order to achieve a quasi-equilibrium structural transition to S-DNA, we gradually increased the length of the spring, and at each length, DNA was allowed to relax long enough time until the extension did not change. Furthermore, to prohibit B-to-ss peeling transition occurring from the two ends, the hydrogen bonds of the first and last base pair were restrained together. The spring constant used in our systems is 100 kJ/mol/nm^2. The restraint DNA is shown in Figure 3.2.5.

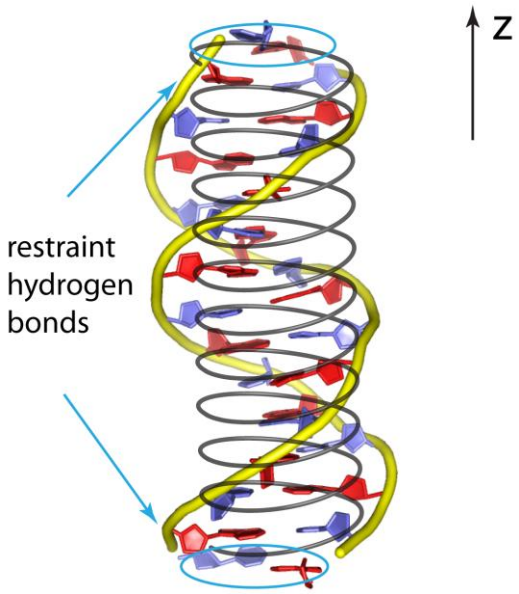


Figure 3.2.6 Cartoon illustration of restraint DNA construct. Solvent molecules are not shown.

3.3 Results and Discussion

3.3.1 Phase diagram

As discussed in section 3.2.1, the free energies of these four DNA structures (B-DNA, S-DNA, 1ssDNA and 2ssDNA) are shown to be dependent on force, temperature and salt concentration. The phase diagram can be plotted by directly comparing the free energies of different DNA forms; therefore we can examine conditions at which the DNA with the lowest free energy can occur at equilibrium.

To simplify the case, here we don't consider the sequence dependence of the base pair stacking energy of B-DNA (g_B) and S-DNA (g_S); rather we choose an average value of entropy and enthalpy value of g_B provided in Table 1.2.1, equivalent to 50% of the GC content in DNA sequence (g_B=1.43 kcal/mol in 1 M NaCl 37 ˚C).

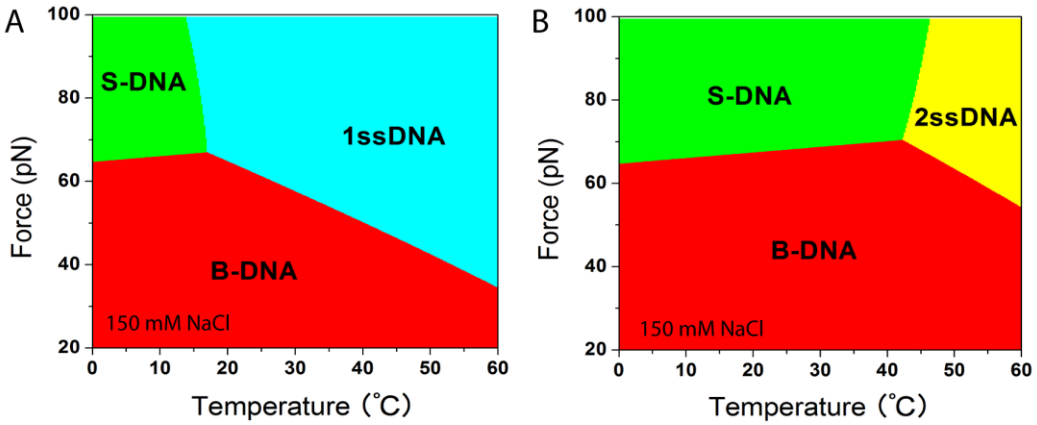


Figure 3.3.1 Force-temperature phase diagram at 150 mM NaCl for a 50% GC content DNA. The boundaries between different DNA structures represent the critical conditions where the DNA structural transition can happen. (A) F-T phase diagram for DNA with open ends, where 2ssDNA structure is not present. (B) F-T phase diagram for topologically end-closed DNA, where the formation of 1ssDNA is prohibited.

Figure 3.3.1A shows the DNA force-temperature phase diagram in 150 mM NaCl for a DNA with open ends, where it can exist as any DNA structure (B-DNA, S-DNA, 1ssDNA or 2ssDNA). As can be seen, under low force, the B-

DNA is always the most stable structure compared with other structures. S-DNA is prominent in low temperature condition while 1ssDNA is more favored at high temperature condition. The boundaries indicate occurring of the DNA structural transitions, where the free energies of different DNA forms are equal. However, 2ssDNA is not present in any of the conditions. This is reasonable since the energy difference between 1ssDNA and 2ssDNA under tension $\Delta G = G_{ss}(f) - G_{2ss}(f)$ is always negative as shown in Figure 3.3.2, which means that 1ssDNA under tension is always more favored compared to 2ssDNA under tension. Therefore, the formation of 2ssDNA structure is repressed as long as the possible existence of 1ssDNA. Hence, the phase diagram for a topologically end-closed DNA is drawn (Fig. 3.3.1B), where the formation of 1ssDNA is prohibited as there are no open ends and DNA can only choose three possible structures (B-DNA, S-DNA and 2ssDNA). Then the 2ssDNA structure is found to exist at much higher temperatures.

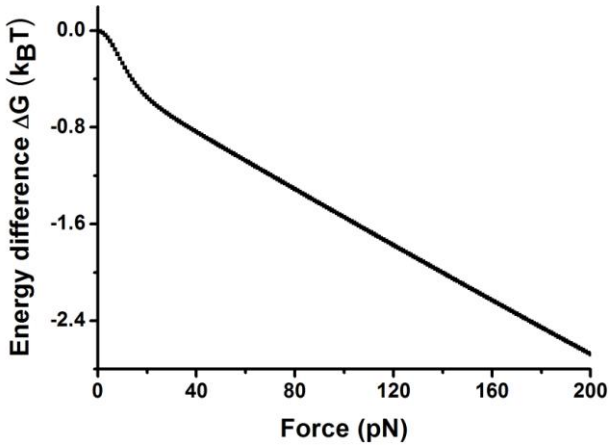


Figure 3.3.2 Energy difference per base pair between 1ssDNA under tension and 2ssDNA under tension.

Similarly, the force-salt phase diagram can be plotted at a constant temperature as shown in Figure 3.3.3. The B-DNA structure is always most stable under low forces at all salt concentrations. At 24 ˚C, 2ssDNA structure is still absent in all conditions for DNA with open ends (Fig. 3.3.3A). In this case, the 1ssDNA structure is more favored under low salt concentrations while S-DNA is more stable at high salt concentrations. For topologically end-closed DNA,

2ssDNA structure is only favored at extremely low salt concentration at 24 ˚C (Fig. 3.3.3B). In the case where temperature is increased to 50 ˚C as shown in Figure 3.3.3C, 2ssDNA then can exist in more physiological salt concentrations.

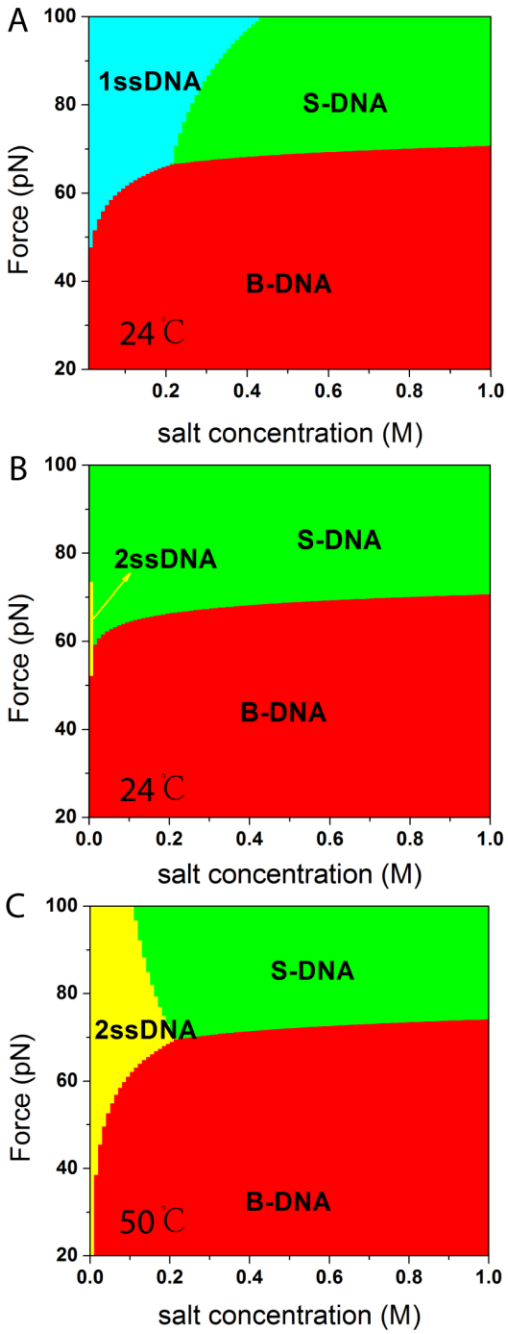


Figure 3.3.3 Force-salt phase diagram at 24 ˚C (A,B) or 50 ˚C (C) for a 50% GC content DNA. The boundaries represent the critical conditions where the DNA structural transition can happen. (A) F-salt phase diagram at 24 ˚C for DNA with open ends; (B) F-salt phase diagram at 24 ˚C for topologically end-closed DNA; (C) F-salt phase diagram at 50 ˚C for topologically end-closed DNA.

The F-T phase diagram and F-salt phase diagram above elucidate the DNA structure distribution over different environmental conditions at equilibrium. In general, B-DNA is the most stable structure under low forces at all temperature and salt concentrations. S-DNA is more favored under high forces at low temperature or high salt concentration, where the base pair is more stable; while 1ssDNA appears under high forces at high temperature or low salt concentrations. Only when the formation of 1ssDNA is forbidden, 2ssDNA can exist under extreme conditions such as very high temperature or extremely low salt concentrations.

3.3.2 Transfer matrix calculation of the stability of B-DNA, 2ssDNA and S-DNA

For a real transition that happens in a long DNA polymer, we have to consider the DNA sequence dependence, which can be solved by transfer matrix due to the nearest-neighbor base pair stacking of the DNA molecule. Therefore, in this section, the transition that occurs in a topologically end-closed DNA has been studied using transfer matrix method as described in section 3.2.2. Here we don't consider peeling transition for DNA with open ends, as its occurrence depends on the location of the open boundary, which is a kinetic process.

A 7535 bp topologically end-closed DNA with 37% GC content was chosen for this study, which can either be the form of B-DNA, S-DNA or 2ssDNA under this DNA construct. The salt-dependence structural coefficient v of the DNA is measured by single molecule experiments; for B-to-S transition v is measured to be ~ 0.48 and for B-to-2ss transition v is ~ 1.15 as indicated in Table 1.2.4. One thing to be mentioned is that the extension per nucleotide for ssDNA used here is not adopted from Cocco *et al.* (45), since that phenomenological formula will not be applicable to low salt below 2 mM NaCl. Instead, we used the experimental

measured value obtained by measuring the force-extension curve of 1ssDNA peeled from λ-DNA in different salt concentration as shown in Figure 3.3.4.

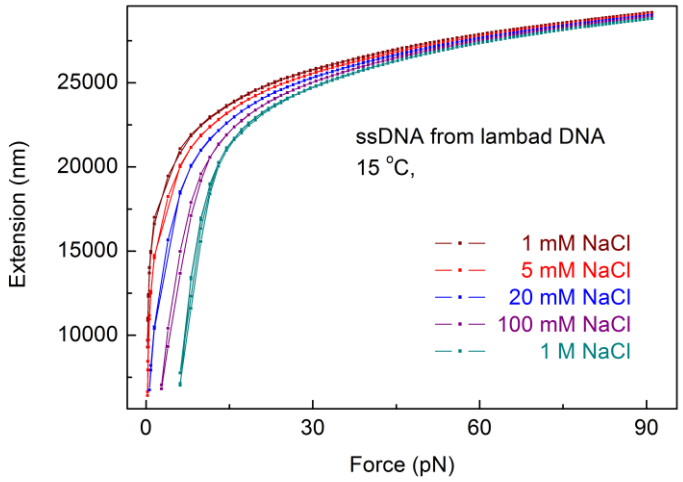


Figure 3.3.4 The force-extension curve of ssDNA peeled from λ-DNA in different salt concentration. This experiment was done by my labmate Dr. Zhang Xinghua.

Firstly, we calculated the fractions of different DNA structures under different forces either in 1 mM or 100 mM NaCl at 23 ˚C as shown in Figure 3.3.5.

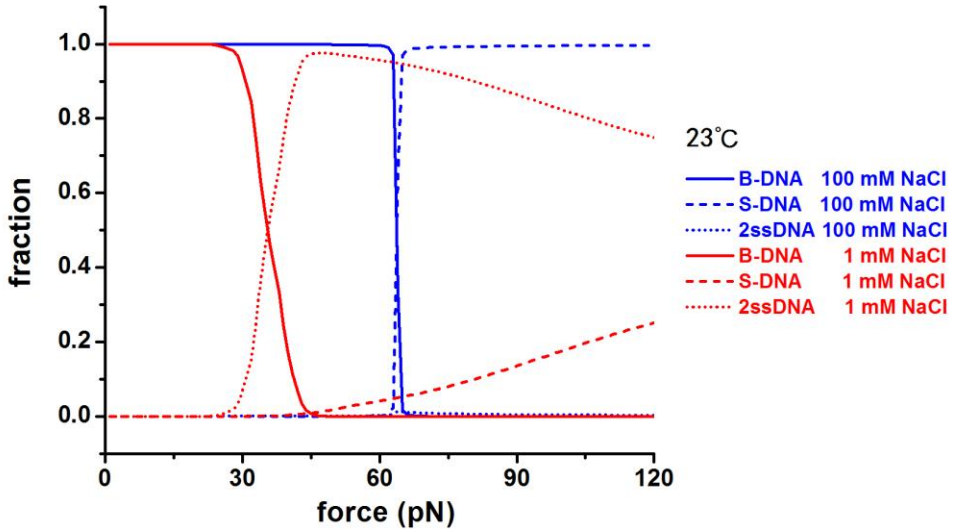


Figure 3.3.5 B-DNA, S-DNA and ssDNA fractions under different forces in 1 or 100 mM NaCl at 23 ˚C.

The B-DNA is overall dominant at force below 30 pN in both salt concentrations. In 100 mM NaCl, B-to-S transition happened at around 65 pN, after which the DNA is basically pure S-DNA as indicated by the blue dashed line. The fraction of 2ssDNA is always very small in this salt concentration as indicated by the blue dotted line. However, in 1 mM NaCl, the internal melting transition is predicted around 40 pN. After the transition, 2ssDNA fraction becomes dominant and the fractions of the other two structures are repressed.

According to Figure 3.3.5, we anticipate that if holding the DNA at high force above 65 pN, the transition between S-DNA and 2ssDNA could be observed if the buffer is switched between the high salt and low salt conditions. One case is shown in Figure 3.3.6, where holding the DNA under 82 pN, changing the buffer from 100 mM to 1 mM NaCl will result in S-to-2ss transition; while reversibly changing the buffer will initiate the transition from 2ssDNA to S-DNA.

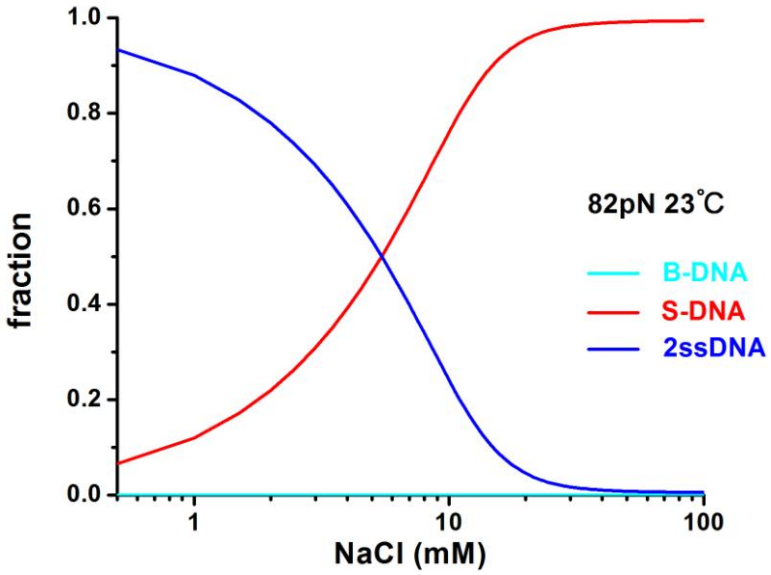


Figure 3.3.6 B-DNA, S-DNA and ssDNA fractions under 82 pN in different salt concentration ranging from 0.5 to 100 mM NaCl at 23 ˚C.

The corresponding single molecule experiment was also carried out as shown in Figure 3.3.7. In 100 mM NaCl, the DNA was basically pure S-DNA indicated by the overlapping of the theoretical and experimental S-DNA force-extension

curve. Changing the buffer to 1 mM NaCl, the DNA became 2ssDNA indicated by the reduction of the extension; reversibly, 2ssDNA could re-transit to S-DNA if changing the buffer back to 100 mM NaCl. This is in perfect agreement with theoretical prediction.

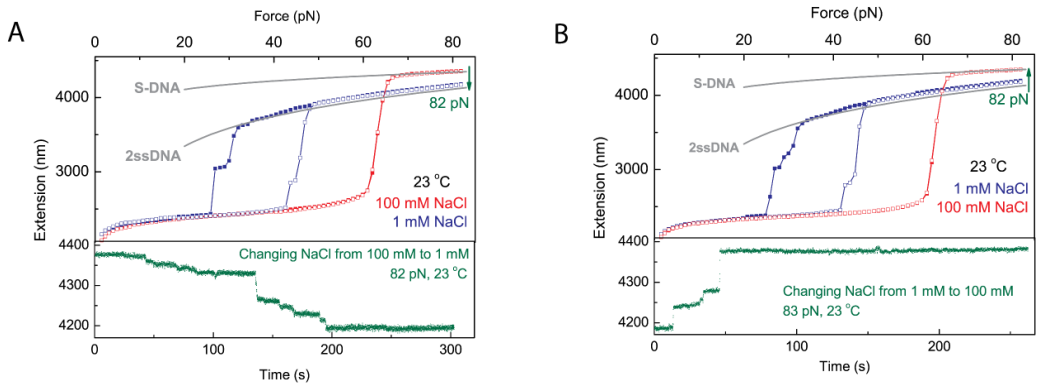


Figure 3.3.7 Transition between S-DNA and two parallel ssDNA under 82 pN at 23 ˚C. This experiment was done by my labmate Dr. Zhang Xinghua.

The salt-driven transition between the S-DNA and 2ssDNA found in this section indicates that S-DNA is not a transient form, but a stable form under physiological conditions under large tension where 2ssDNA will transit to S-DNA.

3.3.3 **Kinetics of DNA overstretching transitions**

DNA overstretching is now known involving three distinct structural transitions: 1) the strand-peeling transition resulting in 1ssDNA structure; 2) the internal melting transition resulting in two parallel ssDNA; 3) the B-to-S transition resulting in the base-paired S-DNA. The previous discussions in the last two sections are purely based on equilibrium statistic, which is solely dependent on the respective stabilities of different DNA structures. However, the kinetics of the respective transitions is also important. Besides being related to the overstretched DNA structures, knowledge of the kinetics of the transition is also

crucial for interpretation of experimental results. For example, under the same solution condition, a B-to-S transition can be observed in quick pulling, while strand-peeling can be observed in slower pulling (11). Such seemingly contradicting results have brought great confusion in previous experiments and also caused a debate on the stability of respective overstretched DNA structures (61). In this section, I systematically investigated the kinetics of all the DNA overstretching transitions and transitions between overstretched DNA structures using the kinetic Monte Carlo simulation introduced in section 2.4.

Previous single-molecule experiments of overstretching long DNA have shown that the peeling transition has a slower kinetics resulting in hysteresis between overstretching and relaxation (10), which was hypothesized to be resulted from fork movement through the sequence dependent energy barriers. To examine whether this hypothesis can fully describe the kinetics observed in strand-peeling, our lab have stretched short DNA tethers (~ a few hundred bp) and monitored the DNA extension changes at constant forces with high spatial resolution of ~ 2nm. In 150 mM NaCl at 24 ˚C, stepwise extension changes were observed at each force during overstretching transition as shown in Figure 3.3.8A (11), which looks similar to what have been observed in DNA unzipping experiments that occurred at lower forces (15-20 pN) (143). After the transition completed, the tether became an ssDNA because the peeled strand drifted away leaving one ssDNA strand under tension (Fig 3.3.8C).

To see whether this stepwise dynamics is indeed caused by fork movement across sequence dependent energy barriers, we performed a KMC simulation for the same sequence of DNA assuming fork diffusion in the DNA. The DNA used in experiments had two open ends with 576 bp; therefore two forks were involved in the simulation, as described in section 3.2.3. In the simulation, the 576 bp DNA was indexed from 1 to 576 along one DNA strand from its 5’ to the 3’ end.

