THERMODYNAMICS AND KINETICS
OF DNA-PROTEIN INTERACTIONS
FROM SINGLE MOLECULE FORCE SPECTROSCOPY

A dissertation presented

by

Leila Shokri

to

The Department of Physics

In partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in the field of

Physics

Northeastern University
Boston, Massachusetts
April 2008
To my parents
THERMODYNAMICS AND KINETICS
OF DNA-PROTEIN INTERACTIONS
FROM SINGLE MOLECULE FORCE SPECTROSCOPY

by

Leila Shokri

ABSTRACT OF DSSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Physics
in the Graduate School of Arts and Sciences
Northeastern University, April 2008
Abstract

The process of replication requires the cooperation of many proteins which associate with each other at the replication fork to form a highly efficient replication machine. Bacterium *E. coli* and bacteriophages that infect it (T4 and T7) have been used extensively in molecular biology research and provide excellent model systems for analyzing the DNA replication. In this work we use single DNA molecule stretching to investigate the degree of alteration in the structure and stability of DNA in the presence of DNA binding proteins which help us to quantify thermodynamics and kinetics of protein-protein and protein-nucleic acid interactions and obtain new insights into the function of proteins in these specific biological system.

Because the nature of the overstretching transition in DNA stretching experiments continues to generate controversy we have undertaken studies of DNA stretching in the presence of glyoxal to solve this dilemma which brought new evidence in favor of force-induced melting theory against the alternative S-DNA model.

One of the classic paradigms of single-stranded DNA binding proteins is bacteriophage T7 gene 2.5 protein (gp2.5), known to have essential roles in DNA replication and recombination in phage-infected cells by binding preferentially to single-stranded DNA and establishing electrostatic interactions with other proteins, recruiting them at the site of the ssDNA and regulating their activity. Varying solution conditions and the pulling dynamics, we could obtain binding affinities to single- and double-stranded DNA for gp2.5 and its deletion mutant lacking 26 C-terminal residues, gp2.5-Δ26C, over a range of salt concentration not available to ensemble studies. We also
obtained rate of cluster growth which is analogous to the growth of clusters that occurs at a replication fork as the helicase unwinds the double helix \textit{in vivo}. We proposed a model to explain proteins’ structure-function relationship, and showed that dimeric gp2.5 must dissociate prior to binding to DNA, a dissociation that consists of a weak non-electrostatic and a strong electrostatic component.

The multisubunit enzyme DNA polymerase III holoenzyme (pol III) is responsible for duplicating the \textit{E. coli} chromosomal DNA. By purifying fragments of the protein that encompass the putative OB-fold domain, and then characterizing those fragments for their DNA binding activity using single molecule DNA stretching experiments, we showed that \(\alpha\) subunit has an affinity for both double and single stranded DNA. However, our results demonstrate that the single-stranded DNA binding component appears to be passive, as the protein does not facilitate melting, but binds instead to regions already separated by force, stabilizing the single-stranded form of DNA. By studying the constructs of segments of the \(\alpha\) subunit we showed that the N-terminal region is responsible for dsDNA stabilization, while the C-terminal region binds to melted DNA suggesting that this domain may interact with ssDNA created during the DNA replication process.

Bacteriophage T4 gene 59 (gp59) protein is a replication-recombination mediator protein that stimulates the activities of the helicase enzyme by promoting its loading onto gene 32 protein (gp32)-saturated single-stranded DNA binding sites. We characterized the interactions of wild type protein gp59 and its site-directed mutant with defects in dsDNA binding (gp59\textsubscript{R12A}) with stretched DNA. We showed that gp59 binds more strongly to double-stranded DNA and determined the equilibrium binding constant to
dsDNA as a function of salt concentration. We also examined the effects of gp59 on the helix-destabilization capabilities of gp32 and its fragment, *I. Our results demonstrate that gp59 is capable of strongly destabilizing both gp32-DNA and *I-DNA interactions. This result confirms previous assumption that the local weakening of gp32-ssDNA binding is a requirement for helicase loading at that site.
Acknowledgements

It has been a great privilege to work with my advisor Prof. Mark Williams whose critical thinking, insight, and ideas have been great sources of inspiration for me. I would like to thank him for his guidance, care, and patience during the entire course of this project.

I would like to thank Prof. Penny J. Beuning from the Department of Chemistry and Chemical Biology at Northeastern University for giving me the opportunity to work in her lab and expand my theoretical knowledge and learn different techniques in molecular biology and biochemistry.

Special thanks are due to Ioulia Rouzina from University of Minnesota for her countless advises on the theoretical foundations of my research.

Prof. Charles C. Richardson and Dr. Boriana Marintcheva from Harvard Medical School have generously provided phage T7 gene 2.5 protein and its mutant for my research. I am grateful for all their assistance and also for the honor they gave me by serving in my thesis committee.

I would like to thank Prof. Andreas Hanke from the Department of Physics and Astronomy, University of Texas at Brownsville for providing theoretical support for my research; Prof. Scott W. Morrical and Amy M. Branagan from University of Vermont, College of Medicine for providing phage T4 gene 59 protein and its mutant; and Prof. Richard L. Karpel from the Department of Chemistry and Biochemistry at University of Maryland Baltimore County for providing phage T4 gene 32 protein and its mutant.

I have enjoyed the company of my lab mates, Dr. Micah J. McCauley, Dr. Jana Sefcikova, Brenna Shurtleff, Thayaparan Paramanathan, Fei Wang, Dr. Margareta
Cruceanu, Dr. Kiran Pant, Dr. Ioana D. Vlădescu, and Kathy Chaurasiya who have been great friends and with whom I have shared many memorable moments and I would like to thank them for their friendship. My special thanks also go to Dr. Micah J. McCauley and Dr. Margareta Cruceanu for their trainings and for helping me to get up to speed on the optical tweezers instrument.

I have gained a lot of confidence and experience in my research during our weekly meetings. Thanks are due to Prof. Paul Champion, Prof. J. Timothy Sage, and Prof. Armen Stepanyants for all the fruitful discussions during my presentations at these meetings.

I would like to thank Prof. Charles C. Richardson, Dr. Boriana Marintcheva, Prof. Armen Stepanyants, and Prof. Sergey Kravchenko for serving in my thesis committee.

I will forever be indebted to my parents who have been a constant source of support and encouragement. I am deeply grateful for all they have done for me and for the extraordinary efforts they put toward setting very high standards for my education. My sisters, Maryam and Farzaneh have always encouraged me and have been valuable friends to me and I heartfully thank them.