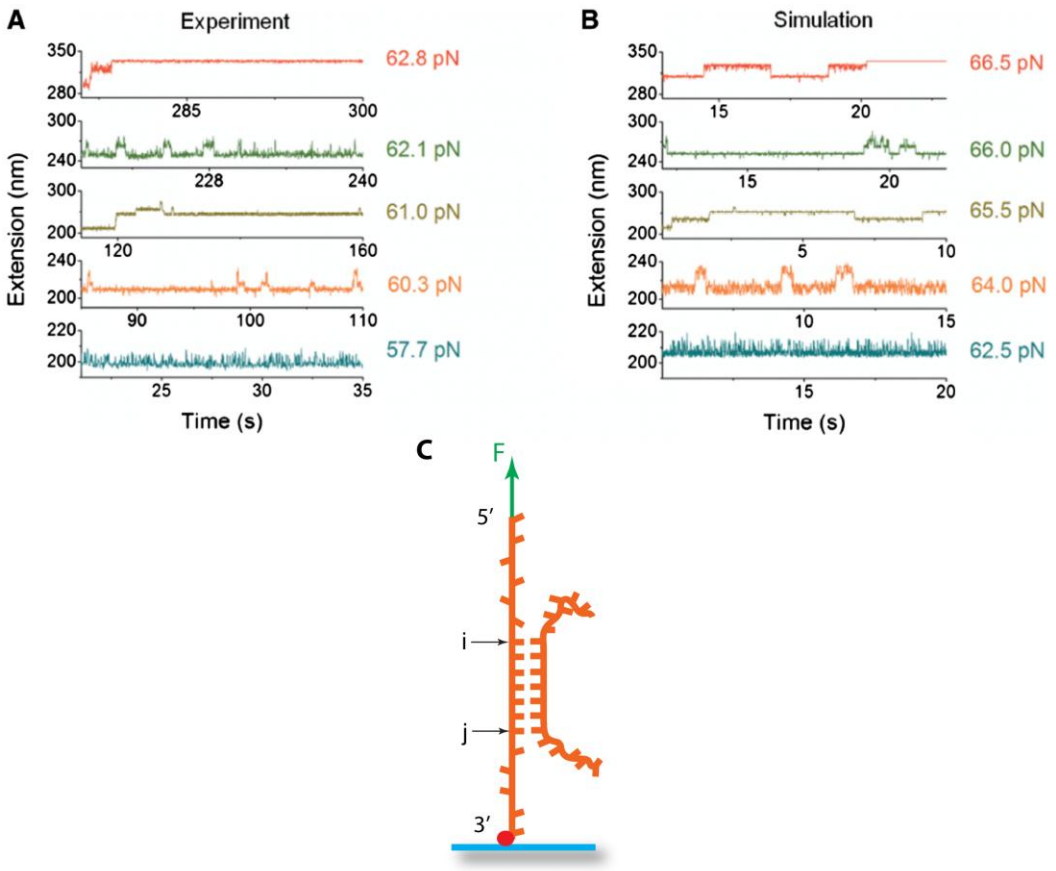


Figure 3.3.8 Detailed dynamics of 576 bp DNA at different forces during overstretching transition in 150 mM NaCl at 24 ˚C. (A) Stepwise dynamics observed in the single-molecule experiment, which was done by my labmate Dr. Fu Hongxia. (B) Stepwise dynamics predicted in B-to-ss peeling kinetics simulation. (C) KMC model for peeling transition.

To provide some insights on the sequence dependent energy landscape during fork movement, we plotted a one dimensional free energy landscape of peeling from the 5' end to the 3' end (indices 1 to 576; blue, Fig. 3.3.9) and peeling from the 3' to the 5' end (indices 576 to 1; red, Fig. 3.3.9). At force > 65.5 pN, both free energy decrease as the 5' fork index *i* increases and the 3' fork *j* decrease, indicating peeling can proceed in both directions. In contrast, opposite trends was observed at force <65.5 pN, indicating B-DNA is stable. Importantly, at any force, the free energy landscapes are rugged, containing many local energy minimums and energy maximums separated by various levels of energy barriers. Such rugged free energy landscape is expected to cause stepwise fork movements.

To see whether the fork movements along these rugged free energy landscapes can explain the experimentally observed stepwise extension changes, we perform KMC simulation. Since the DNA has two forks, its state is defined by the pair of indices (*i*, *j*) that indicate the locations of the 5' fork and the 3' fork respectively. The free energy associated with a state under force f is:

$$\Delta G_{B-ss}(i,j,f) = \Delta G^{5'}_{B-ss}(i,f) + \Delta G^{3'}_{B-ss}(j,f)$$

as introduced in section 3.2.3. The four possible state transitions from the current state are (*i*, *j*+1), (*i*, *j*-1), (*i*+1, *j*), (*i*-1, *j*), which are stochastically selected based on the reaction rate using the Gillespie's algorithm (142) with an attempting rate of $r_{B-ss} = 5 \times 10^8 s^{-1}$. At each state (*i*, *j*), the extension of DNA can be calculated based on known force extension curves of dsDNA and ssDNA.

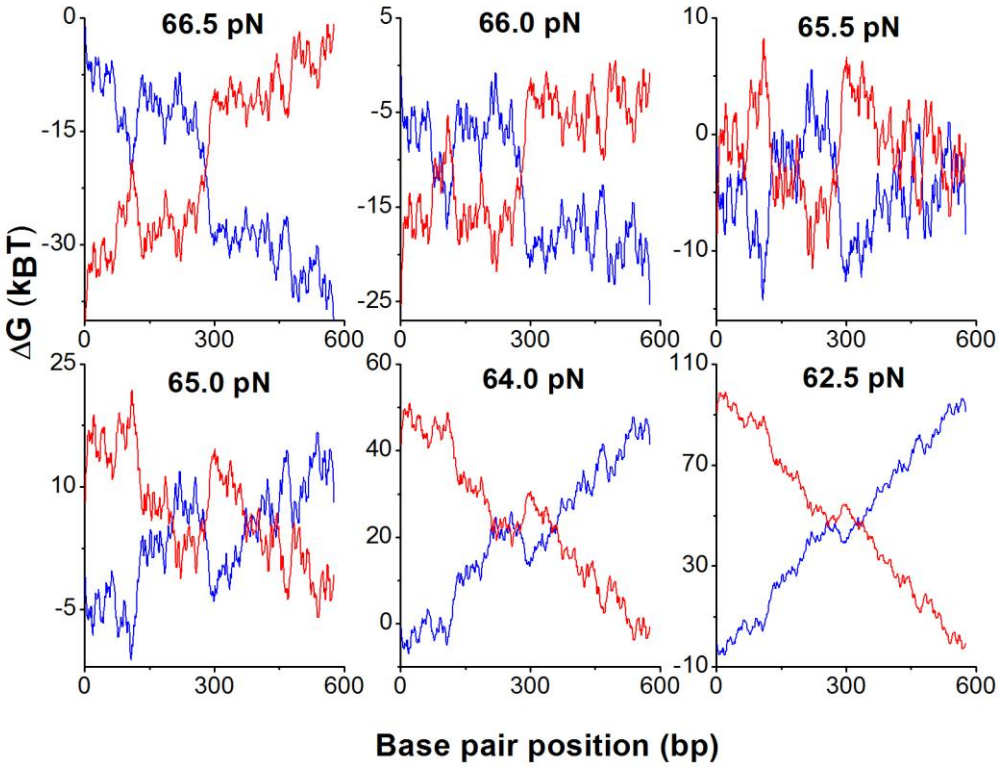


Figure 3.3.9 Free energy landscape of the 576 bp DNA in 150 mM NaCl at 24 ˚C. The blue curve was calculated by $\Delta G_{tot}(n,f) = \sum_{i=1}^{n} \Delta G(i,f)$, which represents the energy required to open successive base pairs from the 5' end. The red curve was calculated by $\Delta G_{tot}(n,f) = \sum_{i=N-n}^{N} \Delta G(i,f)$, which represents the energy required to open successive base pairs from the 3' end.

Figure 3.3.8B shows the simulated time traces at forces corresponding to those in Figure 3.3.8A, which are qualitatively similar to each other. This agreement between experiment and simulation supports the proposed hypothesis that the stepwise dynamics was due to overcoming the sequence-dependent energy barriers.

In 1 M NaCl at 24 ˚C, qualitatively different dynamics was observed between experiments (Fig. 3.3.10A) and simulation (Fig. 3.3.10B). In the experiments, the stepwise dynamics completely disappeared but a fast random fluctuation was observed, indicating that fork movements across sequence-dependent energy barriers are not involved in the transition. This is further supported by the observation that the overstretched DNA was able to go back to B-DNA when force was lowered, indicating that this transition cannot be any DNA melting transition. All together, these results implied that the overstretching transition happened in 1 M NaCl is a B-to-S transition, which has a much faster kinetics than strand-peeling transition.

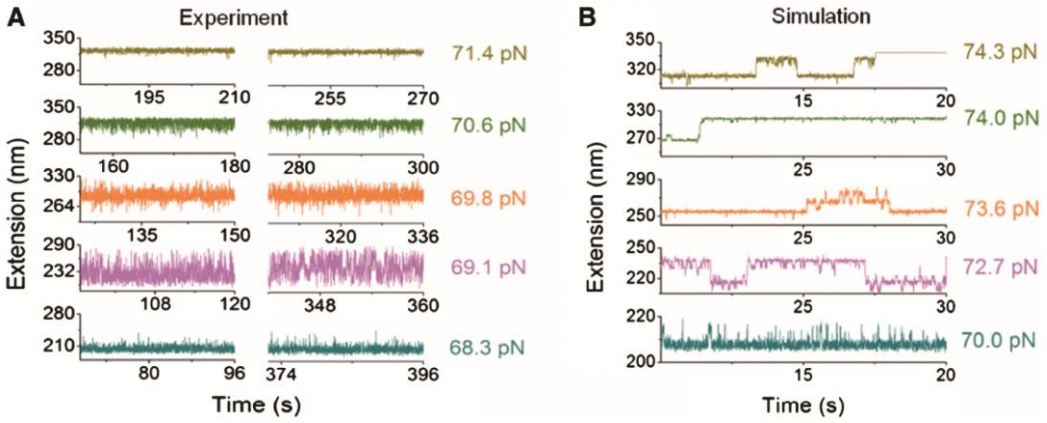


Figure 3.3.10 Detailed dynamics of 576 bp DNA at different forces during overstretching in 1 M NaCl at 24 ˚C. (A) Random fluctuation observed in the single-molecule experiment, which was done by my labmate Dr. Fu Hongxia. (B) Stepwise dynamics predicted in B-to-ss peeling kinetics simulation.

Following, to confirm the transition happened in 1 M NaCl at 24 ˚C is a B-to-S transition, we carried out the KMC simulation for the B-to-S transition using our B-to-S transition algorithm discussed in section 3.2.3 for the same DNA

sequence. As the B-to-S transition does not seem to involve fork movement, in the simulation we treat the state of DNA as an array of 576 indices $\{n_i\}$ ($i = 1,2 \ldots,576$) describing whether the i th base pair is in B-form ($n_i = 0$) or S-form ($n_i = 1$). At each evolution step, there are 576 possible state transitions, which are stochastically selected in the simulation.

The B-to-S transition has a very fast rate (quicker than our sampling rate of ~100 Hz in the experiment). In the simulation, the attempting rate $r_{B-S} = 5 \times 10^8\ s^{-1}$ was adopted, which is the same in the B-to-ss peeling transition. Besides the kinetics, the B-to-S transition occurred within a very narrow force range of around 3 pN (began at 68.3 pN and completed at 71.4 pN), indicating a high cooperativity of the transition. The cooperativity is controlled in simulation by a positive energy cost to create a boundary between B-DNA and S-DNA, which was treated as a free parameter tuned to fit the experiment. With a choice of a B-S boundary energy of $C_{B-S} = 7\mathrm{k_BT}$, the results are shown in Figure 3.3.11, which qualitatively resembles the experimental observations (Fig. 3.3.10A).

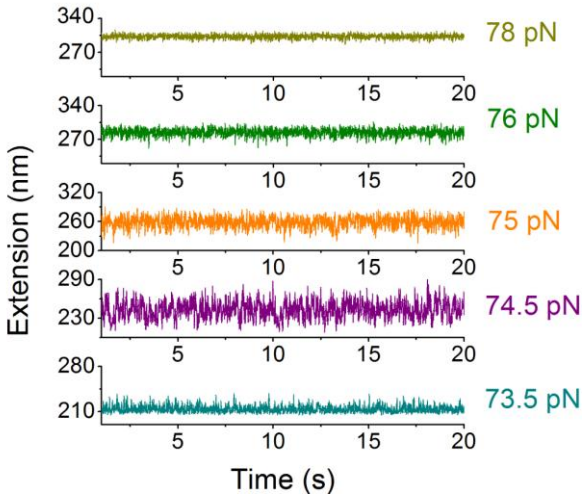


Figure 3.3.11 KMC simulation prediction of the dynamics of B-to-S transition of 576 bp DNA a in 1 M NaCl at 24 ˚C.

The boundary energy of $7\mathrm{k_BT}$ is needed to ensure that the transition occurs over a narrow force range similar to experimental observation. Although the

boundary energy usually originates from the stacking energy, this huge boundary energy between B-S boundary is much larger than the B-DNA stacking energy (~3 k_BT), whose nature is still unclear for now. Further, it is somewhat difficult to directly measure the boundary energy in experiment or speculate it by molecular dynamics simulation. Therefore, more in-depth study is needed but beyond the scope of this thesis.

As our KMC simulation successfully reproduced B-to-ss peeling transition and B-to-S transition, we then take a step further to see whether we can predict the S-to-ss peeling transition by KMC simulation. Furthermore, after the S-to-ss peeling transition, whether the peeled strand could re-anneal back to S-DNA was studied.

An 8015 bp DNA with 58 % GC content was designed for the experiment and simulation. This 8051 bp DNA has two open ends and a super high GC content sequence (~40 bp 100% GC content) located at the 3' end of one strand of the DNA. This super high GC content sequence is meant to favor the S-DNA formation in the low salt concentration, so that two strands could still hold together even in low salt condition, therefore ss-to-S re-annealing can be studied. The DNA is indexed from 1 to 8015 from the 5' end to the 3' end along one strand of DNA (low GC content end to the high GC content end). Similar to the B-to-ss peeling transition, there are also two forks involved in the simulation and the DNA state can be described by a pair of indices (*i, j*) indicating the location of the 5' end fork and the 3' end fork.

Compared to the 5'end fork, the 3' end fork is much harder to move in the presence of super high GC content sequence. Therefore, for simplicity we only plotted a one dimensional free energy landscape of peeling from 5' end to the 3' end (indices 1 to 8015) under high force (97 pN or 74 pN) at 13 ˚C in 200 mM (red, Fig 3.3.12) or 4 mM NaCl (blue, Fig 3.3.12). In 200 mM NaCl (red, Fig 3.3.12), S-DNA is quite stable indicating that huge energy barriers are on the way to open S-DNA base pairs; while in 4 mM NaCl, 1ssDNA is more favorable as indicated by the downward trend of the energy curves. There is an energy

minimum located at the 3' end (insert panel, Fig. 3.3.12), which is due to the super high GC content sequence. This energy minimum will favor the formation of S-DNA in 4 mM NaCl.

According to the energy landscape, we proposed that S-DNA could form at high salt concentration under 97 pN at 13 ˚C; then the preformed S-DNA at high salt concentration would peel to 1ssDNA if switching to low salt buffer; after that if switching back to high salt concentration again, the peeled 1ssDNA could re-anneal to S-DNA.

We then performed the KMC simulation to test our prediction. Assuming S-DNA formed at 200 mM NaCl, stepwise extension elongation was predicted under 97 pN in 4 mM NaCl in the simulation (Fig. 3.3.13A), with an attempting rate of $r_{S-ss} = 5 \times 10^8 s^{-1}$, in agreement with experimental observation under the same condition (Fig. 3.3.13B).

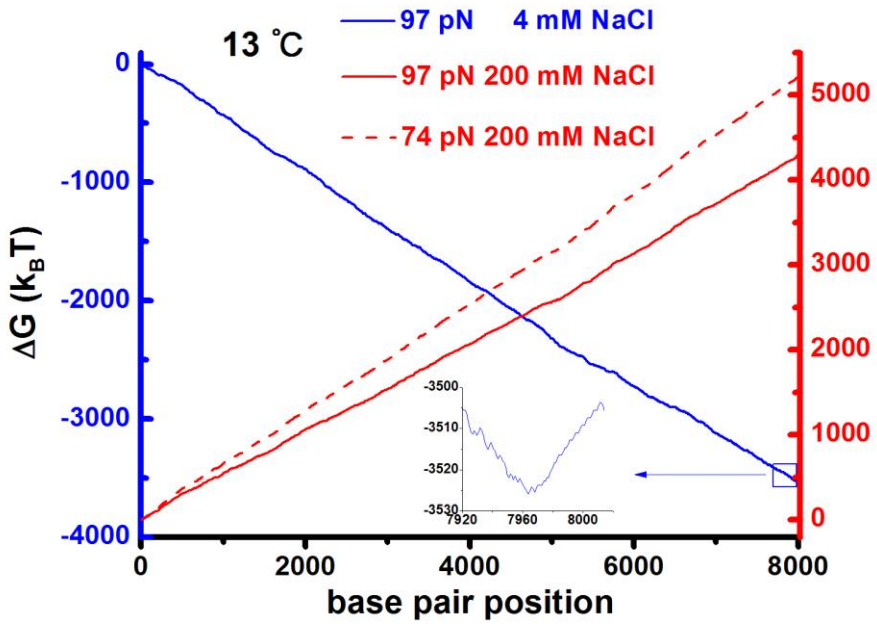


Figure 3.3.12 Free energy landscape of the 8015 bp DNA under high force at 13 ˚C in different salt concentrations. We only show the free energy to open successive base pair from the 5' end to the 3' end (from 1 to 8015 bp).

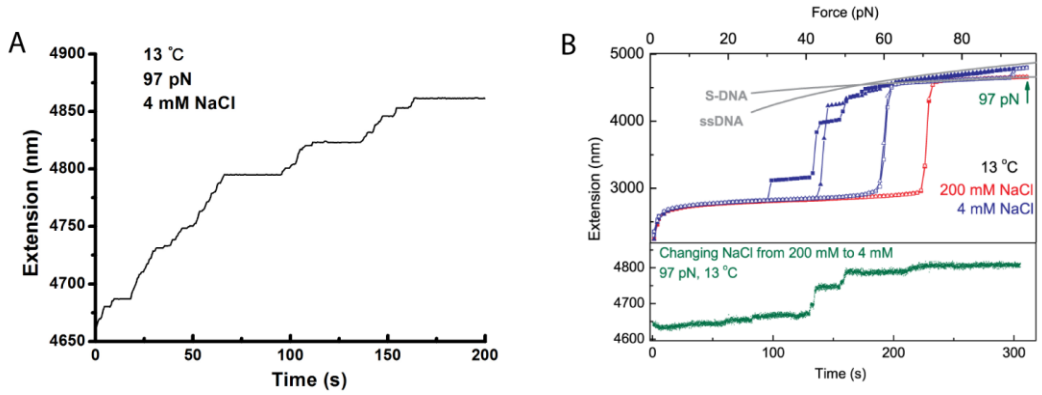


Figure 3.3.13 Simulation results (A) and experimental observation (B) of S-to-ss transition occurred under 97 pN in 4 mM NaCl at 13 ˚C. The experiment shown in (B) was done by my labmate Dr. Zhang Xinghua.

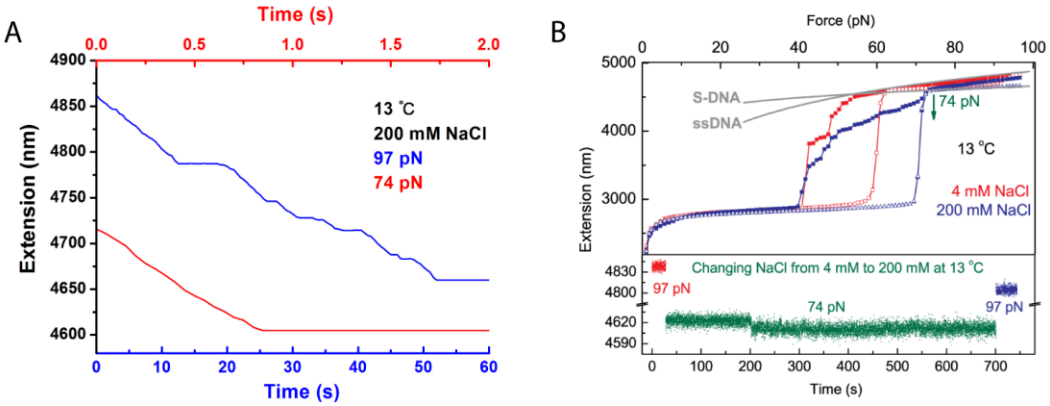


Figure 3.3.14 Simulation results (A) and experimental observation (B) of ss-to-S re-annealing occurred under 97 pN or 74 pN in 200 mM NaCl at 13 ˚C. The experiment shown in (B) was done by my labmate Dr. Zhang Xinghua.

Reversibly, assuming peeled ssDNA created in 4 mM NaCl, stepwise extension reduction was predicted under 97 pN in 200 mM NaCl due to ss-to-S re-annealing (blue, Fig. 3.3.14A). A much faster re-annealing process was predicted under 74 pN in the simulation (red, Fig. 3.3.14A). In contrast, only a very small portion of the DNA is re-annealed to S-DNA even under force 74 pN in the experiment (Fig 3.3.14B). This disagreement between simulation and experiment is understandable. Because in our model, we don't consider the case that the partially peeled ssDNA could form secondary structures, which might prevent the 1ssDNA re-annealing transition back to S-DNA. The formation of the secondary structure by the fallen strand would lower the free energy of 1ssDNA

under tension that tuned the free energy landscape of this transition, resulting in the failure of KMC prediction.

Next, we consider whether we can use the model to predict the internal melting occurred in a B-DNA or S-DNA for an end-closed DNA. The kinetics of these two transitions has not been fully characterized in experiments. Preliminary results from my lab mate have shown that both transitions involved highly cooperative, complicated kinetics with stepwise extension changes. Due to the lack of experimental data for direct comparison, in this part I focused on understanding how reaction rates and boundary energy cost may affect the transition kinetics.

The DNA sequence used here is an end-closed DNA with 7535 bp, which is the same DNA used in section 3.3.2 for the transfer matrix calculation. The DNA state is described by an array of indices $\{n_i\}$ ($i = 1,2, \dots, 7535$), denoting whether the ith base pair is in closed state ($n_i = 0$) or in open state ($n_i = 1$), similar to the simulation for the B-to-S transition. The boundary energy of B-2ss was chosen to be $C_{B-2ss} = 4\,\mathrm{k_B T}$, which is close to the stacking energy of B-DNA. The boundary energy involved in S-DNA internal melting C_{S-2ss} is treated as a free parameter, which was varied from $4 - 7\,\mathrm{k_B T}$ to study its effect.

According to transfer matrix calculation in section 3.3.2, internal melting transition occurred in 1 mM NaCl above 45 pN force and the inter-conversion between S-DNA and 2ssDNA can be achieved by changing the salt concentration between 1 mM and 100 mM NaCl. With a choice of attempting rate $r_{B-2ss} = 5 \times 10^3 s^{-1}$, the typical kinetics of inter-conversion between B-DNA and 2ssDNA was predicted in the simulation (Fig. 3.3.15A) at 60 pN, displaying an exponential extension increase or decrease in 1 mM (red, Fig. 3.3.15A) or 100 mM NaCl (blue, Fig. 3.3.15A) respectively. Similar dynamics were obtained for the inter-conversion between S-DNA and 2ssDNA with attempting rate $r_{S-2ss} = 5 \times 10^5\ s^{-1}$ and boundary energy $C_{S-2ss} = 5\,\mathrm{k_B T}$ as shown in Figure 3.3.15B.

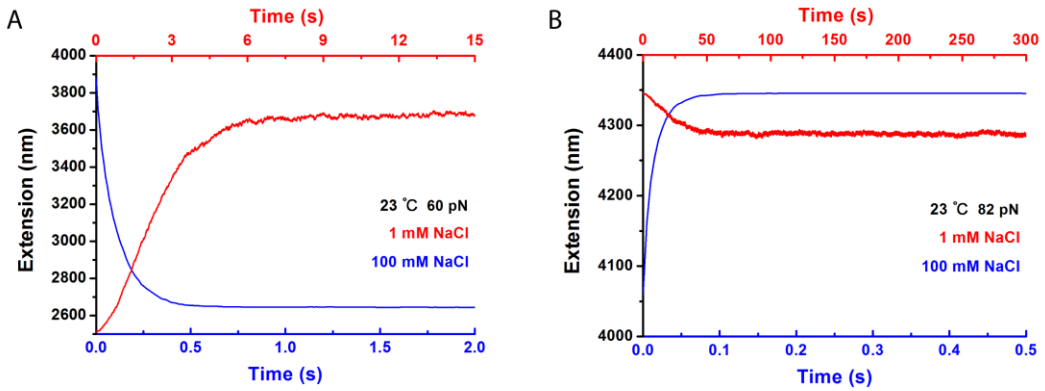


Figure 3.3.15 KMC simulation for internal melting transition occurred in B-DNA (A) and S-DNA (B). (A) $C_{B-2ss} = 4\,k_BT$ and $r_{B-2ss} = 5 \times 10^3 s^{-1}$; (B) $C_{S-2ss} = 5\,k_BT$ and $r_{S-2ss} = 5 \times 10^5 s^{-1}$.

However, the exponentially extension changes predicted in the simulation has never been observed in our preliminary experiment, which on the contrary showed highly cooperative step jump in B-to-2ss transition (preliminary data not shown) and stepwise extension changes in S-to-2ss transition (Fig. 3.3.7). More importantly, the S-to-2ss transition in 1 mM NaCl did not reach equilibrium state as predicted by transfer matrix calculation (Fig. 3.3.6), where DNA is a mixed state with ~ 90 % internal bubbles and ~ 10 % S-DNA. Further, a set of simulations were carried out using different attempting rates and boundary energies, obtaining similar results, failing to reach equilibrium state, indicating that the DNA might be trapped in a meta-stable state (Fig. 3.3.16).

The disagreement between simulation and experiment in the internal melting transition implies that our oversimplified model used in the simulation may not be applicable to internal melting transition. Although lacking sufficient experiment data for the internal melting transition, the stepwise extension changes observed in our preliminary experiment (Fig. 3.3.7) indicate that the internal melting transition may involved with some dominate forks movement similar to the peeling transition. A more realistic simulation by assigning a slower attempting rate for the initial formation of the bubbles might boost this transition and may lead DNA to reach equilibrium state.

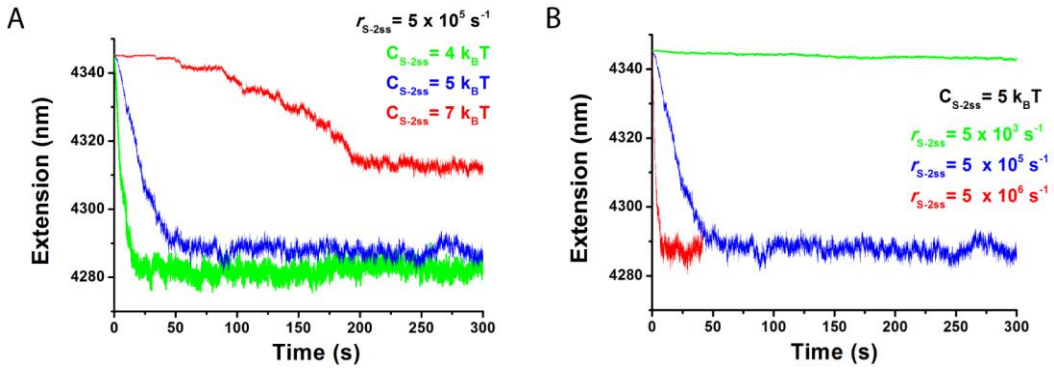


Figure 3.3.16 The effects of different attempting rates and boundary energies for internal melting transition occurred in S-DNA. (A) Dynamics of S-to-2ss transition with boundary energy set to be 4 k_BT (green), 5 k_BT (blue) and 7 k_BT (red) respectively, with $r_{S-2ss} = 5 \times 10^5\ s^{-1}$; (B) Dynamics of S-to-2ss transition with attempting rate set to be $5 \times 10^3\ s^{-1}$ (green), $5 \times 10^5\ s^{-1}$ (blue) and $5 \times 10^6\ s^{-1}$ (red) respectively, with $C_{S-2ss} = 5\ k_B T$.

A simple sketch summarizing the results of the structural transition of DNA is shown in Figure 3.3.17. As can be seen, the overstretching transition can lead DNA from B-DNA to 1ssDNA, 2ssDNA or S-DNA, sensitive to environmental conditions, DNA sequences or DNA topologies. The selection of overstretching transition pathway is determined by the relative stabilities of the overstretched DNA structures. When factor changes tend to increase the DNA base pair stability (such as lower the temperature or increase the salt concentration), it will favor the B-to-S transition over the other two. The selection of the pathway between peeling transition and the internal melting transition highly depends on the DNA topologies, where internal melting will always be suppressed as long as there are open ends or nicks in DNA without extremely heterogenic sequence distribution. The stabilities of respective overstretched DNA structures can also determine the inter-conversion between each other. The inter-conversions at different salt concentration between the S-DNA and 2ssDNA were observed in the experiment as predicted by theoretical calculation. However, the re-annealing from 1ssDNA to S-DNA was hardly achieved in the single-molecule experiment, which might result from the secondary structures formed by the partially peeled ssDNA that was not considered in the simulation.

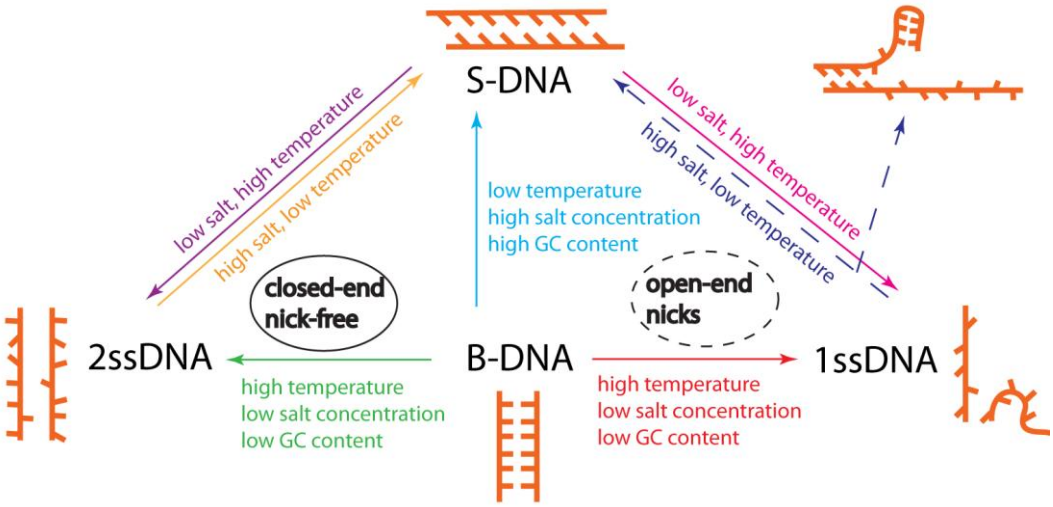


Figure 3.3.17 Sketch summary of the pathway of overstretching transition and the inter-conversion between the overstretched under appropriate environmental conditions.

3.3.4 **Insight for the structure of S-DNA**

In this section, we probed the detailed structure of DNA by steered MD simulation. A 15 bp alternating GC sequence was used as the DNA template. The spring was attached between the first and the last base pair. The original length of the spring was then successively increased by 0.1 times the DNA contour length each step starting from the length equal to DNA contour length. We performed at least 20 ns for each step until the DNA extension did not change and the DNA structure did not drastically fluctuate. The total simulation time was 550 ns and the final length of the spring was 2 times of the DNA contour length.

The extension time course and the force-extension curve were obtained and shown in Figure 3.3.18. The DNA gradually elongated to 1.7 times of the contour length as the increase of the original length of the spring with time as shown in Figure 3.3.18A. Figure 3.3.18B shows that force reached ~ 90 pN when DNA became 1.7 times of the contour length and drastically increased to 180 pN if further stretched. It is comparable though not identical with the experimental measured ~ 65 pN, where overstretching occurs.

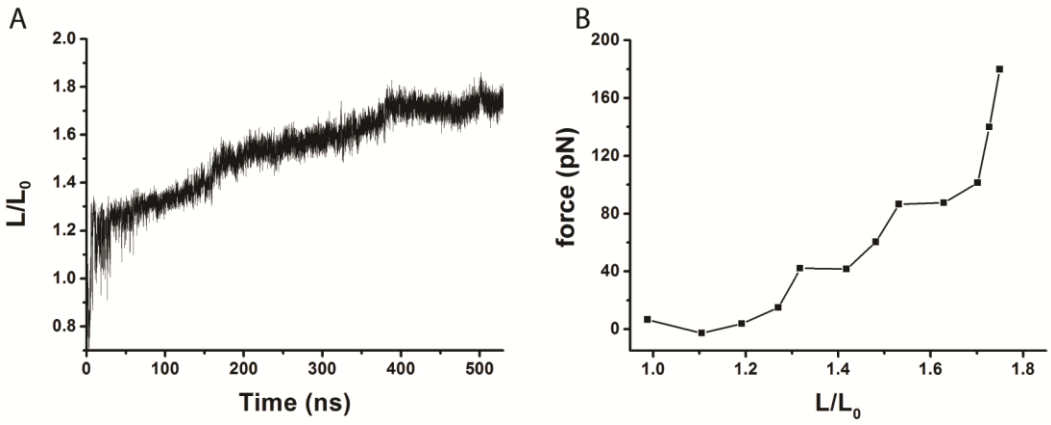


Figure 3.3.18 Steered MD simulation of B-to-S transition. (A) Extension time course and (B) force-extension curve of overstretching DNA.

Next we take a close look at the DNA structure changes during the overstretching transition. Six snapshots in different transition stages were pictured in Figure 3.3.19, which suggested a partially unwinding transition accompanied with an elongation of the end-to-end extension along the central axis of DNA. Further, peeling was suppressed during the transition, where base pairings were still intact for all 15 base pairs.

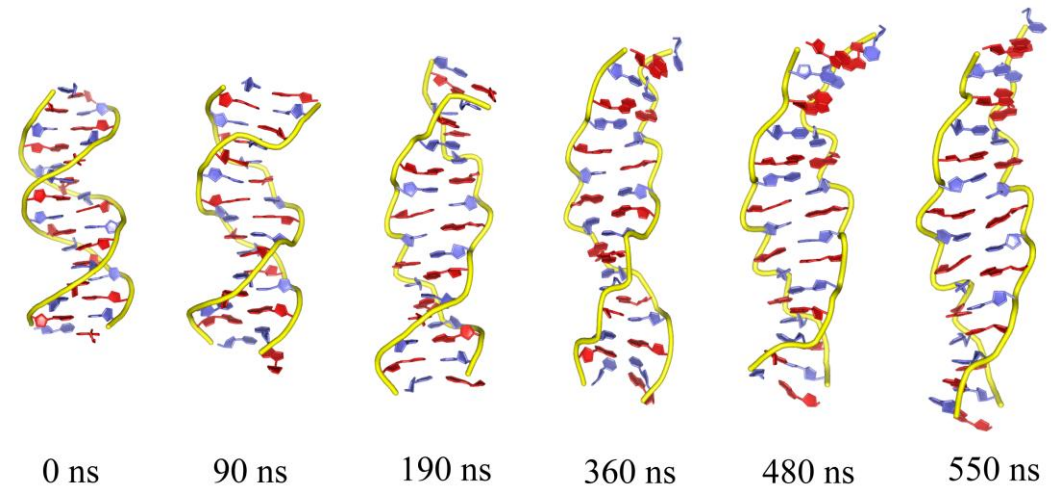


Figure 3.3.19 Snapshots of the DNA structure during the transition. DNA unwinding and elongation were clearly observed in the simulation.

Therefore, the transition happened in the simulation was clearly not the B-to-ss peeling transition but the high cooperative B-to-S transition. This implies that the overstretched form at the end of the simulation may be the structure of the

mysterious S-DNA. As indicated by the DNA structure at 550 ns, the 15 bp DNA with 1.5 helical pitches in B-DNA structure becomes roughly half of the helical pitch in S-DNA, suggesting that S-DNA might have a helical pitch of ~30 bp, which is consistent with previously reported helical pitch of 35 bp (144) or 18 nm (145) for S-DNA.

Following, we take a step forward to examine the detailed structure changes of this transition. In order to exclude the boundary influence, we only focused on five base pairs located in the middle (6^{th} to 10^{th} bp). As we already observed the unwinding by the snapshots of the DNA structure, we then seek ways to quantitatively measure how much the DNA unwound. This requires a definition of the base pair geometry, which has been well characterized by Lu (146) using 16 parameters shown in Figure 3.3.20. These values can be easily calculated by the software package 3DNA (146,147).

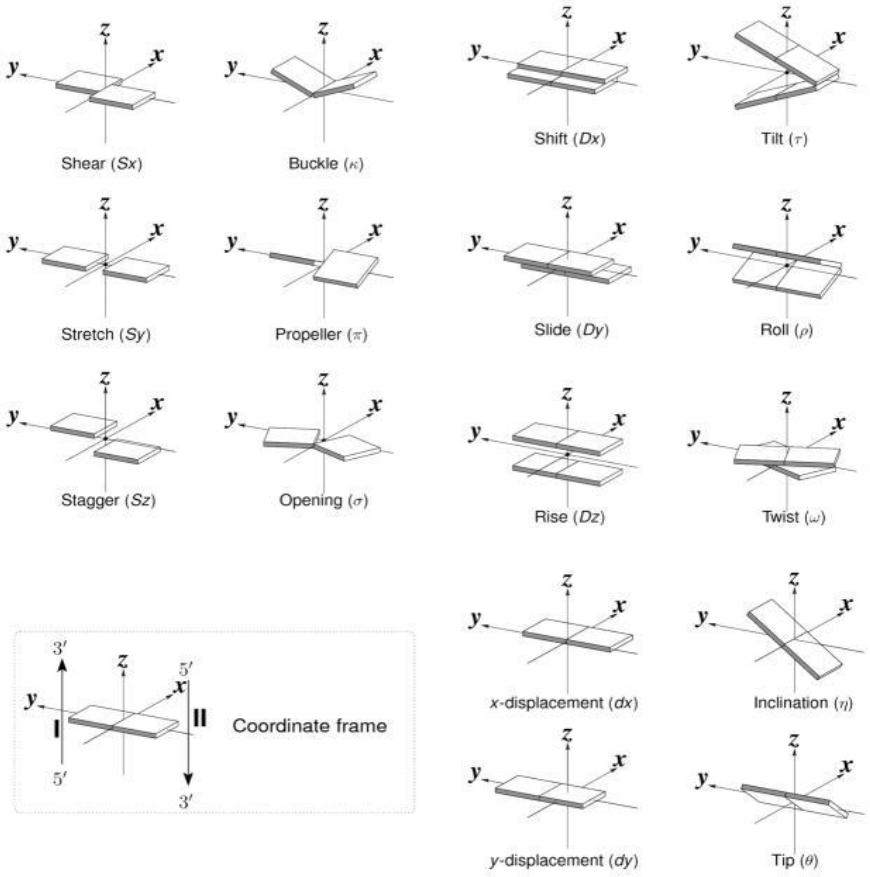


Figure 3.3.20 16 parameters to define the base pair geometry. This picture is adopted from Lu (146).

According to the definition, the parameter suitable for describing the unwinding of the DNA is the twist angle between adjacent base pairs. As shown in Figure 3.3.21, the twist angle between adjacent base pairs indeed decreased as DNA elongated. Moreover, an average of ~20 degrees reduction was observed for 4 twist angles, which further suggests that DNA underwent partially unwinding process when stretched, in agreement with what was observed in the snapshot of the DNA structures.

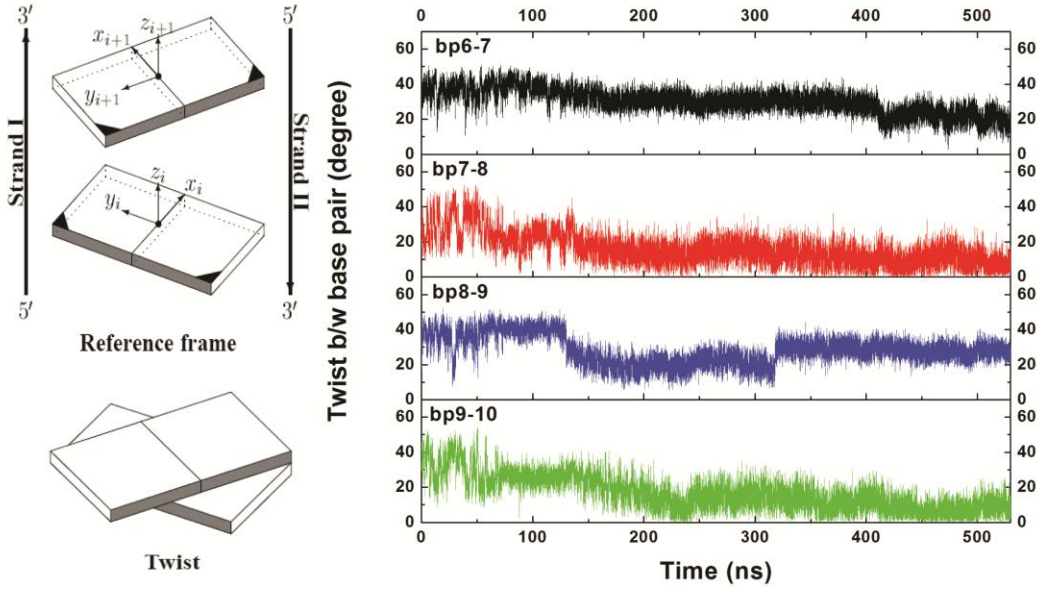


Figure 3.3.21 Change of the twist angle between adjacent base pairs (the middle base pairs from 6th bp to 10th bp) analyzed by 3DNA http://x3dna.org/.