Last, but by no means least, my love, admiration and gratitude goes to my husband, Hamid Sattar without whom I have not been able to manage such an ambitious undertaking. I bestow my gratitude to him for his companionship, patience, and support throughout my PhD studies.
## Table of Contents

Abstract

Acknowledgements

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Nucleic Acid

1.1.1 Deoxyribonucleic Acid (DNA)

1.1.2 Ribonucleic Acid (RNA)

1.2 DNA Replication

1.3 Temperature Dependent DNA Melting

1.4 Polyelectrolyte Theory

1.5 Ionic Effects in DNA-Protein Interaction

1.6 McHee-von Hippel Binding Isotherm

1.7 Effects of Ligand Binding on DNA melting

1.8 Diffusion-controlled Protein-DNA Association

1.9 Single Molecule Force Spectroscopy

1.9.1 Optical Trapping of Dielectric Particles

1.9.2 Models of Polymer Elasticity

1.9.2.1 The Freely-Jointed Chain Model

1.9.2.2 The Worm-Like Chain Model

1.9.3 Force-Induced Melting

1.9.3.1 Force-Induced Melting Model (FIM)

1.9.3.2 Thermodynamics of Force-Induced DNA Melting

1.10 Experimental Procedures

1.10.1 Instrument Setup

1.10.2 DNA Stretching

1.10.3 Protein Preparation and Purification

1.10.3.1 Phage T7 Gene 2.5 Protein

1.10.3.2 α Subunit of E. coli Pol III

1.10.3.3 Phage T4 Gene 59 Protein

1.10.3.4 Phage T4 Gene 32 Protein

1.10.3.5 Glycerol

1.11 Chapter Overview
Chapter 2 ......................................................................................................................... 16
Structural Evidence of Force-Induced DNA Melting......................................................... 43
  2.1 Background............................................................................................................. 43
  2.2 Direct Evidence of the Melting Nature of the Overstretching Transition .......... 49
  2.3 Effect of Solution Conditions on the \( f_a \)............................................................ 55
  2.4 Conclusions........................................................................................................... 59

Chapter 3 ......................................................................................................................... 61
Salt Dependent Binding of Phage T7 Gene 2.5 Protein to Single-stranded DNA .... 61
  3.1 Background........................................................................................................... 61
  3.2 T7 gp2.5 Exhibits Behavior Similar to T4 gp32.................................................. 64
  3.3 Quantifying ssDNA Binding Affinity................................................................. 66
  3.4 Salt-Dependant Binding of T7 gp2.5 and T7 gp2.5-\( \Delta 26C \) to ssDNA .......... 70
  3.5 Conclusions......................................................................................................... 72

Chapter 4 ......................................................................................................................... 80
Kinetics and Thermodynamics of Salt-Dependent T7 gene 2.5 Protein Binding to Single- and Double-stranded DNA ......................................................... 80
  4.1 Background........................................................................................................... 80
  4.2 Rate dependence of DNA melting force in the presence of gp2.5....................... 82
  4.3 Conclusions.......................................................................................................... 92

Chapter 5 ......................................................................................................................... 96
Double- and Single-stranded DNA Binding of the Catalytic \( \alpha \) Subunit of \textit{E. coli} Replicative DNA Polymerase III ............................................................ 96
  5.1 Background........................................................................................................... 96
  5.2 Homology Modeling of the OB-fold Domain of \textit{E. coli} Pol III \( \alpha \) Subunit...... 99
  5.3 Pol III \( \alpha \) Subunit Binds to Force-melted DNA ................................................. 101
  5.4 Pol III \( \alpha \) Subunit Stabilizes dsDNA................................................................. 104
  5.5 N-Terminal Construct \( \alpha 1-917 \) Binds to dsDNA ............................................. 105
  5.6 C-Terminal Constructs \( \alpha 917-1160 \) and \( \alpha 978-1160 \) Bind to ssDNA...... 106
  5.7 Conclusions......................................................................................................... 107

Chapter 6 ....................................................................................................................... 110
Interaction of the Bacteriophage T4 Gene 59 Protein with DNA in the Presence of Gene 32 Protein ............................................................................................................. 110
  6.1 Background........................................................................................................... 110
  6.2 T4 Gene 59 Protein Interactions with DNA ....................................................... 114
  6.3 Destabilization of gp32-ssDNA Interactions by gp59....................................... 118
  6.4 Conclusions......................................................................................................... 123
<table>
<thead>
<tr>
<th>Chapter 7</th>
<th>Conclusions</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>124</td>
<td>124</td>
<td>130</td>
</tr>
</tbody>
</table>
List of Figures

Figure 1.1. Structures of nucleic acid constituents ......................................................... 2
Figure 1.2. Conformations of ribose and deoxyribose found in nucleotides ...................... 3
Figure 1.3. The structure of B-form DNA double helix ................................................... 4
Figure 1.4. The DNA replication fork.............................................................................. 6
Figure 1.5. Ion condensation on highly charged polymer.................................................. 9
Figure 1.6. Schematic diagram of the force of an incident light on a dielectric sphere... 21
Figure 1.7. The freely jointed chain model...................................................................... 23
Figure 1.8. The wormlike chain model.......................................................................... 25
Figure 1.9. Typical force-extension data for double-stranded DNA .............................. 28
Figure 1.10. Schematic diagram of force-induced melting model................................. 30
Figure 1.11. Schematic diagram of optical tweezers instrument ..................................... 34
Figure 2.1. Schematic diagram of the effect of glyoxal on DNA (FIM and S-DNA) .... 47
Figure 2.2. Effect of glyoxal on force-extension data.................................................... 50
Figure 2.3. Mapping DNA force-induced melting in the presence of glyoxal ............. 56
Figure 3.1. The bacteriophage T7 replication fork ......................................................... 62
Figure 3.2. Proposed model for the functional role of gp2.5 acidic C-terminus .......... 63
Figure 3.3. Single-stranded DNA binding affinity of gp2.5 and gp2.5-Δ26C .................. 67
Figure 3.4. Time-dependent equilibrium melting force at constant position............. 68
Figure 3.5. Salt-dependant equilibrium melting force ............................................... 70
Figure 3.6. Salt-dependant equilibrium binding constant to ssDNA (K_{ss}) .......... 71
Figure 3.7. Proposed model for gp2.5 dimer dissociation upon DNA binding .......... 73
Figure 3.8. Salt-dependant free energy of dimerization (ssDNA binding) ............... 75
Figure 4.1. Rate-dependant DNA stretching cycle ...................................................... 83
Figure 4.2. Non-equilibrium DNA melting force as a function of pulling rate .......... 84
Figure 4.3. Protein association rate as a function of protein concentration ............... 85
Figure 4.4. Salt-dependant equilibrium binding constant to ssDNA and dsDNA ........... 86
Figure 4.5. Salt-dependant free energy of dimerization (ssDNA and dsDNA binding) .. 88
Figure 5.1. Catalytic α subunit of E. coli replicative DNA polymerase III .................. 100
Figure 5.2. DNA stretching cycle in the presence of 100 nM full-length α protein..... 102
Figure 5.3. Stabilization of the double helix in low salt ........................................ 104
Figure 5.4. DNA binding of N-terminal α construct ............................................... 105
Figure 5.5. DNA binding of C-terminal α constructs ............................................... 106
Figure 6.1. Double-stranded DNA binding affinity of gp59 and gp59R12A .............. 115
Figure 6.2. The effects of gp59 on single-stranded DNA binding of gp32 and *I .... 119
Figure 6.3. The effects of gp59 on gp32 and *I equilibrium melting force .......... 120
Figure 6.4. gp59 can compete with gp32 in binding to ssDNA .............................. 121
List of Tables

Table 2.1. Permanently glyoxal-modified ssDNA fraction.............................................. 53

Table 6.1. Equilibrium binding constants of gp59 and gp59_{R12A} to dsDNA .................. 117
Chapter 1

Introduction

1.1 Nucleic Acid

Nucleic acids are biological macromolecules composed of strings of monomeric nucleotides found in all living cells and viruses. The most common nucleic acids are deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). Nucleic acids control cellular functions and are the hereditary materials carrying genetic information from one generation to the next. They store and replicate hereditary information, and express them through protein synthesis. Although RNA is considered likely to be the first genetic material very early in evolutionary history and in some viruses\(^1\), the genes of all modern cells and many viruses are made of DNA.