Therefore, we concluded that B-to-S transition might be a process involved elongation and partially unwinding of the B-DNA with base pairing still intact and the new S-DNA is a stable form under physiological condition. Although the relevant biological function of the S-DNA has not been found yet, the possible implications were proposed that S-DNA might be a potential preferred binding substrate for DNA intercalators such as YOYO-1 and RecA, which elongates DNA to a similar level of S-DNA. Another proposed application of S-DNA is that it might be used as a pN range force ruler. As B-to-S transition occurred in a narrow force range and less sensitive to the factors affecting base pair stability, it

would be a perfect nano ruler sensing a few pN force changes in wide range of conditions.

In this chapter, we used three different methods to investigate DNA structural transitions under tension. The transfer matrix method allows us to calculate physical parameters in equilibrium, such as the fractions of different DNA forms or the possibilities of different DNA forms for each base pair. The kinetic Monte Carlo simulation allows us to examine the kinetics of the transition process with proper choices of the transition rates, which then can provide explanations for experimental observation. The SMD simulation can provide atomic level details of DNA structural changes (hydrogen bonds, twist) during transitions. Each method addresses a different aspect of the problem, and by combining these methods together we then are able to gain a thorough understanding of the DNA overstretching transition.

CHAPTER 4 Mechanism of DNA Organization by *Mycobacterium tuberculosis* Protein Lsr2

4.1 Introduction

Lsr2 in *Mycobacterium tuberculosis* is the only identified H-NS family protein in gram-positive bacteria, which has been demonstrated to have structural resemblance and functional complementarities with H-NS in *E coli.*. It takes an important part in chromosomal DNA compaction and organization, involved in many cellular processes and plays as a global gene silencer in *Mycobacterium tuberculosis*.

In this chapter, combining magnetic tweezers and AFM, we show that the *M. tuberculosis* Lsr2 protein binds to DNA through a highly cooperative process resulting in an increase in apparent DNA bending rigidity, similar to H-NS. In addition, Lsr2-induced DNA-folding was also observed. The rigid Lsr2-DNA complex was also found to be stable across physiological environmental changes (salt, pH and temperature). This is more prominent in the case of magnesium salt, an important co-factor in numerous biological processes, where the rigid Lsr2-DNA complex is stable at physiological range of $MgCl_2$ concentration. We also demonstrate that the Lsr2-DNA complex strongly restricts DNA accessibility, a property which is shown in H-NS *E. coli* paralogue, StpA (79). In summary, our work shows that Lsr2 has an intricate DNA-binding mode that encompasses DNA-stiffening and DNA-folding, which provides us a mechanistic platform in understanding how Lsr2 mediates its biological functions *in vivo*.

4.2 Material and Methods

4.2.1 Over-expression and purification of Lsr2

pET expression vector containing the lsr2 gene was expressed according to previous protocol (122). The expressed Lsr2 protein has a C-terminal His-tag to aid protein purification.

4.2.2 Magnetic tweezers experiments

The transverse magnetic tweezers setup utilized in this study was described in section 2.1. A single λ-DNA (48,502bp, NEB), modified with biotin at both ends, was tethered between a streptavidin-functionalized surface and a streptavidin-coated 2.8 μm magnetic bead (Dynabeads M-270 Streptavidin, Invitrogen). The single DNA molecule extension measurement was collected in real-time (100 Hz) using a camera-based centroid tracking software written in LabVIEW program (National Instruments, U.S.A). This setup was previously used to measure changes in DNA rigidity caused by either DNA-stiffening proteins (71) or DNA-bending proteins (135).

4.2.3 Atomic force microscopy imaging

All AFM imaging experiments were done on glutaraldehyde-coated mica surface, which prevents the non-specific aggregation of proteins or DNA-protein complexes because the glutaraldehyde molecules are covalently bound to the surface. Preparation of glutaraldehyde-coated mica was done according to the protocol discussed in section 2.2.2. In all AFM experiments, linearized φX174 DNA (5,386 base pairs, NEB, U.S.A) was used as the DNA template and incubated with the stated Lsr2 concentration or DNA/protein ratio for 20 minutes

in 10 mM Tris-HCl, 50 mM KCl, pH 7.5 buffer condition before depositing on the glutaraldehyde-coated mica. The Lsr2 protein bound on the DNA interacts with the glutaraldehyde-coated mica surface to form covalent bonds thus trapping the Lsr2/DNA complexes on the mica surface without the aid of divalent salts. The sample was then rinsed with deionised water and dried with a clean stream of nitrogen gas before using it for air AFM imaging. Typical AFM image scan size is 1-4 μm square with a scan speed of 1-2 line per second.

4.3 Results

4.3.1 Lsr2 cooperatively binds to extended DNA and stiffens DNA

DNA-distorting proteins can modify the micromechanical properties of DNA and affect its force response. Different types of DNA-distorting proteins cause different DNA force responses, which can be measured by single-DNA stretching experiments (133). Previous AFM experiments revealed that Lsr2 could bridge DNA into DNA hairpins and loops (122); however, it is unclear how Lsr2 initially interacts with an extended DNA before DNA folds, which provides the physical basis for subsequent DNA organization. Further, understanding of the property of Lsr2 on extended DNA will also provide important insights into its gene regulatory function. Therefore, we implemented a quick force-jump measurement that is able to measure the force response of DNA while preventing DNA folding during the measurement (119). This quick force-jump measurement is performed by initially holding the DNA at a high force (~20 pN) which prevents DNA folding, then jumping to a lower force for around 1 second to measure the end-to-end distance of DNA (i.e., DNA extension) before jumping back to the high force. As the DNA is only held at lower forces for very short duration, the level of DNA folding occurred at the lower forces is negligible. Repeating this process for a series of lower forces in the range of 0.3 - 16 pN, a force-extension curve is obtained which quantifies the DNA force response without interference from DNA folding. One thing to be noted is that the equilibrium state is not reached in the force-jump measurement, indicated by progressive folding if holding the DNA at the low force. We used the force-jump procedure to prevent folding in order to obtain the conformational property of the rigid filament. The formation of this rigid filament, which occurred when DNA was held at high force, is assumed having reached a steady state at high force. When it jumped to a lower force, within a short time, we reason that the property of this filament remained unchanged.

Figure 4.3.1 shows the force-extension curves obtained by force-jumping of a λ-DNA (48,502 bp) with increasing Lsr2 concentration in 10 mM Tris-HCl, 50 mM KCl, pH 7.5 buffer condition. At each Lsr2 concentration and each force, this force-jump procedure was repeated three times to get the average values (data points) and the standard deviations (error bars) of the extensions. At 6 nM Lsr2 concentration, the DNA force response is similar to that of the naked DNA, suggesting few Lsr2 binding to DNA, which causes negligible change in the force response of DNA. Increasing Lsr2 concentration to 60 nM, the DNA extension becomes longer than that of the naked DNA, which indicates the increase of the DNA rigidity (133). Further increasing the Lsr2 concentration to 600 nM makes the DNA more extended and increasing to 2,400 nM only gives a slight increase in DNA extension compared to the extension in 600 nM Lsr2, which means the stiffening effect is largely saturated by 600 nM Lsr2. Here we emphasize that the DNA stiffening caused by Lsr2 is not due to steric interaction of overcrowded Lsr2 on DNA. An example is that the *E. coli* IHF, a DNA bending protein that binds to DNA as individual heterodimers, does not cause DNA stiffening even at over saturated IHF concentrations (135). More directly, significant DNA-stiffening by Lsr2 occurred before binding saturation (e.g., at 60 nM).

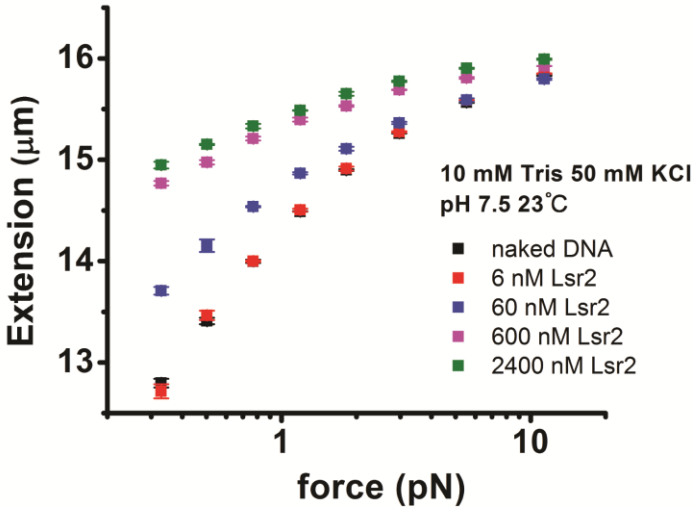


Figure 4.3.1 Formation of rigid Lsr2 nucleoprotein filament on extended 48,502 bp λ-DNA. Force-jump force-extension curves of DNA incubated with varying Lsr2 concentrations, which shows significant DNA stiffening. The error bar at each force is given by three successive force-jump experiments on the same DNA.

The effect of Lsr2 on the DNA bending rigidity can be quantified by fitting the measured force-jump force-extension curves with the curves predicted by the worm-like chain (WLC) polymer model of DNA (49), which only depends on the contour length and the DNA bending rigidity described by a parameter called the DNA bending persistence length. As mentioned in section 1.2.3, for a naked DNA, this parameter has been measured to be around 50 nm in physiological solution conditions (47,49). Figure 4.3.2 shows the bending persistence lengths *A* and contour lengths *L* at different Lsr2 concentrations *C* fitted by the Marko-Siggia formula (inset in Fig. 4.3.2). The effective contour length is nearly constant over the Lsr2 concentration range of 0 – 2,400 nM, while the effective persistence length increases drastically from ~ 50 nm at 0 nM Lsr2 to ~ 490 nm at 2,400 nM Lsr2. The large error bars of the persistence length are due to the stochastic non-specific binding between Lsr2 and DNA. These results reveal that Lsr2-binding does not change the DNA native structure (otherwise one would expect change in the effective contour length); however, it restricts DNA bending resulting in DNA stiffening.

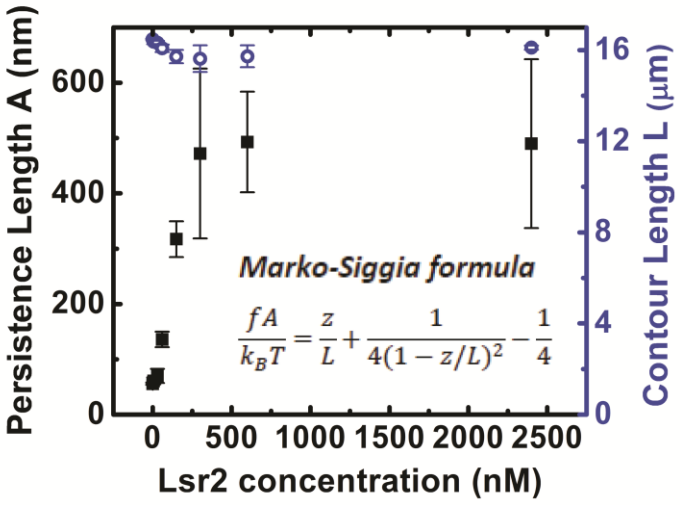


Figure 4.3.2 The bending persistence lengths *A* (black solid square) and the contour lengths *L* (blue open circle) at different Lsr2 concentrations *C* of the resulting extended Lsr2-DNA complex fitted according to the Marko-Siggia formula (inserted formula). Here f denotes the stretching force and z denotes the DNA extension. The error bar at each concentration is given by at least three independent measurements on different DNAs. At saturation (2,400 nM Lsr2), *A* and *L* are determined to be 489 nm ± 152 nm and 16,109 nm ± 94 nm.

As Lsr2 DNA-stiffening only changes the DNA bending persistence length rather the contour length, it is reasonable to assume that both Lsr2 fully coated DNA and naked DNA have different persistence lengths ($A_{saturated}$ and A_{naked}) and the same contour length L . Therefore, if the DNA is partially coated with Lsr2, the measured extension will be the mixture of the extension devoted by the Lsr2-coated part α (occupation fraction) and the naked part $1-\alpha$:

$$z_{measured} = \alpha z_{saturated} + (1-\alpha) z_{naked}$$

Re-arrange the above equation to get α:

$$\alpha = \frac{z_{measured} - z_{naked}}{z_{saturated} - z_{naked}}$$

According to the Marko-Siggia formula at high force range (0.1~10 pN), the extension of the DNA under force can be calculated by the equation $z = L(1 - \sqrt{\frac{k_B T}{4fA}})$. Since the contour length is the same for both the Lsr2-coated DNA and naked DNA, substituting the Marko-Siggia equation to α will lead us to the occupation fraction formula:

$$\alpha = \frac{\sqrt{1/A_{measured}} - \sqrt{1/A_{naked}}}{\sqrt{1/A_{saturated}} - \sqrt{1/A_{naked}}}$$

As the persistence length is obtained by fitting the force-extension curve in a force range, α obtained by above formula is independent of force. With this formula, we are able to calculate the occupation fraction obtained at different Lsr2 concentrations, which allows us to calculate the dissociation constant k_d to be ~ 58 nM and the Hill coefficient n to be ~ 2.3 by fitting to the Hill equation $\alpha = 1/[(k_d/C)^n + 1]$ (Fig. 4.3.3) (148). The above result reveals that Lsr2 binds to extended DNA at nanomolar affinity. A greater than one Hill coefficient indicates that Lsr2 cooperatively binds to DNA. The cooperative nature can be explained by Lsr2 polymerizing along DNA to form a filamentous structure, similar to H-NS family proteins in gram-negative bacteria (71,79,114). Here one should notice that the DNA is forced to be in an extended conformation in the force-jump DNA

stretching experiments; therefore, these values are not directly comparable with values obtained by the conventional electrophoretic mobility shift assay (EMSA). For example, k_d of 1 μM was reported by EMSA before (126), which is about 20 times the k_d value measured in our force-jump experiments.

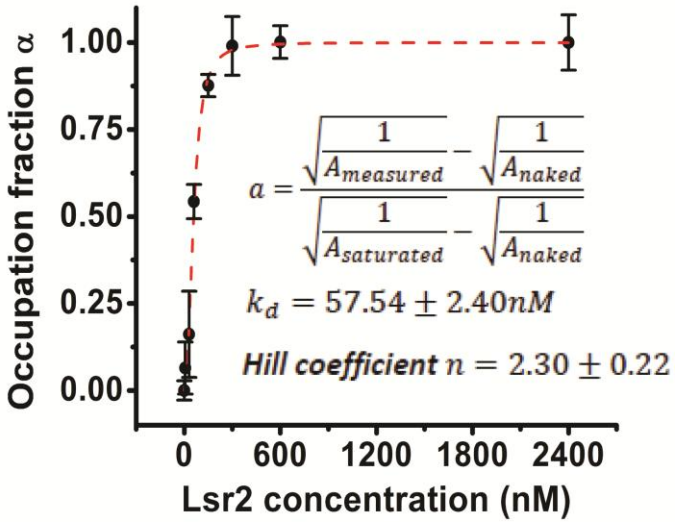


Figure 4.3.3 The fraction of DNA occupied by Lsr2 was calculated according to the apparent bending persistence length (see inserted formula). Its dependency on Lsr2 concentration reveals high binding affinity and cooperativity with k_d of 57.54 ± 2.40 nM and Hill coefficient of 2.30 ± 0.22.

Our own EMSA result yielded similar k_d value of 1.7 μM as shown in Figure 4.3.4, indicating the observed k_d difference is unlikely that of a variation in protein batch activity. The dissociation constant k_d is defined by the protein concentration at which half of the DNA is bound under the condition that the amount of free proteins are in excess of the available binding sites. But in EMSA experiment this condition may not always be satisfied especially under low protein/DNA ratio, as the amount of DNA is usually above 50 ng for better visualization. Therefore, EMSA is usually not suitable for low k_d measurement at high substrate concentration. This has already been demonstrated in the cases of k_d measurements of RecX (149,150) and IHF (151-153). EMSA measurement suggested a k_d value for RecX-ssDNA binding to be ~1 μM (149), while a fluorescence anisotropy based measurement showed a higher ssDNA affinity for

RecX with k_d ~80 nM (150). Similarly, EMSA indicated a k_d value for specific IHF-DNA binding to be 2-20 nM (151), but a recent stop flow measurement at low DNA concentration and a recent single-DNA stretching measurement demonstrated a much lower k_d < 1 nM (152,153). Therefore, the k_d value for Lsr2-DNA binding obtained from our single-DNA stretching experiment is likely more reflective of the actual dissociation constant than the value indicated by EMSA measurement.

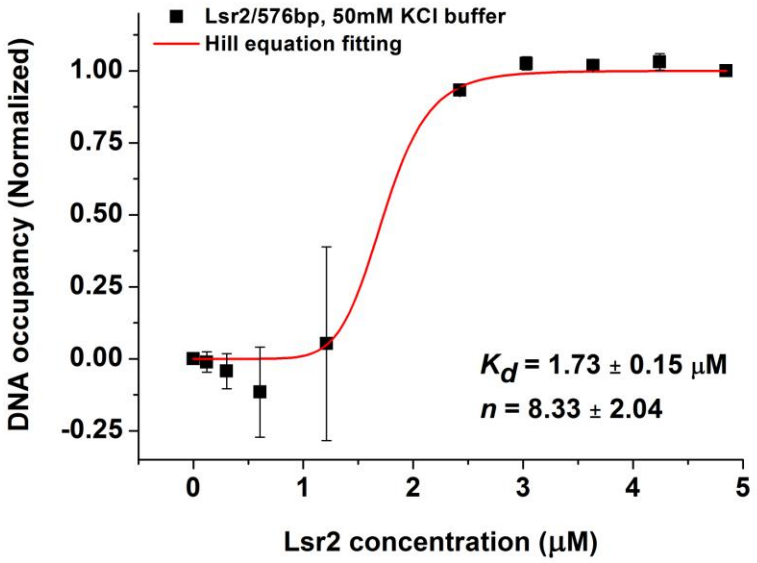


Figure 4.3.4 Electrophoretic mobility shift assay (EMSA) of Lsr2-DNA interaction in 10 mM Tris-HCl, 50 mM KCl, pH 7.5 buffer condition. Fitting the Hill equation yields kd to be 1.7±0.15 μM and Hill coefficient n to be 8.33±2.04.

4.3.2 **The rigid Lsr2-DNA complex condenses under low force**

Previous AFM experiments revealed that Lsr2 could fold large DNA into hairpins and higher order conformations on freshly cleaved mica surface (122). Our independent AFM imaging results are consistent with the previous findings. Fig. 4.3.5 A&B shows AFM images of naked linearized double-stranded φX174 DNA (5,386 bp) on glutaraldehyde-modified mica surface, which exhibits random coiled conformation. At an Lsr2 monomer to DNA base pair ratio of 1: 1 (Lsr2 concentration of 300 nM), DNA is typically folded into highly complex Lsr2-

DNA condensates consisting of large compact globular nucleoprotein structures (white arrows) and extended thick DNA bundles (red arrows) as shown in Fig. 4.3.5 C&D. At a lower Lsr2 monomer to DNA base pair ratio of 1: 10 (Lsr2 concentration of 30 nM), the Lsr2-DNA complex typically have an Lsr2 rich core with higher height (yellow arrow) surrounded by large naked DNA loops (Fig. 4.3.5 E&F).

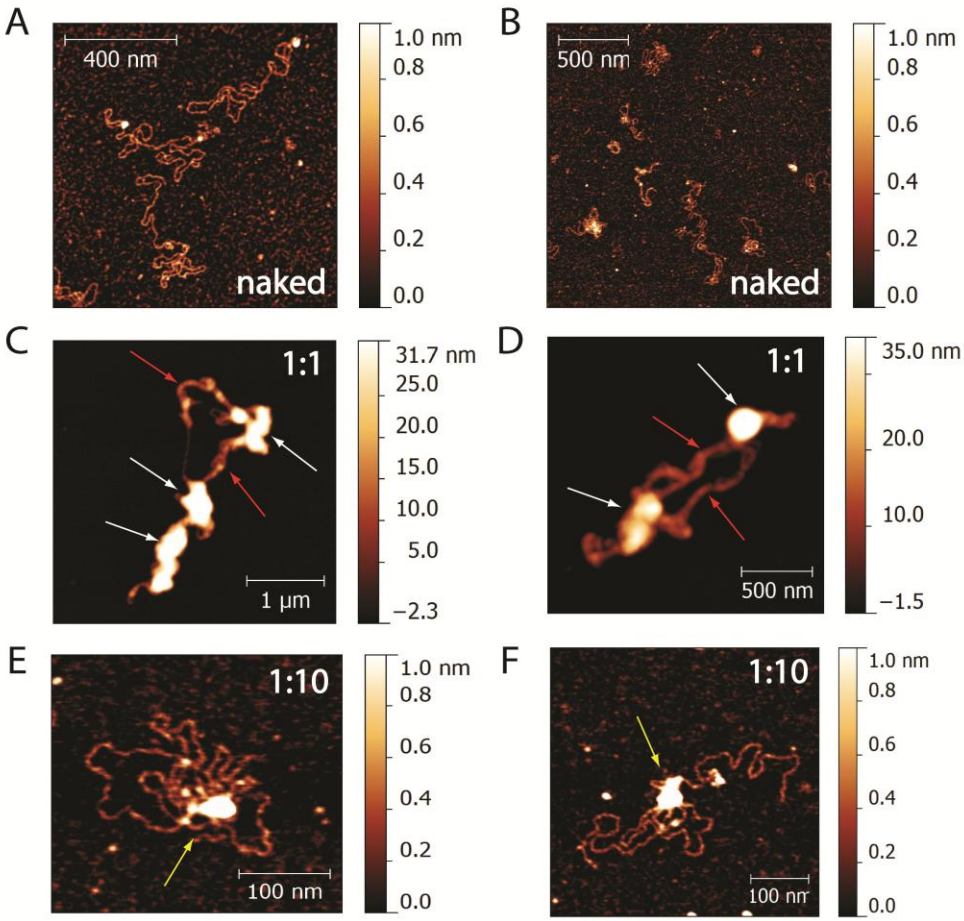


Figure 4.3.5 The rigid Lsr2-DNA complex condenses under low force. AFM images show typical random coiled naked linearized 5,386 bp φX174 DNA (A&B), typical Lsr2-DNA complexes consisting of large condensates (white arrows) and extended thick DNA bundles (red arrows) at an Lsr2 monomer to DNA bp ratio of 1:1 (C&D), and typical Lsr2-DNA complexes consisting of Lsr2-riched cores (yellow arrow) surrounded by large naked DNA loops at a lower Lsr2 monomer to DNA bp ratio of 1:10 (E&F).

As the results in the previous section have demonstrated that Lsr2 can cooperatively bind to extended DNA which stiffens DNA, an interesting question

raised here is whether a preformed rigid extended Lsr2 nucleoprotein structure at high force can fold when the force is dropped to lower values in single-DNA stretching experiments. Figure 4.3.6 shows the force-extension curves obtained in a force-decrease scan (red solid squares) followed by a force-increase scan (red open squares) through the same set of force values of a λ-DNA with 600 nM Lsr2 concentration in 10 mM Tris-HCl, 50 mM KCl, pH 7.5 solution condition. At each force, the DNA was held for 30 seconds, and extension average over this period is plotted in Figure 4.3.6 as a data point. Different from the previous force-jump, this force-scan procedure allowed DNA folding as the DNA was held at lower forces for much longer duration.

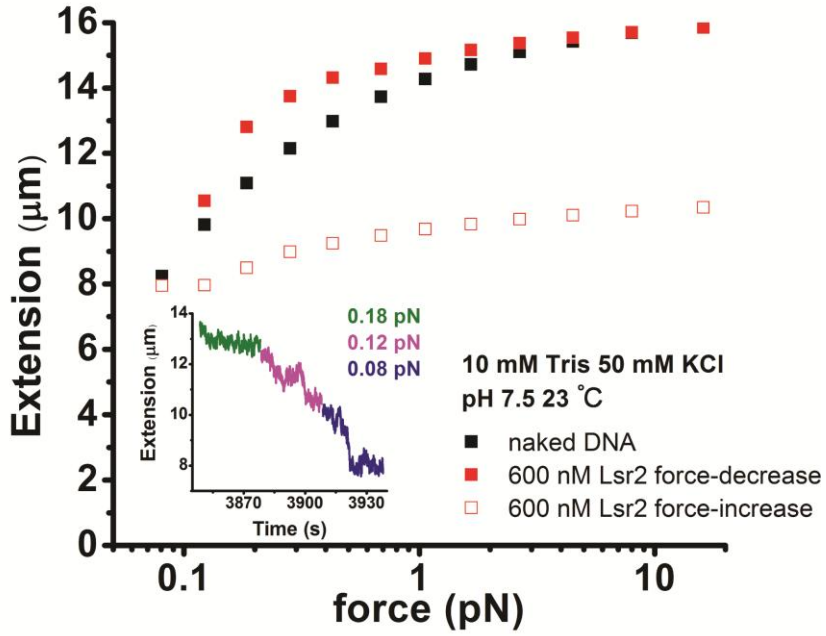


Figure 4.3.6 Force-extension curves obtained by a force-decrease scan (red solid squares) followed by a force-increase scan (red open squares) through the same set of force values of a λ-DNA at 600 nM Lsr2 concentration in 10 mM Tris-HCl, 50 mM KCl, pH 7.5. Inset shows progressive DNA folding at small force (< 0.2 pN). The non-overlapping force-extension curves between the force-decrease and force-increase scans indicate the mixed effects of DNA stiffening and DNA folding.

If DNA folding were to occur during the force-decrease scan at the lower force range, non-overlapping force-extension curves (i.e., hysteresis) between the force-decrease and force-increase scans would be expected, which indeed

occurred (Fig. 4.3.6). The DNA extension obtained in the force-decrease scan is overall longer than the naked DNA, indicating formation of rigid nucleoprotein structure at higher force range; while the shorter than naked DNA extension obtained in the subsequent force-increase scan indicates DNA folding at lower forces. Progressive DNA folding time trace in the force range 0.08 – 0.18 pN was shown in the inset of Figure 4.3.6.

Another independent force-scan stretching experiment showed that the folded Lsr2-DNA complex is extremely stable, which can withstand 20 pN over the experimental time scale of 20 minutes (Fig 4.3.7). Overall, the folding of rigid Lsr2-DNA complexes at low force is consistent with the folded Lsr2-DNA complexes observed in AFM imaging.

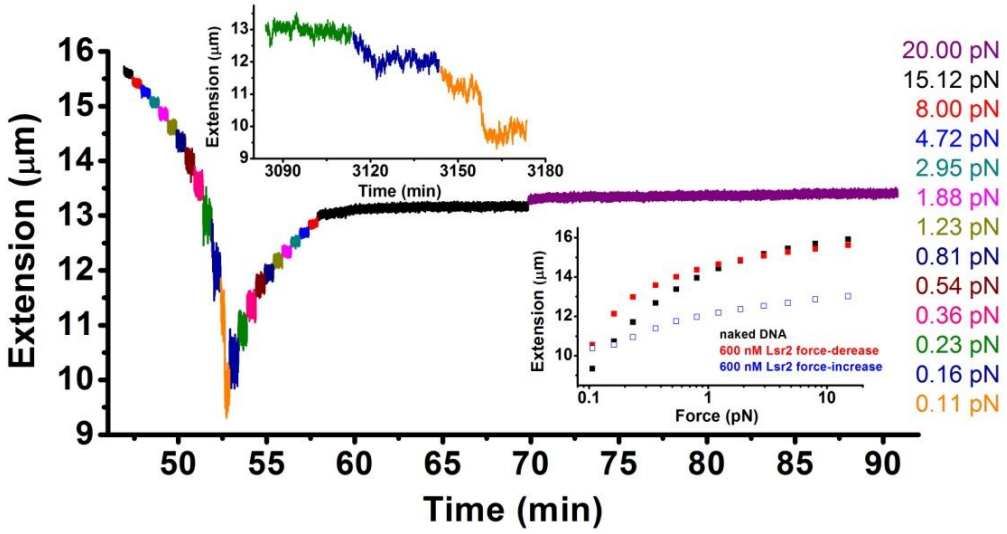


Figure 4.3.7 Mechanical stability of folded Lsr2-DNA complex. A λ-DNA was folded in 600 nM Lsr2 at < 0.3 pN (see top inset) during a force-decrease scan (30 sec for each force). The folded DNA is mechanically stable, able to withstand large force during the subsequent force-increase scan. Holding the DNA at 20 pN for 20 minutes, the DNA remained folded. Corresponding force-extension curve is shown in the lower inset panel.

The results shown in this section suggest that DNA stiffening by Lsr2 at initial binding stage does not exclude DNA folding. This is similar to MvaT, where the formation of rigid MvaT nucleoprotein filament was reported to precede and mediate MvaT-dependent DNA folding (114).

4.3.3 The effects of salt, pH and temperature changes to Lsr2-DNA organization properties

Formation of the rigid nucleoprotein filamentous structures has been shown universal in H-NS family proteins and critical for their gene-silencing functions in gram-negative bacteria (71,79,114,119,120). The formation of the H-NS-like nucleoprotein filaments by those proteins was often regulated by environmental factors such as salt concentrations, pH value, and temperature (71,79,114,119,120). As Lsr2 has been proposed to be the first H-NS family protein in gram-positive bacteria (121), and it also forms rigid nucleoprotein structure on extended DNA by cooperative DNA binding, in this section we focus on how environmental factors affect the formation of the Lsr2 nucleoprotein structure on the extended DNA.

Since Lsr2 can simultaneously stiffen and fold DNA, two sets of experiments were conducted to separately investigate the effects of the environmental factors on these two Lsr2-DNA binding modes. To investigate the effects on the DNA-stiffening property, force-jump procedure described in the previous section was performed to prevent DNA folding. Then a force-scan (force decrease scan followed by force increase scan) experiment was conducted to probe the effects on the DNA-folding property.

Previous studies showed that the function of DNA protection against hydroxyl radical damage of Lsr2 depends on salt concentration and it lost the function in 800 mM NaCl buffer condition (154). Here we show that at 800 mM KCl concentration, Lsr2 at 600 nM is unable to form the rigid nucleoprotein structure on extended DNA (Fig. 4.3.8A), indicated by no changes in the force-extension curves between the naked DNA (blue solid squares) and the same DNA after Lsr2 was introduced (yellow solid squares). DNA folding did not occur either in the force-scan procedure where the DNA was held at lower forces for longer time (Fig. 4.3.8B).

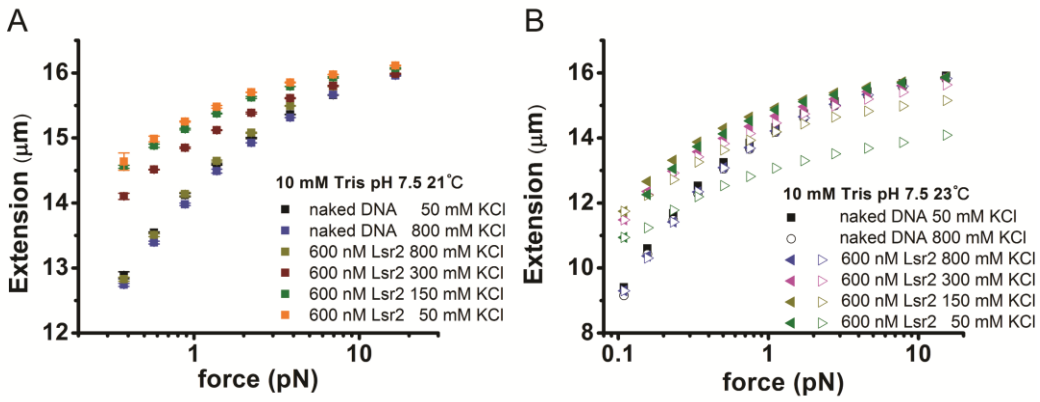


Figure 4.3.8 Effects of KCl concentration on rigid Lsr2 nucleoprotein structure formation at 600 nM Lsr2. Force-jump force-extension curves (A) and force-scan force-extension curves (B) of rigid Lsr2-DNA complex are plotted separately with decreasing KCl concentration. Two independent λ-DNA tethers were used to obtain data in the two panels. Different colors indicate different experimental conditions. Level of DNA stiffening is negatively regulated by increasing KCl concentration. DNA folding occurred under low forces in low KCl concentration.

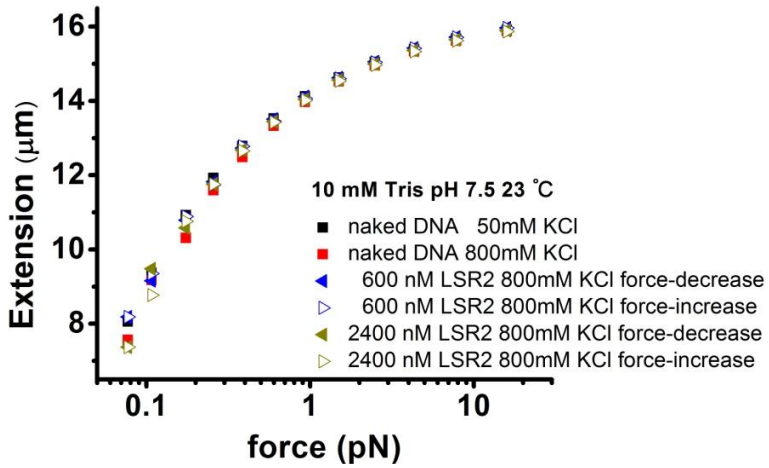


Figure 4.3.9 Reduction of Lsr2 DNA-binding affinity in high salt (800 mM KCl) buffer condition. The overlapping of all force-extension curves obtained by force-decrease and force-increase scans at 0, 600 and 2,400 nM Lsr2 concentrations indicate few Lsr2 proteins bind to DNA, which suggests a reduction of Lsr2 DNA-binding affinity in high salt conditions.

Similar results were obtained at 2,400 nM Lsr2 as shown in Figure 4.3.9, suggesting that at 800 mM KCl, Lsr2 fails to stably bind to DNA at μM concentration range. On the same DNA and at the same Lsr2 concentration of 600 nM, DNA stiffening was observed when the force-jump experiments were

repeated in lower KCl concentrations of 300 mM, 150 mM, 50 mM, successively (Fig. 4.3.8A). The DNA became increasingly stiffer as the KCl concentration was decreased, suggesting the formation of rigid Lsr2-DNA complex in lower KCl concentrations. Similarly, folding was also observed when lowering the KCl concentration (Fig. 4.3.8B), indicated by the hysteric force-extension curves between the force-decrease and force-increase scans.

Previous studies on *E. coli* H-NS have shown that the capability to form rigid nucleoprotein filaments is reduced at higher magnesium concentration. At 10 mM $MgCl_2$, the *E. coli* H-NS in μM range of concentration is unable to stiffen DNA (71,72,155). In contrast, other H-NS family proteins such as StpA and MvaT are able to form rigid nucleoprotein filaments insensitively to magnesium concentration of the same range (79,114). Figure 4.3.10A shows that the force-jump curves obtained with 600 nM Lsr2 in 0, 1, 4 and 10 mM $MgCl_2$ all overlap, indicating that the formation of rigid Lsr2 nucleoprotein structure is insensitive to 0-10 mM $MgCl_2$, similar to StpA and MvaT. Additionally, DNA-folding still occurs under low forces in 0-10 mM MgCl2 in the force-scan procedure (Fig. 4.3.10B)

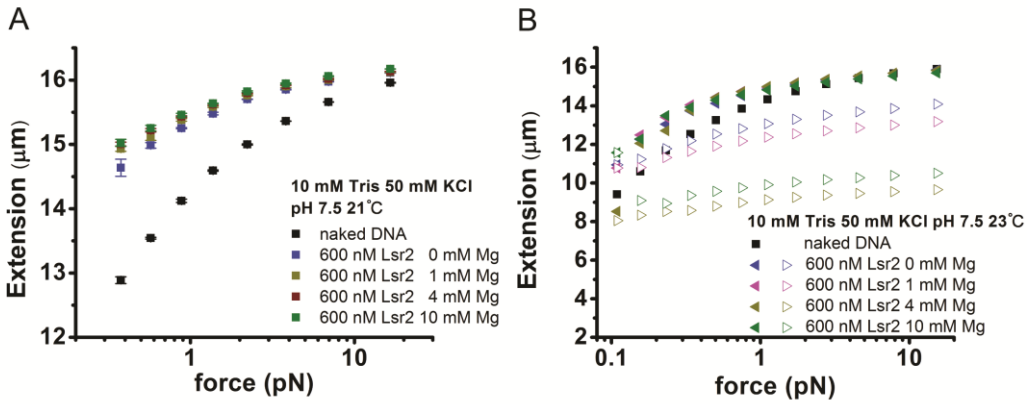


Figure 4.3.10 Effects of magnesium concentration on rigid Lsr2 nucleoprotein structure formation at 600 nM Lsr2. Force-jump force-extension curves (A) and force-scan force-extension curves (B) of rigid Lsr2-DNA complex are plotted separately with increasing magnesium concentration. Two independent λ-DNA tethers were used to obtain data in the two panels. Different colors indicate different experimental conditions. The level of DNA-stiffening is insensitive to magnesium concentration over the range explored in experiments. DNA folding occurred in 0-10 mM magnesium concentration.

Similar studies were performed to investigate the effects of temperature (Fig. 4.3.11) and pH (Fig. 4.3.12). DNA stiffening by Lsr2 was found moderately tuned by temperature (Fig. 4.3.11A). At the human body temperature of 37 ^{o}C, significant reduction in the DNA-stiffening effect of Lsr2 was observed which is also observed in *E. coli* H-NS (71). This effect can be explained by either a disruption in Lsr2 DNA-stiffening ability or a reduction in Lsr2 DNA-binding affinity at 37 ℃, such as in the case of high salt buffer conditions. In contrast, the DNA stiffening effect by Lsr2 was found insensitive to pH values ranging from 6.8 to 8.8 (Fig. 4.3.12A), unlike the highly pH sensitive *E. coli* H-NS (71). We also showed the DNA-folding effect by Lsr2 is not sensitive to changes in buffer temperature (Fig. 4.3.11B) or pH value (Fig. 4.3.12B), as DNA-folding can always be induced in all the conditions explored under low forces.

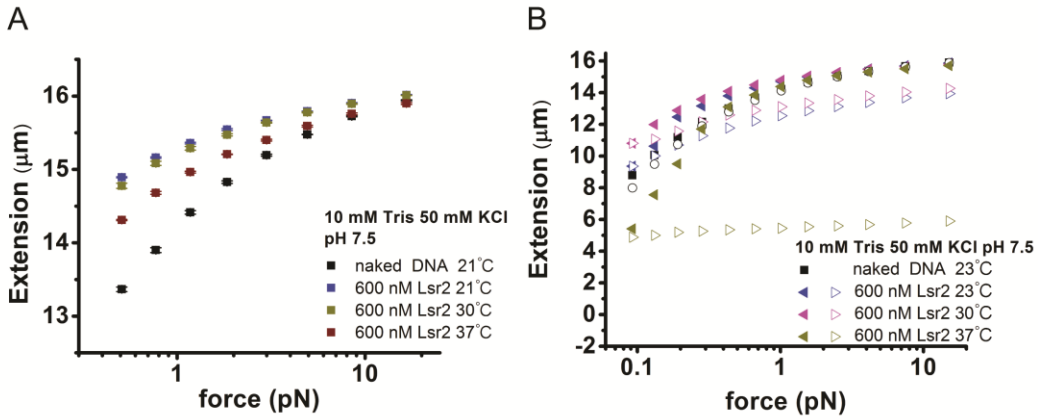


Figure 4.3.11 Effects of buffer temperature on rigid Lsr2 nucleoprotein structure formation at 600 nM Lsr2. Force-jump force-extension curves (A) and force-scan force-extension curves (B) of rigid Lsr2-DNA complex are plotted separately with increasing temperature. Two independent λ-DNA tethers were used to obtain data in the two panels. Different colors indicate different experimental conditions. Level of DNA stiffening is negatively regulated by increasing temperature.

These results suggest that despite of the Lsr2-DNA nucleoprotein complex formation lost under high salt condition (800 mM KCl) and the moderate reduction of the DNA stiffness at the human body temperature; the Lsr2-DNA

complex is a robust structure not sensitive to the changes of environmental conditions over physiological ranges.

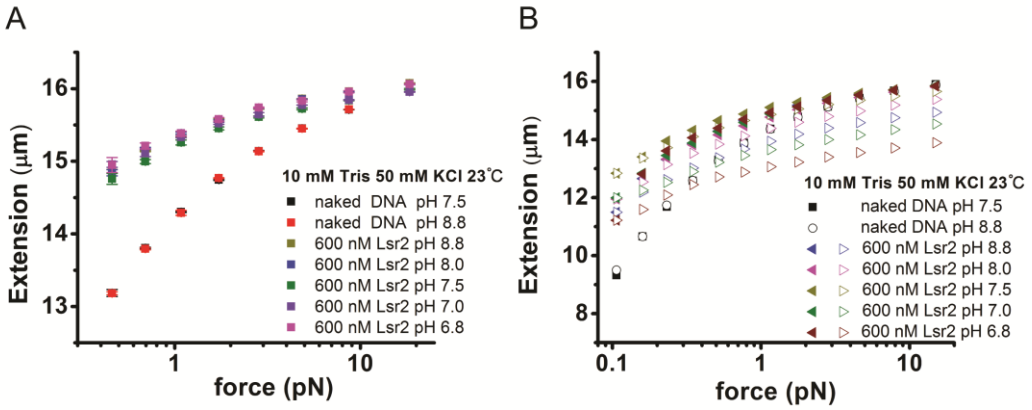


Figure 4.3.12 Effects of pH value on rigid Lsr2 nucleoprotein structure formation at 600 nM Lsr2. Force-jump force-extension curves (A) and force-scan force-extension curves (B) of rigid Lsr2-DNA complex are plotted separately with decreasing pH value. Two independent λ-DNA tethers were used to obtain data in the two panels. Different colors indicate different experimental conditions. Lsr2-DNA nucleoprotein complex is insensitive to pH value over the range explored in experiments.