1.1.1 Deoxyribonucleic Acid (DNA)

DNA is a polymer of deoxyribonucleotide subunits. It consists of two polynucleotide strands twisted into a double helix and joined by hydrogen bonds between the complementary bases, and stacking interactions between neighboring base pairs. Nucleotides are the building blocks of deoxyribonucleic acid (DNA). They are made of a heterocyclic base, a sugar (2'-deoxyribose), and a phosphate group. The unit containing just sugar and base is called a nucleoside. Nucleosides are uncharged at physiological pH.
In contrast all of the phosphate residues in nucleotides are negatively charged. This means that in a polynucleotide chain, there is one negative charge for each residue. Figure 1.1 shows the structures and names of the common bases and sugars.

![Figure 1.1. Structures of nucleic acid constituents](image)

Because of their nitrogen content, these bases are called nitrogenous bases, and are further classified as either purines or pyrimidines. Purines and pyrimidines are two-ringed and one-ringed structures composed of carbon and nitrogen atoms, respectively. The two purine bases found in DNA are guanine (G) and adenine (A), while the pyrimidine bases are cytosine (C) and thymine (T). DNA polymers are constructed by forming phosphodiester bonds between nucleotides. In this arrangement, a phosphate group acts as a bridge between the 5’ position of one sugar and the 3’ position of the next. Guanine nucleotides are base-paired opposite cytosine nucleotides, while Adenine nucleotides are base-paired opposite thymine nucleotides. The GC base pair is stronger than the AT base pairs, because AT forms two hydrogen bonds while GC forms three. These weak
attractions between bases are one of the two main factors responsible for the stability of the DNA double helix. The other factor is the base-stacking interactions between the \( \pi \) orbits of the adjacent bases’ aromatic rings. However, it has been shown that base stacking is the main stabilizing factor in the duplex DNA\(^1\). The stability of DNA is also controlled by the screening of the negative charges of the phosphate ions on the backbone by ions in solution.

The 5 atoms of the sugar ring are never seen to be planar, as this is sterically and energetically very unfavorable. This phenomenon is called sugar puckering and it has two possibilities (Figure 1.2). The first is an envelope (E) form, in which 4 atoms are in a plane and the fifth is out by 0.5 Å. The second is a twist (T) form, in which two adjacent atoms are out of the plane made by the other three atoms. Atoms on the same side as the 5'-C are called \textit{endo} and those on the opposite side are called \textit{exo}\(^2,6\).

B-DNA is the most common form of DNA \textit{in vivo}. The double helix in its canonical B-form is a right-handed spiral formed by two individual DNA strands aligned in an antiparallel fashion. In this form the sugar configuration is known as C2'-\textit{endo} (Figure 1.2). One helical turn of B-DNA contains about 10.5 base pairs and the distance between the bases is 3.4 Å. Base pairs are almost perpendicular to the helical axis and are buried inside the helix. There are two grooves (major and minor) twisting around the surface of
The minor groove is the side of the base pair where the sugars are attached. The sequence of nucleotides determines the genetic code that is used in the development and functioning of all living organisms. These genetic codes can be read when the two strands of the double helix are separated by thermal fluctuations or by specific proteins that uncover the hydrophobic bases. Regulating proteins are able to control how these codes are read in order to create proteins (transcription and translation), or to make more copies of DNA (replication). The structure of the B-form DNA double helix is shown in Figure 1.3.

1.1.2 Ribonucleic Acid (RNA)

RNA is a linear, usually single-stranded chain of ribonucleotides. Ribonucleotides are the building blocks of RNA. They consist of a nitrogen-containing base, a phosphate group, and ribose. The sugars are all ribose and four nucleotides are adenine (A), guanine (G), cytosine (C), and uracil (U)\(^2,3\) (Figure 1.1). The presence of the 2' hydroxyl alters the conformation of the ribose ring and eliminates the formation of B helix. Also, it makes RNA more prone to hydrolysis resulting in a less stable structure compared to DNA.
Its function in the cell is related to protein synthesis. RNA is a direct template for protein synthesis. During protein synthesis, a class of RNA molecules (messenger RNA or mRNA) brings genetic information from a cell’s chromosomes to the ribosomes. Other forms of RNA help ribosomes translate stored genetic information into protein molecules. Ribosomes are small particles composed of ribosomal RNA (rRNA) and proteins. Each successive mRNA’s codon (trinucleotide that specifies an amino acid) is recognized by an anticodon of a transfer RNA (tRNA) through complementary base pairing.

1.2 DNA Replication

Genes are the small heredity units in cells. Before a cell divides, its DNA must be replicated so that the two new cells inherit the genetic code. The process of DNA replication is the basis for biological inheritance in all living organisms. The two strands of a DNA molecule have complementary base pairs and hold the same genetic information so when they are separated, each of them can serve as a template to produce a complementary strand and form a new DNA double helix, identical to the original. This process is called *semiconservative replication* because each daughter double helix contains one parental and one newly synthesized strand. Proofreading mechanism of the newly formed double helix ensures near perfect fidelity and keeps DNA sequences accurate. DNA synthesis requires the cooperation of many proteins to form a highly efficient replication machine.

DNA replication begins from specific nucleotide sequences in the genome, called *origins* and proceeds bidirectionally with two forks moving outward in opposite directions. The replication fork is a Y shape structure which is created through the action
of a helicase. Helicases unwind the helix by breaking the interactions holding the two DNA strands together. Single-stranded DNA-binding proteins help in opening up the DNA helix so that it can be copied. DNA polymerase then moves along the exposed DNA strand in the 5'-to-3' direction, using it as a template for assembling a new DNA strand, complementary to the template, with newly arrived deoxyribonucleotides. Although the substrates for DNA polymerase are the triphosphate forms of the deoxyribonucleotides, polymerase can only add new nucleotides to the 3' end of an existing strand of DNA. Therefore, a primase is required to synthesize a short RNA primer, to which new nucleotides can be added by DNA polymerase. Because of the antiparallel orientation of the two new DNA strands, in one strand (leading strand), polymerase is able to synthesize DNA continuously and with high processivity, using the free 3'-OH group donated by a single RNA primer, in the same 5'-to-3' direction in which the replication fork is moving. Synthesis of the complementary strand, called the lagging strand, which runs in 3'-to-5' direction, is more complex because no 3'-to-5' DNA polymerase has ever been found. To provide a free 3'-
OH starting point, RNA primase builds short RNA primers. DNA polymerases are then able to use these free 3'-OH group and form short stretches in the 5'-to-3' direction away from the replication form. These short 1- to 3-kilobase (kb) stretches are called Okazaki fragments\textsuperscript{12}. The RNA fragments are then removed and the new deoxyribonucleotides are added to fill the gaps where the RNA was present. New fragments are later joined together by an enzyme called DNA ligase. Different proteins are required for DNA synthesis in leading and lagging strands. The structure of the DNA replication fork is shown in Figure 1.4. DNA polymerase also acts as a proofreading enzyme that removes its own polymerization errors as it moves along the DNA. It has a separate catalytic site, a 3'-to-5' proofreading exonuclease, which cuts off any unpaired residues at the primer terminus. Studies of the replication mechanism in all three domain of life (archaea, bacteria, and eucaryota\textsuperscript{13,14}) suggests that the biochemical mechanism of DNA replication is similar. This implies that, although eukaryotic replication machinery is more complicated than its bacterial analog, the basic functions are the same.

1.3 Temperature Dependent DNA Melting

DNA melting, the dissociation of the double-stranded helix into two single-strands, takes place when the weak non-covalent interactions that stabilize double-stranded DNA are broken in a number of ways. One is by heating. The cooperative transition from helix to coil occurs when most of the non-covalent bonds break simultaneously. This highly cooperative phase transition takes place over a very narrow temperature range.

Considering the enthalpies of the coil and helix states as $H_c$ and $H_h$, respectively, the enthalpy of denaturation would be:
\[ \Delta H = H_c - H_h, \]  

(1.1)

which is equal to the heat absorbed by the system at constant pressure. This is the amount of heat needed to melt the DNA. The temperature at which double-stranded DNA melts is called the melting temperature \( T_m \). This temperature depends not only on the DNA sequence composition but also on the pH, ionic strength and other solution conditions. For example, changing the pH of solution changes the net charge of the DNA and can have a marked impact on \( T_m \) and \( \Delta H \).

At constant temperature, the change in the free energy upon melting is given by\textsuperscript{15}:

\[ \Delta G = \Delta H - T \Delta S, \]

(1.2)

where \( \Delta S \) is the change in entropy, and can be given by:

\[ \Delta S = \Delta S(T_m) + \Delta C_p \cdot \ln \left( \frac{T}{T_m} \right), \]

(1.3)

where \( \Delta C_p \) is the change of DNA heat capacity per base pair. It has been shown that there is a significant heat capacity increase associated with DNA melting per base pair\textsuperscript{16,17}.

At the melting temperature, the fraction of molecules in the helix state equals that in the coil state; and the free energy difference between them is zero. Therefore,

\[ T_m = \frac{\Delta H(T_m)}{\Delta S(T_m)}. \]

Because DNA in the helix state is more rigid compared to the coil state, melting due to the disruption of non-covalent interactions causes the entropy to increase, while decreasing the enthalpy. \( T_m \) decreases at low and high pH because the entropy difference between two strands decreases at a faster rate than the enthalpy difference, making \( \frac{\Delta H}{\Delta S} \) progressively smaller. A parameter measured in calorimetric experiments is
Figure 1.5. Ion condensation on highly charged polymer. A highly positively charged polymer in salt solution attracts oppositely charged ions of thickness $R_M$. The net charge $q(r)$ within $r$, per $l_B$, decreases from $q_0$ to zero at infinity. Based on Fig. 1 from O'Shaughnessy, B. and Yang, Q. 21

$$\Delta S(T_m) \approx 25 \text{ cal/K·mol.}$$

The measured heat capacity change is about 65 cal/K·mol, and the free energy change calculated from Equation (1.4) at room temperature is about $2.3k_B T/bp^{16,17}$:

$$\Delta G^0(T) = \Delta S(T_m)(T_m - T) - \frac{\Delta C_p}{2} \frac{(T_m - T)^2}{T_m}.$$  \hspace{1cm} (1.4)

### 1.4 Polyelectrolyte Theory

DNA counterion condensation has been extensively studied and has an intense influence on DNA conformation$^6$. DNA is an example of a highly charged polymer with two electron charges per each monomer unit (a pair of nucleotides). Electrostatic effects play an important role in structural helix-coil transition of DNA. It is shown that the melting point of DNA depends linearly on the logarithm of the salt concentration in the solution$^{18,19}$.

In a model proposed by Manning$^{20}$, the unusual behavior of counterions binding to DNA is described as counterion condensation. In this model, DNA is considered as a highly charged cylinder-like polymer that attracts its counterions very strongly in such a way that a certain fraction of the counterions condenses onto the polymer in the limit of vanishing polymer radius, $(a \to 0)$ (Figure 1.5). Condensation occurs when the Manning
parameter, \( q_0 = l_b / l \) is greater than unity. Here, \( l_b = e^2 / \varepsilon k_B T \) is the Bjerrum length (distance at which the Coulomb energy between two unit charges equals the thermal energy), \( \varepsilon \) is the bulk dielectric constant, \( k_B \) is Boltzmann constant, \( T \) is temperature, and \( e/l \) is the backbone charge density. The Manning number for dsDNA and ssDNA are 4.2 and 1.7, respectively\(^{21}\).

This model is valid as long as the Debye-Hückel screening length (\( \xi = 1/k \)) is smaller than the DNA persistence length (a parameter determining the stiffness of the DNA chain). Debye-Hückel theory is based on the idea that each ion is surrounded by a cloud of opposite charge ions. The inverse measure of the radius of this cloud is known as the Debye length, where \( k^2 = \left( 8\pi q^2 / \varepsilon kT \right) I \). Here, the ionic strength is given by:

\[
I = \frac{1}{2} \sum_{B=1}^{n} c_B z_B^2, \tag{1.5}
\]

where \( c_B \) is the molar concentration of ion B, \( z_B \) is the charge number of that ion, and the sum is taken over all ions in the solution.

The effective charge of DNA or the fraction of ions bound per phosphate group (with the charge of -1) can be given by

\[
N\Psi = 1 - 1/q_0. \tag{1.6}
\]

\( N\Psi \) is 0.76 for dsDNA, where \( N = 1 \) for monovalent ions. This means that in Manning theory 0.76 ions are bound to each phosphate group. However, Record et al.\(^{22}\) showed that the additional cations bound to DNA by Debye-Hückel screening proceeds. Here, the concentration of ions is altered by \((2Nq_0)^{-1}\), and the effective charge of DNA is given by

\[
N\Psi = 1 - 1/2Nq_0. \tag{1.7}
\]
$N\Psi$ is 0.88 for dsDNA where $N = 1$ for monovalent ions. Therefore, in the Record correction to Manning theory, 0.88 ions are bound to each phosphate group.

### 1.5 Ionic Effects in DNA-Protein Interaction

The local concentration of ions within the highly charged cylindrical DNA molecule is about 1.2 M\(^6\). It is independent of bulk concentration and depends only on the DNA charge density. DNA-protein interactions can be considered as a reaction between charged protein molecules (ligands) and the phosphate groups of the nucleotides (lattice)\(^2\). Ligand binding to DNA releases bound counterions. The thermodynamics of the reaction are governed by the entropy of counterions release, and counterions can be treated as a reaction component:

$$DNA + L \rightleftharpoons DNA\bullet L + nM^+. \quad (1.8)$$

Here, $DNA\bullet L$ is the DNA-ligand (protein) complex. The equilibrium binding constant is given by:

$$K = \frac{[DNA\bullet L][M^+]^n}{[DNA][L]}, \quad (1.9)$$

where the observed equilibrium binding constant is defined by:

$$K_{\text{obs}} = \frac{[DNA\bullet L]}{[DNA][L]} = \frac{K_{1M}}{[M^+]^n}, \quad (1.10)$$

such that:

$$\log(K_{\text{obs}}) = \log(K_{1M}) - n \log([M^+]), \quad (1.11)$$
where \( K_{1M} \), the equilibrium binding constant in the standard state( \([M^+] = 1 \text{ M} \) ), depends only on pressure and temperature. A simplified expression for the dependence of \( K_{obs} \) with the concentration of counterions \([M^+]\) is given by\(^{24}\):

\[
\frac{\delta \log K_{obs}}{\delta \log [M^+]} = -Z \Psi,
\]

(1.12)

where, \( Z \Psi \) is equal to the number of ions (cations and anions) displaced in the ligand-DNA binding. \( Z \) is the number of ions formed between the ligand and DNA, and \( \Psi \) is the fraction of cations bound to phosphate group\(^{23}\). For small ligands, the binding of anions to DNA or ligand can be neglected, and the slope given in Eq. (1.12) yields the number of displaced counterions.

### 1.6 McGhee-von Hippel Binding Isotherm

Proteins that regulate several different biological processes in the cell by interacting with the genome can be divided into two groups\(^{25}\). The first are *specific binding proteins* that bind with great affinity to a few operator sites on the genome, and the second are *non-specific binding proteins* that do not have a genetically defined sequence-specific binding site on the genome, interact preferentially with the duplex DNA (stabilizing the helix) or the coil form of DNA (destabilizing the helix), show little base-sequence preference, and cover two or more lattice residues. All proteins studied in this research belong to the latter and are considered as non-specific binding proteins.