4.3.4 The rigid Lsr2-DNA complex is able to restrict access to DNA

Previous biochemical study already demonstrated that Lsr2-DNA complexes are resistant to DNase I digestion (126). As Lsr2 can form rigid nucleoprotein filamentous structure on extended DNA which also mediates higher level of DNA condensation, it is unclear whether the DNA protection from DNase I digestion in that experiment is due to DNA condensation or the formation of the Lsr2-DNA nucleoprotein filamentous structure alone is sufficient for DNA protection. Therefore, in this section we examine the level of restriction of the accessibility to extended DNA covered by Lsr2.

By using a multiplex detection algorithm developed in previous study (79), dozens of DNA tethers were stretched at a force ~ 10 pN and monitored at the

same time as shown in Fig. 4.3.13. The stretched DNA would allow the formation of the Lsr2-DNA filament once Lsr2 was introduced in the channel. As DNA is always invisible to us, the number of the magnetic beads was counted versus time. If DNase I binds to DNA and cut the DNA tether, we then would lose one count of the bead. By recording the number of the remaining beads, the DNase I digestion rate can be obtained.

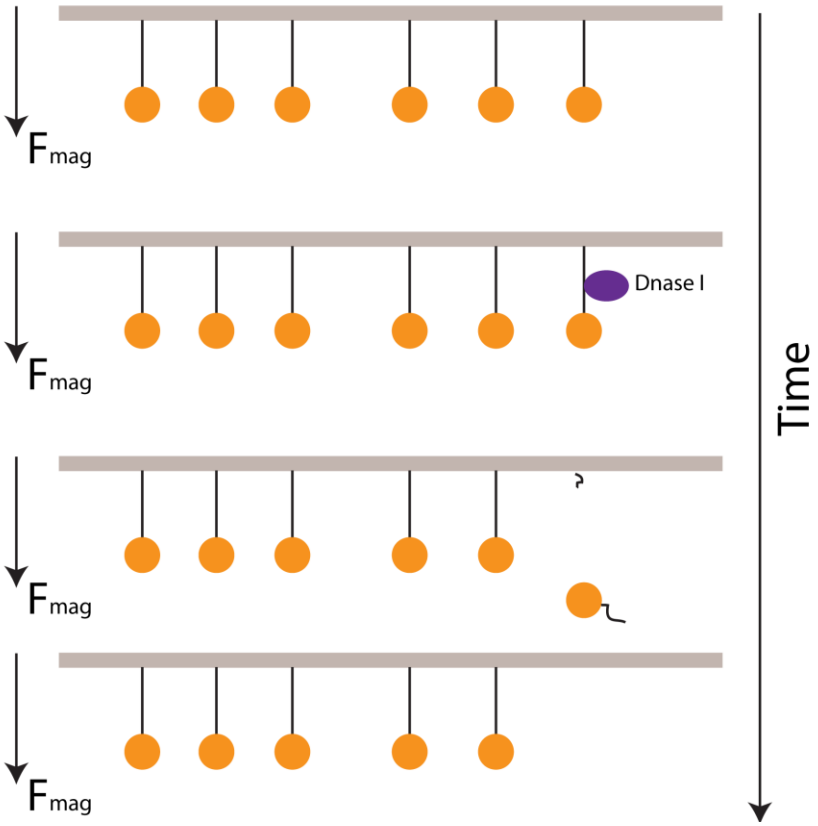


Figure 4.3.13 Illustration of DNase I digestion assay by magnetic tweezers using multiplex detection algorithm.

600 nM of Lsr2 protein was introduced in 10 mM Tris-HCl, 50 mM KCl, pH 7.5, in the absence of $MgCl_2$ (Fig. 4.3.14A) or in the presence of 10 mM $MgCl_2$ (Fig. 4.3.14B), to allow formation of the rigid Lsr2 filament on extended DNA by 15 minutes incubation. Then, 200 nM DNase I in 10 mM Tris-HCl, 50 mM KCl, pH 7.5 buffer conditions (without or with 10 mM $MgCl_2$, respectively) was introduced, and the rate of DNA digestion was monitored. Figure 4.3.14A shows

a typical experiment in the absence of $MgCl_2$, where only 10 percent of Lsr2 nucleoprotein filament tethers were digested within 10 minutes; while in the case of unprotected DNA in the same buffer condition, all tethers were cleaved within two minutes. Similar results were also observed in the presence of 10 mM $MgCl_2$ (Fig. 4.3.14B). The slight increase in the digestion rate in 10 mM $MgCl_2$ could be explained by increased activity of DNase I in the presence of $MgCl_2$. Such experiments were repeated at least three times in each buffer condition with similar results (Fig. 4.3.15).

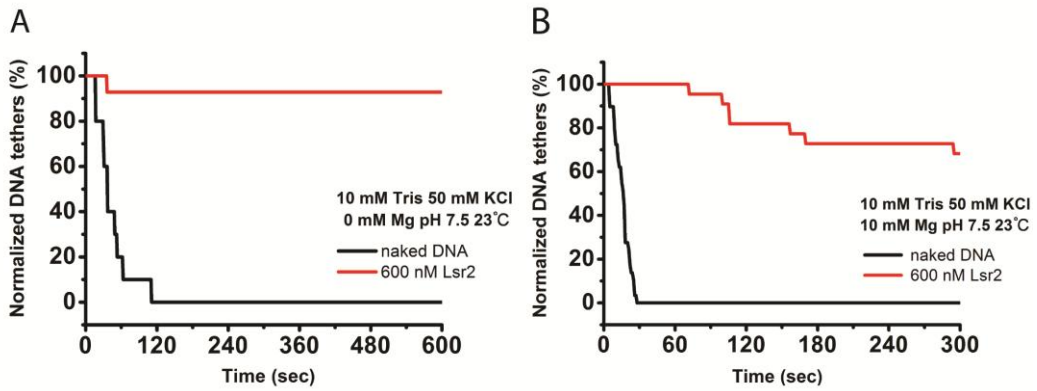


Figure 4.3.14 DNase I digestion assays of DNA accessibility restriction by rigid Lsr2-DNA complex formed on extended DNA. (A) Normalized survival DNA tethers of 10 unprotected naked DNA (black) by 200 nM DNase I in 50 mM KCl or 14 pre-formed rigid Lsr2-DNA complex tethers with 600 nM Lsr2 (red) by 200 nM DNase I in 50 mM KCl. (B) Normalized survival DNA tethers of 22 unprotected naked DNA (black) by 200 nM DNase I in 50 mM KCl and 10 mM $MgCl_2$ or 29 pre-formed rigid Lsr2-DNA complex tethers with 600 nM Lsr2 (red) by 200 nM DNase I in 50 mM KCl and 10 mM $MgCl_2$. In both reaction buffer conditions, compared with the unprotected naked DNAs, formation of the rigid Lsr2-DNA complex on extended DNA drastically slows down the DNA digestion rate.

Overall, these results indicate that extended DNA covered by rigid Lsr2 filament is sufficient to strongly restrict DNA access by DNase I. As DNase I digestion of DNA only requires access to 6 bp of exposed DNA, these results imply that DNA covered by Lsr2 filament should be able to block the access to DNA by RNA polymerase, which requires ~70 bp of exposed DNA (156).

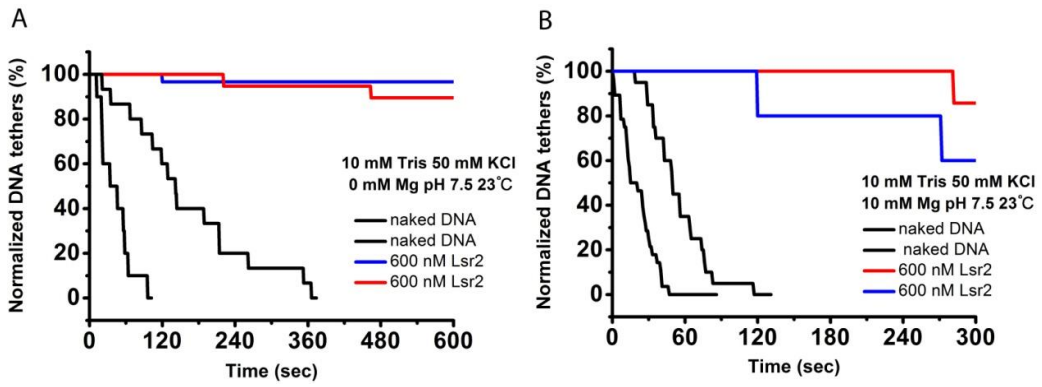


Figure 4.3.15 Multiplex single-DNA DNase I digestion assays of rigid Lsr2-DNA complexes. (A) Naked DNA tethers were rapidly digested by DNase I in 50 mM KCl; while rigid Lsr2-DNA complexes formed in 600 nM Lsr2 dramatically reduced the digestion rate in the same buffer condition. (B) Similar results were obtained in 50 mM KCl and 10 mM $MgCl_2$. Two independent experiments were repeated under each experimental condition, as indicated in figure legends in (A) and (B).

4.4 Discussion

4.4.1 Structural implication of cooperative Lsr2 binding on extended DNA

This work shows *M. tuberculosis* Lsr2 protein cooperatively binds to extended DNA resulting in a rigid Lsr2-DNA complex that has a much higher bending stiffness than a typical B-form DNA. A cooperative binding on an extended linear DNA track implies formation of a nucleoprotein filament. Results in Figure 4.3.2 also showed that the fully covered rigid Lsr2 nucleoprotein filament has similar contour length to that of a naked DNA, suggesting that the formation of Lsr2 filament on extended DNA does not cause significant distortion of the DNA backbone. Based on these observations, we propose that Lsr2 wraps around DNA and buries DNA inside. This also explains why DNA covered by Lsr2 can drastically restrict DNA accessibility. All these observations have been found in our previous studies of H-NS family proteins in gram-negative bacteria (79,114), further highlighting the universality and potential physiological importance of such nucleoprotein filamentous structures. Here we want to emphasize that the cooperative nucleoprotein filament formation by Lsr2 is not artificially triggered by a forced extended DNA conformation, as on short length scale comparable to the persistence length of DNA (~ 50 nm or 150 bp), DNA is always locally extended in the absence of tensile force due to the intrinsic DNA bending rigidity (157).

4.4.2 Mechanism of Lsr2 mediated physical DNA organization

Other than the DNA-stiffening effect of Lsr2 at high DNA tension, we also showed that Lsr2 caused DNA folding at low DNA tension which was complemented with the observed aggregation of Lsr2-DNA complexes in our AFM imaging experiments where no tension was applied to the DNA molecules.

This is consistent with previous AFM studies that showed Lsr2 can bridge DNA (122). As such, we see that DNA tension can regulate the observed DNA-binding properties of Lsr2; high DNA tension (high DNA stretching force) favors the DNA-stiffening mode while low DNA tension favors DNA-folding. In the cell, many DNA processing motors such as DNA or RNA polymerases can exert force on DNA up to 30 pN (3,54,158) while the occasional interaction between nucleoid and cell membrane may also impose a certain mechanical constraint on the chromosomal DNA (159). This suggests the chromosomal DNA is always under a variation of tension and mechanical stress which potentially affect protein-DNA interactions and thus bring forth the physiological relevancy in how Lsr2 organizes DNA under force constraint.

Although it is likely that high DNA tension (> 10 pN) applied during protein introduction prevents the DNA from Lsr2 folding, we cannot exclude the possibility that DNA tension and thus a more extended DNA conformation might preferentially select binding of an Lsr2 species that stiffens DNA. A relaxed DNA conformation at low DNA tension may then allow dominant binding by another Lsr2 species that causes DNA-folding. This scenario is possible given the solution oligomeric nature of Lsr2 (122,128) and can also explain the observation of tension regulated DNA-stiffening and –folding by Lsr2.

In addition, as revealed in the previous studies on StpA and MvaT (79,114), DNA-folding and DNA-stiffening do not necessarily exclude each other. In fact, DNA-folding can be mediated by DNA-binding proteins in numerous pathways. The StpA nucleoprotein filaments, for an example, can mediate DNA-bridging when this filament meets another naked DNA segment in low $MgCl_2$ concentration, while higher-ordered DNA organization occurs at higher $MgCl_2$ concentration through interactions between StpA nucleoprotein filaments (79). Similarly, the MvaT nucleoprotein filaments were able to mediate complex higher-ordered DNA organization (114). These similarities suggest that DNA-folding by Lsr2 maybe also be mediated by locally-formed Lsr2 nucleoprotein filaments which interact with each other. Future studies using systematic Lsr2

mutations to isolate its individual DNA organization mechanism will help to address this question.

4.4.3 Implication of Lsr2 DNA-binding properties in its physiological functions

The results from this work provide a platform to study how Lsr2 may perform its functions *in vivo*. Lsr2 DNA-folding ability suggests that it may potentially play an important role in DNA organization in *M. tuberculosis*. This is seen in many cases of DNA-folding bacterial NAPs that are involved in chromosomal DNA packaging (160-162). In particular, *E. coli* H-NS was shown highly localized in *E. coli* chromosome (163,164), and deletion of H-NS results in global reorganization of the *E. coli* chromosome DNA (163). Lsr2 prefers binding to AT-rich DNA sequences and its binding sites are correlated with low CG-content segments of the genomic DNA (124). We also performed single-DNA stretching experiments on truncated fragments of λ-DNA; a 19 kb 57 % CG-rich fragment and a 15 kb 54 % AT-rich fragment and found no change in Lsr2 DNA-binding modes on either fragments (Fig. 4.4.1). We still observed DNA-stiffening effects of Lsr2 whether with high CG- or AT-rich fragments while in both cases, Lsr2 DNA-folding was induced at low force. This suggests DNA sequence has no significant effect on how Lsr2 organizes DNA but rather tunes its DNA-binding affinity as previously shown (124).

In addition, the resistance of the rigid Lsr2-DNA complexes to salt and pH within physiological range suggests Lsr2 may not take part in gene regulatory regions that are pH and salt sensitive. This is contrary to gram-negative H-NS, which was shown to be sensitive to salt and pH changes (71,75) and involved in the regulation of salt sensitive proU operon (165). The differential response of Lsr2 and H-NS to pH might be due to the fact that Lsr2 is more basic than H-NS with a predicted pI of 10.69 compared to 5.25 for H-NS, therefore is more likely

to bind negatively charged DNA at the range of pH tested (pH 6.8-8.8). On the other hand, the rigid Lsr2-DNA complexes are shown to be more sensitive to temperature than salt or pH as indicated by a drop in Lsr2 DNA-stiffening effect at 37 ℃.

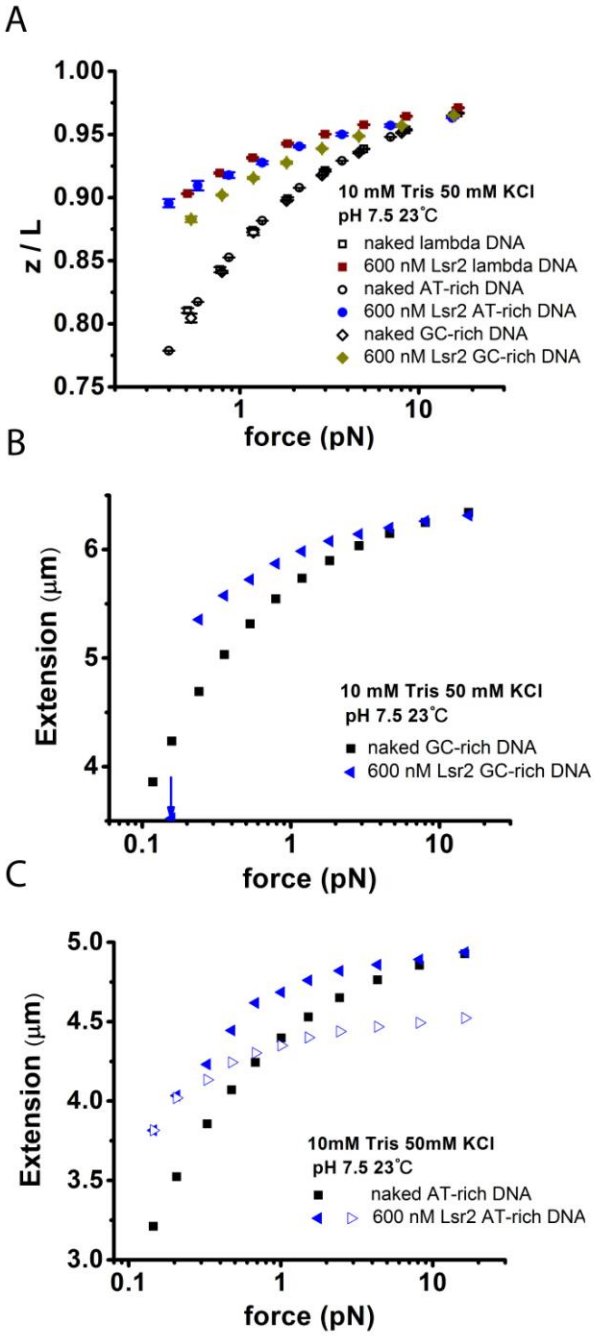


Figure 4.4.1 Lsr2-DNA nucleoprotein complex formation on 19,327 bp GC-rich DNA (GC = 57 %) and 15,003 bp AT-rich DNA (AT = 54 %). (A) Force-jump force-extension curves shows Lsr2 cause DNA stiffening on both GC-rich and AT-rich DNA. The level

of stiffness is reduced on the GC-rich DNA, which may due to the reduction of binding affinity of Lsr2 on GC-rich DNA since it preferentially bind to AT-rich DNA. (B-C) Lsr2-DNA nucleoprotein complex condenses under low forces for both GC-rich sequence (B) and AT-rich sequence (C). The solid left triangles represent the force decrease scan and the open right triangles represent the force increase scan. The blue arrow in (B) indicates that the DNA is folded to the edge that cannot be unfolded and the extension was unable to measure, therefore, no force increase curve was shown. The GC-rich construct was digested from λ-DNA (1–19,327 bp of the λ-DNA), and the AT-rich construct was also digested from λ-DNA (33,499–48,502 bp of the λ-DNA). For the preparation of the truncated DNA, please refer to (10) for more details.

This shows certain similarity to H-NS which is also a temperature sensor (71,75) although H-NS has a much more drastic response to temperature. This suggests that Lsr2 may potentially be involved in regulating operons that are temperature sensitive. Transcription of Lsr2 was also found to be up-regulated at high temperatures (166).

The emerging discovery of bacterial NAPs (H-NS, StpA, MvaT and Dan) nucleoprotein filament structures is intriguing (71,79,114,167). Strikingly, all of these proteins are known to involve in regulating DNA transcriptions, mainly repressive actions. For example, H-NS is a known global gene silencer (96), StpA represses RpoS (sigma 38) regulon (168) and loss of MvaT expression resulted in higher expression of Pf4 genes (169). As all of these proteins form similar rigid nucleoprotein filaments, it suggests that such nucleoprotein filaments may play an important role in repressing gene expressions. Given the numerous similarities between Lsr2 and H-NS family proteins in gram-negative bacteria as revealed from this work, the proposed rigid Lsr2 nucleoprotein filament has the potential to be the structural basis for its gene silencing function.

CHAPTER 5 Conclusions

My thesis has described results from two main research projects that I have completed during my Ph.D. studies: 1) the stability and kinetics of transitions between four DNA structures induced by large DNA tension; and 2) the mechanism of DNA organization by *Mycobacterium tuberculosis* protein Lsr2. These results are summarized in Chapter 3 and 4, respectively.

My results on the DNA structural transition under tension show that the three possible overstretched structures, namely, the 1ssDNA structure from the force-induced strand-peeling transition, the 2ssDNA from the force-induced internal melting transition, and the base-paired S-DNA from the B-to-S transition, can all exist as thermodynamically stable structures at large tension under appropriate conditions. Conditions that affect the relative stability between these structures include the DNA sequences, temperature, salt, and DNA topologies. Besides determining the selection of the transition pathways from B-DNA during overstretching transition, the relative stabilities of the overstretched DNA structures also determine the transitions from one overstretched structure to another by tuning environmental factors.

In general, my theoretical analysis has shown that for DNA with open ends, free from extremely heterogenic sequence distribution, such as an extremely AT-rich region sandwiched between two extremely GC-rich regions, the force-induced internal melting transition is always disfavored compared to the strand-peeling and the B-to-S transitions in physiological salt and temperature conditions. The selection between the strand-peeling and the B-to-S transitions is very sensitive to any changes of factors affecting DNA base pair stability. Generally speaking, if a factor is changed toward increasing the base pair stability, it will favor the selection of the B-to-S transition over the strand-peeling transition.

On topologically closed, nick-free DNA, the strand-peeling transition is prohibited; therefore the internal melting can occur. Due to the high-energy cost

to create internal bubbles, the B-to-S transition is always favored on DNA with normal sequence distribution in physiological salt and temperature ranges. The internal melting is selected only under extreme conditions, such as in < 2 mM NaCl, or under high temperature, or for extremely AT-rich DNA sequences. These conclusions are summarized in Figures 3.3.17.

It should be noted that the above discussion of the selection of the transitions is based on the equilibrium statistics, which is solely determined by the stability of the overstretched DNA. However, the actual transition type observed in an experiment also depends on the kinetics of the transition. The B-to-S transition is found to be a fast and highly cooperative process while the peeling transition is slow, requesting overcoming sequence-dependent energy barriers. Therefore, even under the condition that favors peeling transition, B-to-S transition can occur first under high rate of pulling, and then followed by peeling from the S-DNA to 1ssDNA. Although it was not discussed in my thesis work, such phenomenon was observed in experiments.

Because of the differential stabilities of the overstretched DNA structures (S-DNA, 1ssDNA and 2ssDNA), inter-conversion from one to another may occur. Such transition from one overstretched DNA structure to another was not experimentally investigated before our ongoing experiments. The inter-conversion between S-DNA and 2ssDNA is theoretically predicted for a topologically end-closed DNA by switching the buffer between high salt and low salt condition under large tension, which was observed in our experiments. Importantly, transition from 2ssDNA to S-DNA indicates that S-DNA can be a thermodynamic stable structure under physiological solution conditions. In addition, I predicted conditions where the transition between S-DNA and 1ssDNA may occur. The prediction for the transition from S-DNA to 1ssDNA was confirmed in our experiments, while those for the reverse transition from 1ssDNA to S-DNA were not consistent with experiments. We reason the difficulty of 1ssDNA to S-DNA transition is mainly due to the formation of the secondary structures on the partially peeled ssDNA that was not under tension.

Besides the stability analysis of the DNA structures, I have been particularly interested in the possible structures of the mysterious S-DNA. To gain some insights to the S-DNA structure, I developed a novel quasi-equilibrium full-atom steered molecular dynamics simulation, which revealed an elongated DNA structure with all base pairs maintained but significantly untwisted. This structure agrees with all known experimental results; therefore it is a highly possible candidate for the S-DNA (Fig.3.3.19).

Not only force can change the DNA structure, the DNA conformation can also be altered when binding with NAPs. This provides a physical basis for NAPs to perform their biological functions, such as DNA organization and gene regulation. In our study, we found that the gram-positive H-NS paralogue Lsr2, from *Mycobacterium tuberculosis* (actinobacteria phylum group), is able to form rigid nucleoprotein filament along DNA through a highly cooperative process similar to H-NS family proteins found in gram-negative bacteria (Fig. 4.3.1 & Fig. 4.3.3). This finding is the first evidence showing that the nucleoprotein filament formation capability is conserved for H-NS family proteins across different phylum groups (proteobacteria and actinobacteria). In addition, Lsr2-induced DNA-folding was also observed (Fig. 4.3.5 & Fig. 4.3.6), which may be involved in mediating physical DNA organization. Moreover, compared to the H-NS proteins form *E. coli*, Lsr2 nucleoprotein filament is found to be more resistant to environmental factor (salt, pH and temperature) (section 4.3.3). Despite of its weakly dependence on KCl concentration and temperature, it is insensitive to changes of magnesium concentration and pH values.

We also demonstrated that Lsr2 nucleoprotein filament formation is able to strongly restrict DNA accessibility across a single large piece of DNA fragment (Fig. 4.3.13 & Fig. 4.3.14), a property that is shown in H-NS *E. coli* paralog, StpA. This suggests that Lsr2 filament may restrict the RNA polymerase access to DNA, hence blocking the transcription process. Overall, this finding is consistent with the gene-silencing function of the Lsr2 protein.

BIBLIOGRAPHY

1. Fredrickson, J.K., Zachara, J.M., Balkwill, D.L., Kennedy, D., Li, S.M., Kostandarithes, H.M., Daly, M.J., Romine, M.F. and Brockman, F.J. (2004) Geomicrobiology of high-level nuclear waste-contaminated vadose sediments at the hanford site, washington state. *Applied and environmental microbiology*, **70**, 4230-4241.
2. Thanbichler, M., Wang, S.C. and Shapiro, L. (2005) The bacterial nucleoid: a highly organized and dynamic structure. *Journal of cellular biochemistry*, **96**, 506-521.
3. Wang, M.D., Schnitzer, M.J., Yin, H., Landick, R., Gelles, J. and Block, S.M. (1998) Force and Velocity Measured for Single Molecules of RNA Polymerase. *Science (New York, N.Y.)*, **282**, 902-907.
4. Watson, J.D. and Crick, F.H. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, **171**, 737-738.
5. Tomonaga, T. and Levens, D. (1996) Activating transcription from single stranded DNA. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 5830-5835.
6. Pohl, F.M. (1986) Dynamics of the B-to-Z transition in supercoiled DNA. *Proceedings of the National Academy of Sciences of the United States of America*, **83**, 4983-4987.
7. Balasubramanian, S., Hurley, L.H. and Neidle, S. (2011) Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nat Rev Drug Discov*, **10**, 261-275.
8. Smith, S.B., Cui, Y. and Bustamante, C. (1996) Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science (New York, N.Y.)*, **271**, 795-799.
9. Cluzel, P., Lebrun, A., Heller, C., Lavery, R., Viovy, J.-L., Chatenay, D. and Caron, F. (1996) DNA: An Extensible Molecule. *Science (New York, N.Y.)*, **271**, 792-794.
10. Fu, H., Chen, H., Marko, J.F. and Yan, J. (2010) Two distinct overstretched DNA states. *Nucleic acids research*, **38**, 5594-5600.
11. Fu, H., Chen, H., Zhang, X., Qu, Y., Marko, J.F. and Yan, J. (2011) Transition dynamics and selection of the distinct S-DNA and strand unpeeling modes of double helix overstretching. *Nucleic acids research*, **39**, 3473-3481.
12. Zhang, X., Chen, H., Fu, H., Doyle, P.S. and Yan, J. (2012) Two distinct overstretched DNA structures revealed by single-molecule thermodynamics measurements. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 8103-8108.
13. Zhang, X., Chen, H., Le, S., Rouzina, I., Doyle, P.S. and Yan, J. (2013) Revealing the competition between peeled ssDNA, melting bubbles, and S-DNA during DNA overstretching by single-molecule calorimetry. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 3865-3870.
14. Paik, D.H. and Perkins, T.T. (2011) Overstretching DNA at 65 pN Does Not Require Peeling from Free Ends or Nicks. *Journal of the American Chemical Society*, **133**, 3219-3221.

15. Bosaeus, N., El-Sagheer, A.H., Brown, T., Smith, S.B., Åkerman, B., Bustamante, C. and Nordén, B. (2012) Tension induces a base-paired overstretched DNA conformation. *Proceedings of the National Academy of Sciences*, **109**, 15179-15184.
16. Devoe, H. and Tinoco, I., Jr. (1962) The stability of helical polynucleotides: base contributions. *Journal of molecular biology*, **4**, 500-517.
17. Šponer, J., Leszczyński, J. and Hobza, P. (1996) Nature of Nucleic Acid−Base Stacking: Nonempirical ab Initio and Empirical Potential Characterization of 10 Stacked Base Dimers. Comparison of Stacked and H-Bonded Base Pairs. *The Journal of Physical Chemistry*, **100**, 5590-5596.
18. Yakovchuk, P., Protozanova, E. and Frank-Kamenetskii, M.D. (2006) Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic acids research*, **34**, 564-574.
19. SantaLucia, J., Allawi, H.T. and Seneviratne, P.A. (1996) Improved Nearest-Neighbor Parameters for Predicting DNA Duplex Stability†. *Biochemistry*, **35**, 3555-3562.
20. SantaLucia, J., Jr. (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc Natl Acad Sci U S A*, **95**, 1460-1465.
21. Gray, D.M. and Tinoco, I. (1970) A new approach to the study of sequence-dependent properties of polynucleotides. *Biopolymers*, **9**, 223-244.
22. Uhlenbeck, O.C., Borer, P.N., Dengler, B. and Tinoco Jr, I. (1973) Stability of RNA hairpin loops: A6-Cm-U6. *Journal of molecular biology*, **73**, 483-496.
23. Borer, P.N., Dengler, B., Tinoco Jr, I. and Uhlenbeck, O.C. (1974) Stability of ribonucleic acid double-stranded helices. *Journal of molecular biology*, **86**, 843-853.
24. Tinoco, I., Jr., Borer, P.N., Dengler, B., Levin, M.D., Uhlenbeck, O.C., Crothers, D.M. and Bralla, J. (1973) Improved estimation of secondary structure in ribonucleic acids. *Nature: New biology*, **246**, 40-41.
25. Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, J.C., Hutchison, C.A., Slocombe, P.M. and Smith, M. (1977) Nucleotide sequence of bacteriophage [phi]X174 DNA. *Nature*, **265**, 687-695.
26. Lyubchenko, Y.L., Vologodskii, A.V. and Frank-Kamenetskii, M.D. (1978) Direct comparison of theoretical and experimental melting profiles for RFII [Phi]X174 DNA. *Nature*, **271**, 28-31.
27. Vologodskii, A.V. and Frank-Kamenetskii, M.D. (1978) Theoretical melting profiles and denaturation maps of DNA with known sequence: fd DNA. *Nucleic acids research*, **5**, 2547-2556.
28. Vizard, D.L., White, R.A. and Ansevin, A.T. (1978) Comparison of theory to experiment for DNA thermal denaturation. *Nature*, **275**, 250-251.
29. Ueno, S., Tachibana, H., Husimi, Y. and Wada, A. (1978) "Thermal stability" maps for several double-stranded DNA fragments of known sequence. *Journal of biochemistry*, **84**, 917-924.
30. Wada, A., Ueno, S., Tachibana, H. and Husimi, Y. (1979) Stability mapping along the DNA double strand and its relation to the genetic map. *Journal of biochemistry*, **85**, 827-832.
31. Tong, B.Y. and Battersby, S.J. (1978) Melting fine structure of φX174 DNA and its fragments: A theroetical study. *Biopolymers*, **17**, 2933-2937.

32. Tong, B.Y. and Battersby, S.J. (1979) Melting curves, denaturation maps, and genetic map of φX174: Their relations and applications. *Biopolymers*, **18**, 1917-1936.
33. Gotoh, O. and Tagashira, Y. (1981) Stabilities of nearest-neighbor doublets in double-helical DNA determined by fitting calculated melting profiles to observed profiles. *Biopolymers*, **20**, 1033-1042.
34. Vologodskii, A.V., Amirikyan, B.R., Lyubchenko, Y.L. and Frank-Kamenetskii, M.D. (1984) Allowance for heterogeneous stacking in the DNA helix-coil transition theory. *Journal of biomolecular structure & dynamics*, **2**, 131-148.
35. Wartell, R.M. and Benight, A.S. (1985) Thermal denaturation of DNA molecules: A comparison of theory with experiment. *Physics Reports*, **126**, 67-107.
36. Breslauer, K.J., Frank, R., Blöcker, H. and Marky, L.A. (1986) Predicting DNA duplex stability from the base sequence. *Proceedings of the National Academy of Sciences*, **83**, 3746-3750.
37. Quartin, R.S. and Wetmur, J.G. (1989) Effect of ionic strength on the hybridization of oligodeoxynucleotides with reduced charge due to methylphosphonate linkages to unmodified oligodeoxynucleotides containing the complementary sequence. *Biochemistry*, **28**, 1040-1047.
38. Delcourt, S.G. and Blake, R.D. (1991) Stacking energies in DNA. *The Journal of biological chemistry*, **266**, 15160-15169.
39. Doktycz, M.J., Goldstein, R.F., Paner, T.M., Gallo, F.J. and Benight, A.S. (1992) Studies of DNA dumbbells. I. Melting curves of 17 DNA dumbbells with different duplex stem sequences linked by T4 endloops: Evaluation of the nearest-neighbor stacking interactions in DNA. *Biopolymers*, **32**, 849-864.
40. Sugimoto, N., Nakano, S., Yoneyama, M. and Honda, K. (1996) Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes. *Nucleic acids research*, **24**, 4501-4505.
41. Dong, F., Allawi, H.T., Anderson, T., Neri, B.P. and Lyamichev, V.I. (2001) Secondary structure prediction and structure-specific sequence analysis of single-stranded DNA. *Nucleic acids research*, **29**, 3248-3257.
42. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic acids research*, **31**, 3406-3415.
43. Iserte, J.A., Stephan, B.I., Goni, S.E., Borio, C.S., Ghiringhelli, P.D. and Lozano, M.E. (2013) Family-specific degenerate primer design: a tool to design consensus degenerated oligonucleotides. *Biotechnology research international*, **2013**, 383646.
44. Matveeva, O.V., Shabalina, S.A., Nemtsov, V.A., Tsodikov, A.D., Gesteland, R.F. and Atkins, J.F. (2003) Thermodynamic calculations and statistical correlations for oligo-probes design. *Nucleic acids research*, **31**, 4211-4217.
45. Cocco, S., Yan, J., Leger, J.F., Chatenay, D. and Marko, J.F. (2004) Overstretching and force-driven strand separation of double-helix DNA. *Phys Rev E Stat Nonlin Soft Matter Phys*, **70**, 011910.
46. Huguet, J.M., Bizarro, C.V., Forns, N., Smith, S.B., Bustamante, C. and Ritort, F. (2010) Single-molecule derivation of salt dependent base-pair free energies in DNA. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 15431-15436.

47. Smith, S.B., Finzi, L. and Bustamante, C. (1992) Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads. *Science (New York, N.Y.)*, **258**, 1122-1126.
48. Bustamante, C., Marko, J.F., Siggia, E.D. and Smith, S. (1994) Entropic elasticity of lambda-phage DNA. *Science (New York, N.Y.)*, **265**, 1599-1600.
49. Marko, J.F. and Siggia, E.D. (1995) Stretching DNA. *Macromolecules*, **28**, 8759-8770.
50. Rouzina, I. and Bloomfield, V.A. (2001) Force-induced melting of the DNA double helix 1. Thermodynamic analysis. *Biophysical journal*, **80**, 882-893.
51. Wenner, J.R., Williams, M.C., Rouzina, I. and Bloomfield, V.A. (2002) Salt dependence of the elasticity and overstretching transition of single DNA molecules. *Biophysical journal*, **82**, 3160-3169.
52. Williams, M.C., Rouzina, I. and Bloomfield, V.A. (2002) Thermodynamics of DNA Interactions from Single Molecule Stretching Experiments. *Accounts of Chemical Research*, **35**, 159-166.
53. van Mameren, J., Gross, P., Farge, G., Hooijman, P., Modesti, M., Falkenberg, M., Wuite, G.J.L. and Peterman, E.J.G. (2009) Unraveling the structure of DNA during overstretching by using multicolor, single-molecule fluorescence imaging. *Proceedings of the National Academy of Sciences*.
54. Maier, B., Bensimon, D. and Croquette, V. (2000) Replication by a single DNA polymerase of a stretched single-stranded DNA. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 12002-12007.
55. Bustamante, C., Smith, S.B., Liphardt, J. and Smith, D. (2000) Single-molecule studies of DNA mechanics. *Current Opinion in Structural Biology*, **10**, 279-285.
56. Wuite, G.J.L., Smith, S.B., Young, M., Keller, D. and Bustamante, C. (2000) Single-molecule studies of the effect of template tension on T7 DNA polymerase activity. *Nature*, **404**, 103-106.
57. Rief, M., Clausen-Schaumann, H. and Gaub, H.E. (1999) Sequence-dependent mechanics of single DNA molecules. *Nat Struct Mol Biol*, **6**, 346-349.
58. Zhang, Y., Zhou, H. and Ou-Yang, Z.C. (2001) Stretching single-stranded DNA: interplay of electrostatic, base-pairing, and base-pair stacking interactions. *Biophysical journal*, **81**, 1133-1143.
59. Montanari, A. and Mezard, M. (2001) Hairpin formation and elongation of biomolecules. *Physical review letters*, **86**, 2178-2181.
60. Rouzina, I. and Bloomfield, V.A. (2001) Force-induced melting of the DNA double helix. 2. Effect of solution conditions. *Biophysical journal*, **80**, 894-900.
61. Williams, M.C., Rouzina, I. and McCauley, M.J. (2009) Peeling back the mystery of DNA overstretching. *Proceedings of the National Academy of Sciences*, **106**, 18047-18048.
62. Romano, F., Chakraborty, D., Doye, J.P., Ouldridge, T.E. and Louis, A.A. (2013) Coarse-grained simulations of DNA overstretching. *J Chem Phys*, **138**, 085101.
63. Ishihama, Y., Schmidt, T., Rappsilber, J., Mann, M., Hartl, F.U., Kerner, M. and Frishman, D. (2008) Protein abundance profiling of the Escherichia coli cytosol. *BMC Genomics*, **9**, 102.
64. Murphy, L.D. and Zimmerman, S.B. (1997) Isolation and characterization of spermidine nucleoids from Escherichia coli. *Journal of structural biology*, **119**, 321-335.

65. Portalier, R. and Worcel, A. (1976) Association of the folded chromosome with the cell envelope of E. coli: characterization of the proteins at the DNA-membrane attachment site. *Cell*, **8**, 245-255.
66. Azam, T.A. and Ishihama, A. (1999) Twelve species of the nucleoid-associated protein from Escherichia coli. Sequence recognition specificity and DNA binding affinity. *The Journal of biological chemistry*, **274**, 33105-33113.
67. Ali Azam, T., Iwata, A., Nishimura, A., Ueda, S. and Ishihama, A. (1999) Growth phase-dependent variation in protein composition of the Escherichia coli nucleoid. *Journal of bacteriology*, **181**, 6361-6370.
68. Zhang, A.X., Rimsky, S., Reaban, M.E., Buc, H. and Belfort, M. (1996) Escherichia coli protein analogs StpA and H-NS: Regulatory loops, similar and disparate effects on nucleic acid dynamics. *Embo Journal*, **15**, 1340-1349.
69. Xiao, B., Johnson, R.C. and Marko, J.F. (2010) Modulation of HU–DNA interactions by salt concentration and applied force. *Nucleic acids research*, **38**, 6176-6185.
70. van Noort, J., Verbrugge, S., Goosen, N., Dekker, C. and Dame, R.T. (2004) Dual architectural roles of HU: formation of flexible hinges and rigid filaments. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 6969-6974.
71. Liu, Y., Chen, H., Kenney, L.J. and Yan, J. (2010) A divalent switch drives H-NS/DNA-binding conformations between stiffening and bridging modes. *Genes & development*, **24**, 339-344.
72. Dame, R.T., Wyman, C. and Goosen, N. (2000) H-NS mediated compaction of DNA visualised by atomic force microscopy. *Nucleic acids research*, **28**, 3504-3510.
73. Dame, R.T. and Wuite, G.J.L. (2003) *On the role of H-NS in the organization of bacterial chromatin: from bulk to single molecules and back*.
74. Dame, R.T., Noom, M.C. and Wuite, G.J. (2006) Bacterial chromatin organization by H-NS protein unravelled using dual DNA manipulation. *Nature*, **444**, 387-390.
75. Amit, R., Oppenheim, A.B. and Stavans, J. (2003) Increased Bending Rigidity of Single DNA Molecules by H-NS, a Temperature and Osmolarity Sensor. *Biophysical journal*, **84**, 2467-2473.
76. Tapias, A., Lopez, G. and Ayora, S. (2000) Bacillus subtilis LrpC is a sequence-independent DNA-binding and DNA-bending protein which bridges DNA. *Nucleic acids research*, **28**, 552-559.
77. Jafri, S., Evoy, S., Cho, K., Craighead, H.G. and Winans, S.C. (1999) An Lrp-type transcriptional regulator from Agrobacterium tumefaciens condenses more than 100 nucleotides of DNA into globular nucleoprotein complexes. *Journal of molecular biology*, **288**, 811-824.
78. Beloin, C., Jeusset, J., Revet, B., Mirambeau, G., Le Hegarat, F. and Le Cam, E. (2003) Contribution of DNA conformation and topology in right-handed DNA wrapping by the Bacillus subtilis LrpC protein. *The Journal of biological chemistry*, **278**, 5333-5342.
79. Lim, C.J., Whang, Y.R., Kenney, L.J. and Yan, J. (2012) Gene silencing H-NS paralogue StpA forms a rigid protein filament along DNA that blocks DNA accessibility. *Nucleic acids research*, **40**, 3316-3328.
80. Skoko, D., Yoo, D., Bai, H., Schnurr, B., Yan, J., McLeod, S.M., Marko, J.F. and Johnson, R.C. (2006) Mechanism of Chromosome Compaction and Looping by

the Escherichia coli Nucleoid Protein Fis. *Journal of molecular biology*, **364**, 777-798.
81. Pan, C.Q., Finkel, S.E., Cramton, S.E., Feng, J.A., Sigman, D.S. and Johnson, R.C. (1996) Variable structures of Fis-DNA complexes determined by flanking DNA-protein contacts. *Journal of molecular biology*, **264**, 675-695.
82. Cho, B.K., Knight, E.M., Barrett, C.L. and Palsson, B.O. (2008) Genome-wide analysis of Fis binding in Escherichia coli indicates a causative role for A-/AT-tracts. *Genome research*, **18**, 900-910.
83. Oberto, J., Nabti, S., Jooste, V., Mignot, H. and Rouviere-Yaniv, J. (2009) The HU regulon is composed of genes responding to anaerobiosis, acid stress, high osmolarity and SOS induction. *PloS one*, **4**, e4367.
84. Morales, P., Rouviere-Yaniv, J. and Dreyfus, M. (2002) The histone-like protein HU does not obstruct movement of T7 RNA polymerase in Escherichia coli cells but stimulates its activity. *Journal of bacteriology*, **184**, 1565-1570.
85. Prieto, A.I., Kahramanoglou, C., Ali, R.M., Fraser, G.M., Seshasayee, A.S. and Luscombe, N.M. (2012) Genomic analysis of DNA binding and gene regulation by homologous nucleoid-associated proteins IHF and HU in Escherichia coli K12. *Nucleic acids research*, **40**, 3524-3537.
86. Cho, B.K., Barrett, C.L., Knight, E.M., Park, Y.S. and Palsson, B.O. (2008) Genome-scale reconstruction of the Lrp regulatory network in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 19462-19467.
87. Newman, E.B. and Lin, R. (1995) Leucine-responsive regulatory protein: a global regulator of gene expression in E. coli. *Annual review of microbiology*, **49**, 747-775.
88. Braaten, B.A., Platko, J.V., van der Woude, M.W., Simons, B.H., de Graaf, F.K., Calvo, J.M. and Low, D.A. (1992) Leucine-responsive regulatory protein controls the expression of both the pap and fan pili operons in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, **89**, 4250-4254.
89. Hung, S.P., Baldi, P. and Hatfield, G.W. (2002) Global gene expression profiling in Escherichia coli K12. The effects of leucine-responsive regulatory protein. *The Journal of biological chemistry*, **277**, 40309-40323.
90. Dorman, C.J. (2009) Nucleoid-associated proteins and bacterial physiology. *Advances in applied microbiology*, **67**, 47-64.
91. Arfin, S.M., Long, A.D., Ito, E.T., Tolleri, L., Riehle, M.M., Paegle, E.S. and Hatfield, G.W. (2000) Global gene expression profiling in Escherichia coli K12. The effects of integration host factor. *The Journal of biological chemistry*, **275**, 29672-29684.
92. Kelly, A., Goldberg, M.D., Carroll, R.K., Danino, V., Hinton, J.C. and Dorman, C.J. (2004) A global role for Fis in the transcriptional control of metabolism and type III secretion in Salmonella enterica serovar Typhimurium. *Microbiology (Reading, England)*, **150**, 2037-2053.
93. Lucchini, S., Rowley, G., Goldberg, M.D., Hurd, D., Harrison, M. and Hinton, J.C. (2006) H-NS mediates the silencing of laterally acquired genes in bacteria. *PLoS pathogens*, **2**, e81.
94. Navarre, W.W., Porwollik, S., Wang, Y., McClelland, M., Rosen, H., Libby, S.J. and Fang, F.C. (2006) Selective silencing of foreign DNA with low GC content by the H-NS protein in Salmonella. *Science (New York, N.Y.)*, **313**, 236-238.