To study non-specific binding of large ligands (proteins) to one-dimensional homogeneous lattices (DNA) we should consider the cooperativity factor (\( \omega \)), the number of lattice sites covered by one ligand (\( n \)), and the ligand intrinsic binding constant
(K). Using the conditional probability method, McGhee and von Hippel developed a model to predict the binding isotherm for multi-site ligands\textsuperscript{26}. In this model, lattice sites are either free or blocked by a bound ligand. For any successive free sites (n) on a lattice of N sites, the binding equilibrium is:

\[
K = \frac{[\text{Bound ligand}]}{[\text{Free ligand}][\text{Free ligand binding sites}]} = \frac{b}{c \cdot f}. \tag{1.13}
\]

Using conditional probabilities, the number of locations for an n-mer ligand (f) is given by\textsuperscript{27}:

\[
f = Np(f) p(f|f)^{n-1} = (N-nb) \left( \frac{N-nb}{N-(n-1)b} \right)^{n-1}. \tag{1.14}
\]

Here, \(Np(f)\) is the number of sites that are free and \(p(f|f)\) is the conditional probability of finding a site \(i+1\) free, given that the site \(i\) is free. Substituting Eq. (1.13) in Eq. (1.14), the binding isotherm can be obtained as:

\[
\Theta = K \cdot n \cdot c \cdot \frac{(1-\Theta)^n}{(1-\Theta + \Theta/n)^{n-1}}, \tag{1.15}
\]

where \(\Theta = \frac{nb}{N}\) is the non-cooperative degree of ligand (protein) binding. To analyze the competition-type binding of a mixture of different ligands to the same lattice, the above formula can be extended to:

\[
\Theta_i = K_i \cdot n_i \cdot c_i \cdot \frac{(1-\sum \Theta_i)^n}{\left(1-\sum (\Theta_i + \Theta_i/n_i)\right)^{n-1}}, \tag{1.16}
\]
in which \( i \) refers to the \( i^{th} \) ligand. To consider the next-neighbor interactions among ligands, McGhee and von Hippel generalized the model to:

\[
\Theta = K \cdot \omega \cdot n \cdot c \cdot (1 - \Theta) \left[ \left( \frac{(2\omega+1)(1-\Theta)+\frac{\Theta}{n} \cdot n - R}{2(\omega-1)(1-\Theta)} \right)^{n-1} \left[ \frac{1 - \frac{(n+1)\Theta}{n} + R}{2(1-\Theta)} \right]^2 \right],
\]

(1.17)

where \( R \equiv \left\{ \left[ 1 - \frac{(n+1)\Theta}{n} \right]^2 + 4\omega \cdot \Theta (1-\Theta) \right\}^{1/2} \). For \( \omega > 1 \), the ligands attract each other and exhibit positive cooperativity, while for \( \omega < 1 \), the ligands repel each other and the binding is negatively cooperative. For \( \omega = 1 \), the binding is non-cooperative, Eq. (1.15), and can be exemplified by the weak binding of single-stranded binding proteins to double-stranded DNA (see chapter 4).

### 1.7 Effects of Ligand Binding on DNA melting

Ligands (proteins) that bind preferentially to the double helix or to single-stranded forms of DNA affect the helix-coil melting transition and result in a shift in the DNA melting temperature. For a simple equilibrium process like:

\[
\text{helix} \xrightleftharpoons[K]{\kappa} \text{coil},
\]

(1.18)

the equilibrium constant is given by:

\[
K = \frac{[\text{coil}]}{[\text{helix}]}.
\]

(1.19)

However, in the presence of ligands that bind unequally to helix or coils, the equilibrium process is:

\[
\text{helix} \xrightleftharpoons[K']{\kappa'} \text{coil} + (r_h - r_c) \text{ligands},
\]

(1.20)
and therefore the equilibrium constant will be

\[ K' = \frac{[\text{coil}]}{[\text{helix}]} \cdot [\text{ligand}]^{(n-r_c)} = K \cdot c^{(n-r_c)}, \]

(1.21)

where \( r_{h/c} \) is the average degree of binding to helix/coil and \( c \) is the concentration (more precisely, the activity) of free ligand. For \( r_h > r_c \), ligands stabilize DNA by binding to the helix from of the DNA. Similarly, for \( r_c > r_h \), ligands can destabilize DNA by binding to a greater extent to the coil form than to the helix. Using the van’t Hoff relation, and considering that \( K \) is independent of ligand concentration, we have:

\[ \frac{\partial \ln K'}{\partial (1/T_m)} = \frac{-\Delta H_m^0}{R} = (r_h - r_c) \frac{\partial \ln c}{\partial (1/T_m)}, \]

(1.22)

where \( \Delta H_m^0 \) is the heat of melting in the presence of ligands. Integrating of Eq. (1.22) yields:

\[ \frac{1}{T_m^0} - \frac{1}{T_m} = \frac{R}{\Delta H_m^0} \int (r_h - r_c) d \ln c. \]

(1.23)

The approach proposed by Crothers (1971)\(^{28}\) can be used to determine the expected shift in DNA melting temperature due to ligand binding in which the cooperativity of melting and the entropy of denatured units are not affected by the binding. In principle, \( r_{h/c} \) can be evaluated from equilibrium binding studies. We assume a simple binding reaction as:

\[ \text{helix} / \text{coil} + \text{ligand} \underset{K_{bc}}{\rightleftharpoons} \text{complex}, \]

(1.24)

and the equilibrium constant as:

\[ K_{h/c} = \frac{[\text{complex}]}{[\text{free sites}][\text{free ligand}]} = \frac{r_{h/c}}{(n_{h/c} - r_{h/c})c}. \]

(1.25)
which gives rise to a Scatchard plot. Here, $K_h$, $n_h$, $K_c$, and $n_c$ are apparent binding constants and site sizes for binding to helix and coil, respectively. At constant $T$:

$$\int_0^c r \, d \ln c' = n \ln (1 + Kc).$$  \hspace{1cm} (1.26)

Substituting Eq. (1.26) in Eq. (1.23)

$$\frac{1}{T_m} - \frac{1}{T_m} = \frac{R}{\Delta H_m^0} \int_0^c \left( r_h - r_c \right) \, d \ln c' = \frac{R}{\Delta H_m^0} \left[ \frac{(1 + K_h c)^{n_h}}{(1 + K_c c)^{n_c}} \right],$$  \hspace{1cm} (1.27)

where $K_h$, $n_h$, $K_c$, and $n_c$ are apparent binding constant and site sizes for binding to helix and coil, respectively.

### 1.8 Diffusion-controlled Protein-DNA Association

Diffusion is the expression of random motion (Brownian motion) of molecules in the system and has a key role in regulating the biological processes that depend on the association of proteins to binding sites on nucleic acid chains. Because the thermodynamic tendency in every system is toward uniform concentrations, diffusion help equalizing the concentration differences between different parts. This behavior is summarized by Fick’s first and second laws of diffusion:

$$\mathbf{J} = -D \nabla c,$$  \hspace{1cm} (1.28)

and

$$\left( \frac{\partial c}{\partial t} \right) = D \nabla^2 c,$$  \hspace{1cm} (1.29)

where $\mathbf{J}$ is the flux vector, $D$ is the diffusion coefficient, and $c$ is the number of particles per unit volume. The relation between the diffusion and friction is given by Einstein-Stokes equation for a sphere of radius $R$ as:
\[ D = \frac{k_b T}{\xi} = \frac{k_b T}{6\pi \eta R}, \]  

where \( \xi \) is the frictional coefficient, \( k_b \) is the Boltzmann coefficient, \( T \) is the temperature, and \( \eta \) is the solvent viscosity.