95. Dorman, C.J. (2004) H-NS: a universal regulator for a dynamic genome. *Nature reviews. Microbiology*, **2**, 391-400.
96. Dorman, C.J. (2007) H-NS, the genome sentinel. *Nature reviews. Microbiology*, **5**, 157-161.
97. Varshavsky, A.J., Nedospasov, S.A., Bakayev, V.V., Bakayeva, T.G. and Georgiev, G.P. (1977) Histone-like proteins in the purified Escherichia coli deoxyribonucleoprotein. *Nucleic acids research*, **4**, 2725-2746.
98. Dorman, C.J., Hinton, J.C. and Free, A. (1999) Domain organization and oligomerization among H-NS-like nucleoid-associated proteins in bacteria. *Trends in microbiology*, **7**, 124-128.
99. Rimsky, S. (2004) Structure of the histone-like protein H-NS and its role in regulation and genome superstructure. *Current opinion in microbiology*, **7**, 109-114.
100. Esposito, D., Petrovic, A., Harris, R., Ono, S., Eccleston, J.F., Mbabaali, A., Haq, I., Higgins, C.F., Hinton, J.C., Driscoll, P.C. *et al.* (2002) H-NS oligomerization domain structure reveals the mechanism for high order self-association of the intact protein. *Journal of molecular biology*, **324**, 841-850.
101. Shindo, H., Iwaki, T., Ieda, R., Kurumizaka, H., Ueguchi, C., Mizuno, T., Morikawa, S., Nakamura, H. and Kuboniwa, H. (1995) Solution structure of the DNA binding domain of a nucleoid-associated protein, H-NS, from Escherichia coli. *FEBS letters*, **360**, 125-131.
102. Falconi, M., Gualtieri, M.T., La Teana, A., Losso, M.A. and Pon, C.L. (1988) Proteins from the prokaryotic nucleoid: primary and quaternary structure of the 15-kD Escherichia coli DNA binding protein H-NS. *Molecular microbiology*, **2**, 323-329.
103. Smyth, C.P., Lundback, T., Renzoni, D., Siligardi, G., Beavil, R., Layton, M., Sidebotham, J.M., Hinton, J.C., Driscoll, P.C., Higgins, C.F. *et al.* (2000) Oligomerization of the chromatin-structuring protein H-NS. *Molecular microbiology*, **36**, 962-972.
104. Vanessa, B., Yinshan, Y., Emmanuel, M., Alain, C., Marie Thérèse, A., Bruno, R., Stefan, A., Sylvie, R. and Michel, K. (2003) The H-NS dimerization domain defines a new fold contributing to DNA recognition. *Nature Structural & Molecular Biology*, **10**, 212-218.
105. Schroder, O. and Wagner, R. (2002) The bacterial regulatory protein H-NS--a versatile modulator of nucleic acid structures. *Biological chemistry*, **383**, 945-960.
106. Hommais, F., Krin, E., Laurent-Winter, C., Soutourina, O., Malpertuy, A., Le Caer, J.P., Danchin, A. and Bertin, P. (2001) Large-scale monitoring of pleiotropic regulation of gene expression by the prokaryotic nucleoid-associated protein, H-NS. *Molecular microbiology*, **40**, 20-36.
107. White-Ziegler, C.A. and Davis, T.R. (2009) Genome-wide identification of H-NS-controlled, temperature-regulated genes in Escherichia coli K-12. *Journal of bacteriology*, **191**, 1106-1110.
108. Shi, X. and Bennett, G.N. (1994) Plasmids bearing hfq and the hns-like gene stpA complement hns mutants in modulating arginine decarboxylase gene expression in Escherichia coli. *Journal of bacteriology*, **176**, 6769-6775.

109. Zhang, A. and Belfort, M. (1992) Nucleotide sequence of a newly-identified Escherichia coli gene, stpA, encoding an H-NS-like protein. *Nucleic acids research*, **20**, 6735.
110. Castang, S. and Dove, S.L. (2010) High-order oligomerization is required for the function of the H-NS family member MvaT in Pseudomonas aeruginosa. *Molecular microbiology*, **78**, 916-931.
111. Goyard, S. and Bertin, P. (1997) Characterization of BpH3, an H-NS-like protein in Bordetella pertussis. *Molecular microbiology*, **24**, 815-823.
112. Tendeng, C., Badaut, C., Krin, E., Gounon, P., Ngo, S., Danchin, A., Rimsky, S. and Bertin, P. (2000) Isolation and characterization of vicH, encoding a new pleiotropic regulator in Vibrio cholerae. *Journal of bacteriology*, **182**, 2026-2032.
113. Ueguchi, C., Seto, C., Suzuki, T. and Mizuno, T. (1997) Clarification of the dimerization domain and its functional significance for the Escherichia coli nucleoid protein H-NS. *Journal of molecular biology*, **274**, 145-151.
114. Winardhi, R.S., Fu, W., Castang, S., Li, Y., Dove, S.L. and Yan, J. (2012) Higher order oligomerization is required for H-NS family member MvaT to form gene-silencing nucleoprotein filament. *Nucleic acids research*.
115. Schroder, O. and Wagner, R. (2000) The bacterial DNA-binding protein H-NS represses ribosomal RNA transcription by trapping RNA polymerase in the initiation complex. *Journal of molecular biology*, **298**, 737-748.
116. Dame, R.T., Wyman, C. and Goosen, N. (2001) Structural basis for preferential binding of H-NS to curved DNA. *Biochimie*, **83**, 231-234.
117. Dame, R.T., Wyman, C., Wurm, R., Wagner, R. and Goosen, N. (2002) Structural basis for H-NS-mediated trapping of RNA polymerase in the open initiation complex at the rrnB P1. *The Journal of biological chemistry*, **277**, 2146-2150.
118. Grainger, D.C., Hurd, D., Goldberg, M.D. and Busby, S.J. (2006) Association of nucleoid proteins with coding and non-coding segments of the Escherichia coli genome. *Nucleic acids research*, **34**, 4642-4652.
119. Lim, C.J., Lee, S.Y., Kenney, L.J. and Yan, J. (2012) Nucleoprotein filament formation is the structural basis for bacterial protein H-NS gene silencing. *Scientific reports*, **2**, 509.
120. Walthers, D., Li, Y., Liu, Y., Anand, G., Yan, J. and Kenney, L.J. (2011) Salmonella enterica response regulator SsrB relieves H-NS silencing by displacing H-NS bound in polymerization mode and directly activates transcription. *The Journal of biological chemistry*, **286**, 1895-1902.
121. Gordon, B.R., Imperial, R., Wang, L., Navarre, W.W. and Liu, J. (2008) Lsr2 of Mycobacterium represents a novel class of H-NS-like proteins. *Journal of bacteriology*, **190**, 7052-7059.
122. Chen, J.M., Ren, H., Shaw, J.E., Wang, Y.J., Li, M., Leung, A.S., Tran, V., Berbenetz, N.M., Kocincova, D., Yip, C.M. *et al.* (2008) Lsr2 of Mycobacterium tuberculosis is a DNA-bridging protein. *Nucleic acids research*, **36**, 2123-2135.
123. Gordon, B.R., Li, Y., Wang, L., Sintsova, A., van Bakel, H., Tian, S., Navarre, W.W., Xia, B. and Liu, J. (2010) Lsr2 is a nucleoid-associated protein that targets AT-rich sequences and virulence genes in Mycobacterium tuberculosis. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 5154-5159.
124. Gordon, B.R., Li, Y., Cote, A., Weirauch, M.T., Ding, P., Hughes, T.R., Navarre, W.W., Xia, B. and Liu, J. (2011) Structural basis for recognition of AT-rich DNA by

unrelated xenogeneic silencing proteins. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 10690-10695.
125. Chen, J.M., German, G.J., Alexander, D.C., Ren, H., Tan, T. and Liu, J. (2006) Roles of Lsr2 in colony morphology and biofilm formation of Mycobacterium smegmatis. *Journal of bacteriology*, **188**, 633-641.
126. Colangeli, R., Helb, D., Vilchèze, C., Hazbón, M.H., Lee, C.-G., Safi, H., Sayers, B., Sardone, I., Jones, M.B., Fleischmann, R.D. *et al.* (2007) Transcriptional Regulation of Multi-Drug Tolerance and Antibiotic-Induced Responses by the Histone-Like Protein Lsr2 in <named-content xmlns:xlink="http://www.w3.org/1999/xlink" content-type="genus-species" xlink:type="simple">M. tuberculosis</named-content>. *PLoS pathogens*, **3**, e87.
127. Kocincova, D., Singh, A.K., Beretti, J.L., Ren, H., Euphrasie, D., Liu, J., Daffe, M., Etienne, G. and Reyrat, J.M. (2008) Spontaneous transposition of IS1096 or ISMsm3 leads to glycopeptidolipid overproduction and affects surface properties in Mycobacterium smegmatis. *Tuberculosis (Edinburgh, Scotland)*, **88**, 390-398.
128. Summers, E.L., Meindl, K., Uson, I., Mitra, A.K., Radjainia, M., Colangeli, R., Alland, D. and Arcus, V.L. (2012) The structure of the oligomerization domain of Lsr2 from Mycobacterium tuberculosis reveals a mechanism for chromosome organization and protection. *PloS one*, **7**, e38542.
129. Neuman, K.C. and Nagy, A. (2008) Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nature methods*, **5**, 491-505.
130. Fernandez, J.M. and Li, H. (2004) Force-Clamp Spectroscopy Monitors the Folding Trajectory of a Single Protein. *Science (New York, N.Y.)*, **303**, 1674-1678.
131. Lee, G.U., Chrisey, L.A. and Colton, R.J. (1994) Direct measurement of the forces between complementary strands of DNA. *Science (New York, N.Y.)*, **266**, 771-773.
132. Kellermayer, M.S., Grama, L., Karsai, A., Nagy, A., Kahn, A., Datki, Z.L. and Penke, B. (2005) Reversible mechanical unzipping of amyloid beta-fibrils. *The Journal of biological chemistry*, **280**, 8464-8470.
133. Yan, J. and Marko, J.F. (2003) Effects of DNA-distorting proteins on DNA elastic response. *Physical review. E, Statistical, nonlinear, and soft matter physics*, **68**, 011905.
134. Ali, B.M.J., Amit, R., Braslavsky, I., Oppenheim, A.B., Gileadi, O. and Stavans, J. (2001) Compaction of single DNA molecules induced by binding of integration host factor (IHF). *Proceedings of the National Academy of Sciences*, **98**, 10658-10663.
135. Lin, J., Chen, H., Droge, P. and Yan, J. (2012) Physical Organization of DNA by Multiple Non-Specific DNA-Binding Modes of Integration Host Factor (IHF). *PloS one*, **7**, e49885.
136. Bennink, M.L., Schärer, O.D., Kanaar, R., Sakata-Sogawa, K., Schins, J.M., Kanger, J.S., de Grooth, B.G. and Greve, J. (1999) Single-molecule manipulation of double-stranded DNA using optical tweezers: Interaction studies of DNA with RecA and YOYO-1. *Cytometry*, **36**, 200-208.
137. Wang, H., Bash, R., Yodh, J.G., Hager, G.L., Lohr, D. and Lindsay, S.M. (2002) Glutaraldehyde modified mica: a new surface for atomic force microscopy of chromatin. *Biophysical journal*, **83**, 3619-3625.

138. Fu, H., Freedman, B.S., Lim, C.T., Heald, R. and Yan, J. (2011) Atomic force microscope imaging of chromatin assembled in Xenopus laevis egg extract. *Chromosoma*, **120**, 245-254.
139. Yan, J., Kawamura, R. and Marko, J.F. (2005) Statistics of loop formation along double helix DNAs. *Physical review. E, Statistical, nonlinear, and soft matter physics*, **71**, 061905.
140. Young, W.M. and Elcock, E.W. (1966) Monte Carlo studies of vacancy migration in binary ordered alloys: I. *Proceedings of the Physical Society*, **89**, 735.
141. Bortz, A.B., Kalos, M.H. and Lebowitz, J.L. (1975) A new algorithm for Monte Carlo simulation of Ising spin systems. *Journal of Computational Physics*, **17**, 10-18.
142. Gillespie, D.T. (1977) Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, **81**, 2340-2361.
143. Cocco, S., Marko, J.F. and Monasson, R. (2003) Slow nucleic acid unzipping kinetics from sequence-defined barriers. *The European physical journal. E, Soft matter*, **10**, 153-161.
144. Léger, J.F., Romano, G., Sarkar, A., Robert, J., Bourdieu, L., Chatenay, D. and Marko, J.F. (1999) Structural Transitions of a Twisted and Stretched DNA Molecule. *Physical review letters*, **83**, 1066-1069.
145. Maaloum, M., Beker, A.F. and Muller, P. (2011) Secondary structure of double-stranded DNA under stretching: elucidation of the stretched form. *Physical review. E, Statistical, nonlinear, and soft matter physics*, **83**, 031903.
146. Lu, X.J. and Olson, W.K. (2008) 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nature protocols*, **3**, 1213-1227.
147. Zheng, G., Lu, X.J. and Olson, W.K. (2009) Web 3DNA--a web server for the analysis, reconstruction, and visualization of three-dimensional nucleic-acid structures. *Nucleic acids research*, **37**, W240-246.
148. Hill, A.V. (1910) The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *Journal of Physiology*, **40**, iv-vii.
149. Drees, J.C., Lusetti, S.L., Chitteni-Pattu, S., Inman, R.B. and Cox, M.M. (2004) A RecA Filament Capping Mechanism for RecX Protein. *Molecular Cell*, **15**, 789-798.
150. Gruenig, M.C., Stohl, E.A., Chitteni-Pattu, S., Seifert, H.S. and Cox, M.M. (2010) Less is more: Neisseria gonorrhoeae RecX protein stimulates recombination by inhibiting RecA. *The Journal of biological chemistry*, **285**, 37188-37197.
151. Yang, S.W. and Nash, H.A. (1995) Comparison of protein binding to DNA in vivo and in vitro: defining an effective intracellular target. *The EMBO journal*, **14**, 6292-6300.
152. Sugimura, S. and Crothers, D.M. (2006) Stepwise binding and bending of DNA by Escherichia coli integration host factor. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 18510-18514.
153. Le, S., Chen, H., Cong, P., Lin, J., Dröge, P., Yan, J. (2013) Mechanosensing of DNA bending in a single specific protein-DNA complex. *Scientific reports*, **In press**.
154. Colangeli, R., Haq, A., Arcus, V.L., Summers, E., Magliozzo, R.S., McBride, A., Mitra, A.K., Radjainia, M., Khajo, A., Jacobs, W.R., Jr. *et al.* (2009) The multifunctional histone-like protein Lsr2 protects mycobacteria against reactive oxygen intermediates. *Proc Natl Acad Sci U S A*, **106**, 4414-4418.

155. Dame, R.T., Luijsterburg, M.S., Krin, E., Bertin, P.N., Wagner, R. and Wuite, G.J.L. (2005) DNA bridging: a property shared among H-NS-like proteins. *Journal of Bacteriology*, **187**, 1845-1848.
156. Talkington, C. and Pero, J. (1979) Distinctive nucleotide sequences of promoters recognized by RNA polymerase containing a phage-coded "sigma-like" protein. *Proceedings of the National Academy of Sciences of the United States of America*, **76**, 5465-5469.
157. Edwards, M.D.a.S.F. (1988) *The Theory of Polymer Dynamics*. Oxford University Press.
158. John, R., Davenport, Wuite, G.J.L., Landick, R. and Bustamante, C. (2000) Single-Molecule Study of Transcriptional Pausing and Arrest by E. coli RNA Polymerase. *Science (New York, N.Y.)*, **287**, 2497-2500.
159. Toro, E. and Shapiro, L. (2010) Bacterial chromosome organization and segregation. *Cold Spring Harbor perspectives in biology*, **2**, a000349.
160. Kim, J., Yoshimura, S.H., Hizume, K., Ohniwa, R.L., Ishihama, A. and Takeyasu, K. (2004) Fundamental structural units of the Escherichia coli nucleoid revealed by atomic force microscopy. *Nucleic Acids Res*, **32**, 1982-1992.
161. Kaidow, A., Wachi, M., Nakamura, J., Magae, J. and Nagai, K. (1995) Anucleate cell production by Escherichia coli delta hns mutant lacking a histone-like protein, H-NS. *J Bacteriol*, **177**, 3589-3592.
162. Ghatak, P., Karmakar, K., Kasetty, S. and Chatterji, D. (2011) Unveiling the role of Dps in the organization of mycobacterial nucleoid. *Plos One*, **6**, e16019.
163. Wang, W., Li, G.W., Chen, C., Xie, X.S. and Zhuang, X. (2011) Chromosome organization by a nucleoid-associated protein in live bacteria. *Science (New York, N.Y.)*, **333**, 1445-1449.
164. Azam, T.A., Hiraga, S. and Ishihama, A. (2000) Two types of localization of the DNA-binding proteins within the Escherichia coli nucleoid. *Genes to cells : devoted to molecular & cellular mechanisms*, **5**, 613-626.
165. Lucht, J.M., Dersch, P., Kempf, B. and Bremer, E. (1994) Interactions of the nucleoid-associated DNA-binding protein H-NS with the regulatory region of the osmotically controlled proU operon of Escherichia coli. *The Journal of biological chemistry*, **269**, 6578-6578.
166. Stewart, G.R., Wernisch, L., Stabler, R., Mangan, J.A., Hinds, J., Laing, K.G., Young, D.B. and Butcher, P.D. (2002) Dissection of the heat-shock response in Mycobacterium tuberculosis using mutants and microarrays. *Microbiology (Reading, England)*, **148**, 3129-3138.
167. Lim, C.J., Lee, S.Y., Teramoto, J., Ishihama, A. and Yan, J. (2012) The nucleoid-associated protein Dan organizes chromosomal DNA through rigid nucleoprotein filament formation in E. coli during anoxia. *Nucleic Acids Research*.
168. Lucchini, S., McDermott, P., Thompson, A. and Hinton, J.C. (2009) The H-NS-like protein StpA represses the RpoS (sigma 38) regulon during exponential growth of Salmonella Typhimurium. *Mol Microbiol*, **74**, 1169-1186.
169. Castang, S. and Dove, S.L. (2012) Basis for the essentiality of H-NS family members in Pseudomonas aeruginosa. *J Bacteriol*, **194**, 5101-5109.

LIST OF PUBLICATIONS

1. Fu, H., Chen, H., Zhang, X., **Qu, Y.**, Marko, J.F. and Yan, J. (2011) Transition dynamics and selection of the distinct S-DNA and strand unpeeling modes of double helix overstretching. *Nucleic acids research*, **39**, 3473-3481.

2. **Qu, Y.**, Lim, C.J., Whang, Y.R., Liu, J. and Yan, J. (2013) Mechanism of DNA organization by Mycobacterium tuberculosis protein Lsr2. *Nucleic acids research*, **41**, 5263-5272.