At the microscopic level, diffusion results from Brownian motion (due to collisions with neighboring) in which molecules make random walks (characterized by a mean free path) about a central location. By comparing the microscopic random walk with the macroscopic diffusion, we can obtain the mean-square distance in which a particle moves in a time \( t \) as:

\[ \langle r^2 \rangle = 6Dt. \]

Solving the Eq. (1.29) for a spatial concentration gradient, we can obtain the ligand association rate by:

\[ k_{a,3D} = 4\pi DR, \]  

in which \( R \) is the ligand size.

Diffusion theory is important in understanding the rates of bimolecular reactions in solution. In protein-DNA interaction studies, much interest has been focused on diffusion-controlled reaction by the finding that a simple three-dimensional search is not sufficient to explain experimentally measured target search rates by some proteins. This rate-enhancement is assigned to a two-step binding process. The first step is due to the non-specific binding of the protein to the rest of the nucleic acid, and the second step involves thermally driven protein translocation events, that serves to speed up the search process. The process that may be involved in these protein translocation events are inter-segment transfer, hopping, and sliding. However, in DNA stretching experiments,
the first mechanism is impossible to occur since single DNA molecules are straightened by force. In section 4.2, the mechanism of rate-enhancement in the interactions of phage T7 single-stranded binding proteins with ssDNA will be discussed.

1.9 Single Molecule Force Spectroscopy

Single molecule manipulation techniques have provided new information about the physical and biomechanical properties of both nucleic acids and proteins\textsuperscript{30-34}. In order to obtain a better understanding of biological processes involving DNA, it is crucial to study the properties of single DNA molecules and molecules that interact with them under different conditions presented \textit{in vivo}. By isolating a single molecule, we can simplify the system we are studying in the absence of interactions between molecules, and obtain fine details which otherwise could be lost due to ensemble averaging in a bulk experiment. In addition, because DNA is stretched, DNA melting studies can be performed over a wide range of temperature and solutions, thus avoiding protein denaturation and protein-DNA aggregation.

\textit{In vivo}, the double helical structure of DNA can be disrupted by enzymes that are responsible for processing DNA in the cell. These biological motors transfer chemical energy into mechanical work to stretch or unzip the double helix. Similarly, single molecule manipulation techniques can perform work and produce molecular extension \textit{in vitro} on an energy scale which is comparable to the thermal energy ($k_B T$). Using these techniques, we can measure the force required to produce a given molecular extension, and study the response of an individual biomolecule to an applied force. The DNA helix-coil transition is associated with a decrease in enthalpy due to disruption of stacking
interactions and hydrogen bonds, and an increase in entropy due to increased mobility of ssDNA vs. dsDNA. Single molecule techniques can measure thermodynamic quantities associated with the conversion of nucleic acids from one structure to another.

In this study, we performed single molecule force measurements using a dual beam optical tweezers instrument to measure the elasticity of the DNA, as well as the forces under which the DNA breaks or undergoes a phase transition in the presence of different ligands, which allows us to quantify the thermodynamics and kinetics of the protein-protein and protein-nucleic acid interaction.

1.9.1 Optical Trapping of Dielectric Particles

The optical trap (also known as optical tweezers) instrument has been used in physical science for years for the manipulation and study of micron sized dielectric particles\textsuperscript{35-40}. More recently it has become a powerful tool for assessment of small forces in biological systems due to the necessity of knowing all the possible states of a single molecule in the absence of interactions between a large number of such units\textsuperscript{30,33,41-43}.

Optical trapping of small particles by the force of photon pressure was first demonstrated by Ashkin \textit{et al.}\textsuperscript{35-40} They discovered that by focusing a laser beam through a microscope objective, dielectric particles with high indexes of refraction were attracted to the highest intensity region in the beam and could be held at the beam focal point. Optical tweezers use the forces of radiation pressure to trap small, polarizable particles in a three-dimensional potential well near the beam focus. For dielectric particles large compared with the wavelength, simple ray optics can be used to derive the net radiation force from the change of incident light momentum. In this regime, light beams are composed of individual rays each having proper direction, state of polarization, and
intensity. In a medium with uniform refractive index, each ray propagates in a straight line. However, when an incident light reflects or refracts on the surface of a particle (the boundary surface between two medium), its optical path is changed, the direction and polarization of each ray changes according to Fresnel laws, and therefore the incident light exerts an optical force on the particle. The fundamental principle behind the operation of optical tweezers comes from the fact that a beam light consists of photons each carrying a momentum \( \vec{P} \):

\[
\vec{P} = \hbar \vec{k},
\]

where \( \hbar \) is the Plank’s constant and \( \vec{k} \) is the wave vector. This momentum is proportional to the propagation vector of the electromagnetic field:

\[
d(\frac{d\vec{P}}{dt}) = (n / c)\vec{S}dA,
\]

where \( \vec{S} \) is the Poynting’s vector, \( c \) is the speed of light, and \( dA \) is an element of area normal to \( \vec{S} \). When parallel rays enter the sphere, they deflect. Because momentum is conserved, an equal and opposite momentum change is conveyed to the sphere. The total force on the particle due to the refraction of light is given by:

\[
\vec{F} = d(\vec{P}_{in} / dt) - d(\vec{P}_{out} / dt) = (n/c) \iint (\vec{S}_{in} - \vec{S}_{out})dA,
\]

where \( \vec{S}_{in} \) and \( \vec{S}_{out} \) are the light intensity entering and leaving the particle, respectively. When a beam Light is reflected and refracted on the particle surface according to Snell's law, the particle feels a reaction impulse that is equal but opposite to the change in the photon’s momentum. This means that light exerts a force on all objects that refract or reflect light. The resultant forces due to refraction and reflection are in the opposite direction. In Figure 1.6, the action of trap on a dielectric sphere is described in terms of
both reflection and refraction of a typical pair of rays $a$ and $b$ of the converging beam.

The attractive force due to refraction of the light at the surface of the bead is in the direction of the focus light, while the force due to reflection tries to push the sphere out of trap. This problem can be solved by having a high trapping force that is maintained by using a microscope objective with a high numerical aperture (NA) and increasing the index of refraction of the medium immediately following the objective (using water or oil). Alternatively, the loss due to reflection at the bead surface can be resolved by using dual beam optical tweezers. In this design (see section 1.10.1), two counter-propagating beams focus to the same spot. Therefore, the approximately same scattering forces produced by the two laser beams can cancel each other out. In dual beam optical tweezers, both lasers contribute to lateral stability and give a three dimensional trap of very high stiffness.

Because thermal fluctuations cause the bead to execute Brownian motion while trapped, an optical trap can be considered as a virtual spring or a harmonic potential well that pulls the bead toward the trap center or its equilibrium position at the bottom of the trap’s potential well. This force is proportional to the displacement of the particle within the trap ($\Delta x$) by the relation $\vec{F} = -k\Delta \vec{x}$, where $k$ is the trap stiffness. So, to

![Figure 1.6. Schematic diagram of the resultant force of an incident light on a dielectric sphere. The attractive force due to refraction of the light at the surface of the bead is in the direction of the focus light, while the force due to reflection tries to push the sphere out of trap.](image-url)
measure force on a bead in an optical trap, it is necessary to be able to measure the trap stiffness. There are two ways to calibrate the trap stiffness, passive and active methods. The passive method involves monitoring the thermal motion of a trapped bead and then calculating the variance. The equipartition theorem can be used to obtain the trap stiffness:

\[
\frac{1}{2} k \langle \Delta x^2 \rangle = \frac{1}{2} k_B T,
\]

(1.36)

where \( \langle \Delta x^2 \rangle \) is the variance. This method requires an accurately calibrated detector with a sufficient bandwidth. Furthermore, the signal to noise ratio must be high enough to prevent any additional noise. Finally, the bead’s motion must be only in the linear region of trap. In the active method, a known external dragging force is applied on the bead in the trap (by moving the stage or the trap) and the response of the bead is analyzed. This method requires an accurate calibration of the stage and depends on the size, shape and the orientation of the trapped particles. We should also take into account that the bead undergoes thermal motion during the experiment, so averaging is necessary to remove the thermal component.

1.9.2 Models of Polymer Elasticity

The morphology of DNA suggests how DNA can recognize and bind to other molecules, such as proteins, and how it packs into cellular components or viral capsids. Over the recent past years\textsuperscript{30,45}, single molecule studies on single- and double-stranded DNA have been made impressive progress in understanding the physics of DNA by resolving local stresses and strains in a single DNA molecule. A number of models try to predict the elastic properties of DNA. Among them, the worm-like chain (WLC) and freely-jointed
chain (FJC) models have worked well to explain the elastic properties of DNA molecules\textsuperscript{46,47}. By fitting optical trap measurements of force-extension relation of a λ-DNA molecule to these models, we can obtain the DNA’s average elastic properties.

### 1.9.2.1 The Freely-Jointed Chain Model

Long macromolecules have different mechanisms of flexibility. The Freely-jointed chain (FJC) or random-walk chain model\textsuperscript{47} is the simplest model to describe these polymers. In this model, the interactions among successive monomers are neglected and the molecule is considered as a chain of rigid subunits with length \( l \) that are joined by perfectly flexible hinges. As it is shown in Figure 1.7, the directions of bond vectors in the chain are completely uncorrelated to each other and the flexibility of the molecule is concentrated at the connection points. The end-to-end distance vector \( \vec{R} \) can be calculated as the sum of bond vectors \( l\hat{u}_i \):

\[
\vec{R} = l \sum_{i=1}^{N} \hat{u}_i, \tag{1.37}
\]

and the mean square end-to-end distance can be calculated as:

\[
\langle R^2 \rangle = Nl^2, \tag{1.38}
\]

where \( \sum_{j \neq i} \langle \hat{u}_i \cdot \hat{u}_j \rangle = 0 \) when segment directions in a freely-jointed chain are not correlated. The effective Kuhn segment length \( l \) is longer than the actual monomer size and represents more than one monomer. This parameter describes the flexibility of the macromolecule and can be obtained from the fit of force-extension measurements to FJC.
model and depends on the external conditions that affect intramolecular interactions. The number of these segments \((N)\) and their length \((l)\) are determined by the requirement that both Kuhn model chain and real chain have the same value of \(\langle R^2 \rangle\) and contour length (total stretched-out length of the chain)\(^{27,48}\).

In the random-walk chain model, statistical distribution of polymer end-to-end distance vector is Gaussian and the relative number of all polymer conformations is given by\(^{27,48}\):

\[
P_N(\vec{R}) = \left[ \frac{3}{2\pi NL^2} \right]^{3/2} e^{-3R^2/2NL^2}.
\]

Here, the prefactor is found from the normalization condition \(\int P_N(\vec{R}) d^3\vec{R} = 1\) and decays at \(R \sim N^{l/2}\) which agrees with Eq. (1.38).

In the presence of an external force \((F_z)\), polymers act like a Hooke’s law spring where the entropic refractive force is proportional to the extension. In the low force limit \((F < 5 \text{ pN})\), all the conformations of the freely-jointed chains have the same energy resulting in a pure entropic free energy\(^{27,47}\):

\[
\frac{\epsilon_{\text{FJC}}}{k_B T} = -\sum_{i=1}^{N} \frac{Fli}{k_B T} \hat{u}_i = -\ln P_N(\vec{R}) = \frac{3}{2NL^2} R^2 + \text{constant.}
\]

The refractive force can be derived as:

\[
F = -\frac{d\epsilon_{\text{FJC}}}{dL} = -\frac{3k_B T}{NL^2} R,
\]

where \(\epsilon_{\text{FJC}}\) is the freely-jointed chain free energy, \(k_B\) is the Boltzmann constant, and \(T\) is temperature. The applied force reduces the conformational entropy of the polymer. The
resulting entropic elastic behavior can be summarized in the force-extension relation using the well-known Langevin function:

\[
z(F) = L_c \left[ \coth \left( \frac{Fl}{k_B T} \right) - \frac{k_B T}{Fl} \right],
\]

(1.42)

where, \( z \) is the extension and \( L_c \) is the contour length. The force-extension measurements of single-stranded DNA per base pair can be fit to the extensible freely-jointed chain model (FJC) by:

\[
b(F) = b_{ss}^{\text{max}} \left[ \coth \left( \frac{2FP_{ss}}{k_B T} \right) - \frac{1}{2} \frac{k_B T}{FP_{ss}} \right] \left[ 1 + \frac{F}{K_{ss}} \right],
\]

(1.43)

where, \( b \) is the extension per base pair, \( b_{ss}^{\text{max}} \) is the ssDNA contour length per base pair, \( P_{ss} \) is the persistence length, and \( K_{ss} \) is the elastic stretch modulus. Experimental values for ssDNA in 150 mM [Na⁺] for \( b_{ss}, P_{ss}, \) and \( K_{ss} \) are 0.56 nm, 0.75 nm, and 800 pN, respectively and are affected by solution conditions.

1.9.2.2 The Worm-Like Chain Model

In the worm-like chain model, a polymer is treated as a continuously flexible rod. This model is similar to the freely-jointed chain model, but the polymer is treated as a continuous elastic medium. This means that at room temperature, the polymer is like a stiff rod over short lengths, but it gradually curves over long lengths. The worm-like chain model is particularly suited for describing stiffer polymers like double-
stranded DNA in which successive segments of dsDNA display a sort of cooperativity and are constrained to point in nearly the same direction. Figure 1.8 shows a cartoon of worm-like chain model. For a polymer of length $L_c$, the path of the polymer is parameterized as $s \in (0, L_c)$. $\hat{t}(s)$ is the unit tangent vector to the chain at $s$, and $\hat{r}(s)$ is the position vector along the chain. Assuming that the chain is inextensible ($|\hat{t}(s)|=1$), the local tangent and curvature vectors ($\hat{t}(s)$ and $\hat{w}(s)$, respectively) are given by:

$$\hat{t}(s) = \frac{d\hat{r}(s)}{ds}, \quad \hat{w}(s) = \frac{d\hat{t}(s)}{ds}.$$

The effective energy of a stretched worm-like chain taking into account the resistance of the chain to bending can be driven as:

$$\frac{E_{WLC}}{k_BT} = \int_0^{L_c} ds \left[ \frac{P_{ds}}{2} |\hat{w}(s)|^2 - \frac{F}{k_BT} \hat{t}(s) \cdot \hat{w}(s) \right],$$

where $P_{ds}$ is a measure of the persistence length of the chain reflecting the bend stiffness at zero stretching force. It is the characteristic distance along the worm-like chain over which the tangent vectors all pointing roughly the same direction and their correlation function follows an exponential decay:

$$\left\langle \hat{t}(s) \cdot \hat{t}(0) \right\rangle_{WLC} \sim e^{-\frac{s}{\xi}}.$$

The end-to-end distance is:

$$\bar{R} = \int_0^{L_c} \hat{t}(s) ds.$$

So, the mean square end-to-end distance of the polymer is:

$$\left\langle R^2 \right\rangle = \left\langle \bar{R} \cdot \bar{R} \right\rangle = 2P_{ds}L_c \left[ 1 - \frac{P_{ds}}{L_c} \left( 1 - e^{-L_c/\xi} \right) \right] \approx 2P_{ds}L_c.$$
Here, the last approximation holds for a polymer with large contour length \( L_c \gg P_{\text{sw}} \), and can be used to show that a Kuhn segment is equal to twice the persistence length of a worm-like chain.

The analytical solution for the force-extension relation of worm-like chains is not currently known. However, the Eq. (1.45) has been solved numerically by Marko and Siggia\(^49\), and a high-precision interpolation formula to the numerical solution is used to fit DNA force-extension measurements. At high force limit, we have:

\[
b(F) = b_{\text{ds}}^{\text{max}} \left[ 1 - \frac{1}{2} \left( \frac{k_B T}{P_{\text{ds}} F} \right)^{1/2} + \frac{F}{K_{\text{ds}}} \right]. \quad (1.49)
\]

where \( b \) is the extension per base pair. Typical measured values for \( b_{\text{ds}}^{\text{max}} \) (the dsDNA contour length), \( P_{\text{ds}} \) (the persistence length), and \( K_{\text{ds}} \) (the elastic stretch modulus) are 0.34 nm, 48 nm, and 1200 pN, respectively\(^50\), though they depend upon salt concentration and pH\(^45,49,51\). At low force limit, Eq. (1.49) displays a Hookean linear relation, but as the extension nears the contour length of the molecule, it scales not as \( \frac{1}{F} \) (predicted by freely jointed chain model), but as \( \frac{1}{\sqrt{F}} \), in significantly better agreement with the data.

### 1.9.3 Force-Induced Melting

Optical tweezers have been used extensively for studying the biomechanical properties of single DNA molecules by stretching the molecules and measuring the required force for a given extension under various conditions\(^30,32,33,42,50-52\). The mechanical work performed by stretching has an energy scale of the noncovalent interactions that hold the two DNA strands together\(^53\) and can therefore be used to induce conversion of double-stranded
DNA (dsDNA) into single-stranded DNA (ssDNA). When this mechanical process is reversible, the calculated work is equal to the equilibrium melting free energy\(^{51,53}\).

The method of ssDNA stretching in which two strands are melted by force is referred to as force-induced melting (FIM). This method provides valuable information regarding the interaction between nucleic acids and proteins or small molecules that bind to DNA. In our experiments, a single \(\lambda\)-phage DNA molecule of 48,500 base pairs was stretched to extensions that were almost twice its B-form contour length, resulting in a FIM transition. Extended regions of dsDNA melt cooperatively, and the midpoint of the melting transition, \(F_m\), is analogous to the DNA melting temperature, \(T_m\), obtained in thermal melting studies (see section 1.3). \(F_m\), like \(T_m\), is similarly affected by solution conditions such as pH, temperature, and ionic strength\(^{50,51,54,55}\). DNA-binding proteins and small molecules that affect the thermal melting equilibrium of dsDNA affect the FIM transition in a similar manner\(^{51,56-64}\).
Typical force-extension curves\textsuperscript{65} for $\lambda$-DNA in 10 mM HEPES pH 7.5, 100 mM $[\text{Na}^{+}]$ are shown in Figure 1.9. At low extensions, the tension increases as the DNA is first uncoiled. Beyond this entropic stretching regime, as the molecule end to end extension nears the canonical B-form contour length, DNA begins to unwind and the backbone reveals an elastic response that results in an abrupt rise in the force-extension curve. At these forces, DNA molecule can be stretched slightly further beyond its contour length. The properties of double-stranded force-extension data can be successfully characterized by the worm-like chain model (WLC) which describes an elastic polymer with a fixed dihedral bond angle (see section 1.9.2.2).

As the DNA is uncoiled, the dsDNA elastic effect ends with onset of the overstretching region. The resulting plateau presents the overstretching transition, where significant cooperative base pair melting occurs and it continuous until the molecule is stretched to 1.7 times its B-form contour length. Extending DNA further, the force again rises rapidly with a slope that depends on stretching rate\textsuperscript{66,67} and yields ssDNA, such that the two ssDNA strands will separate at applied forces at about 150 pN. The data of single-stranded DNA can be fit to the extensible freely jointed chain model (FJC) that describes an elastic polymer with a varying bond angle (see section 1.9.2.1). For a reversible process, the melting free energy as a function of force can be estimated as the area between the double-stranded and single-stranded stretching curves:

$$
\Delta G(F) = -\int_{0}^{F} \left[ b_{ss} \left( F' \right) - b_{ds} \left( F' \right) \right] \cdot dF'.
$$

(1.50)

The hysteresis observed upon reannealing is strongly dependent on the solution conditions, such as temperature, salt, and pH and can be explained by a slow reannealing process on the time scale of our experiment.
1.9.3.1 Force-Induced Melting Model (FIM)

To quantitatively analyze the helix-coil transition, we use the *force-induced melting model*[^FIM][^FIM2][^FIM3][^FIM4]. The model shown in Figure 1.10, consists of two stages. The first is a reversible equilibrium melting transition. When a single DNA molecule is stretched at forces less than the overstretching force, the bases remain paired and the helical form is maintained. It is always thermodynamically favorable for the molecule to maintain a significant number of domain boundaries. During the transition, domains of melted DNA are separated by helical sections. The second stage consists of nonequilibrium melting, in which for forces more than the overstretching force, the domain boundaries are removed and the two DNA strands separate in an irreversible rate dependent manner. This two-stage melting model only applies to a torsionally relaxed DNA. If the DNA is torsionally constrained, more complex curves depending on helical strain are obtained[^FIM5][^FIM6]. The model also explains the hysteresis observed under conditions that inhibit reannealing, like high temperature, low salt, and high pH[^FIM7][^FIM8][^FIM9][^FIM10].

[^FIM]: [51,53,68]
[^FIM2]: [51,53,68]
[^FIM3]: [51,53,68]
[^FIM4]: [51,53,68]
[^FIM5]: [69,70]
[^FIM6]: [69,70]
[^FIM7]: [50,51,54,55]
1.9.3.2 Thermodynamics of Force-Induced DNA Melting

To describe the thermodynamics of DNA overstretching, theoretical and experimental studies have been carried out that lead to a new model, so called as force-induced melting model (FIM). Detailed measurements of the dependence of this transition on solution conditions such as pH, temperature, ionic strength, and the presence of DNA binding ligands, follow that expected for force-induced melting\(^5\)\(^0\),\(^5\)\(^1\),\(^5\)\(^3\)-\(^5\)\(^5\),\(^6\)\(^0\)-\(^6\)\(^4\),\(^6\)\(^8\),\(^7\)\(^1\). Any solution condition that destabilizes DNA, and therefore lowers its melting temperature, must also lower the overstretching force, which we define as the force required to stretch the molecule halfway through the overstretching transition.

The expected decrease in the overstretching force in extremes of low or high pH is shown experimentally\(^5\)\(^5\) which is consistent with the experimental results of the decrease in thermal melting point of DNA\(^7\)\(^2\). It is shown that high and low pH do not affect the secondary structure of dsDNA\(^7\)\(^3\),\(^7\)\(^4\). The main protonation and deprotonation sites are accessible only to ssDNA and are the same ones that participate in hydrogen bonding. This means that protonation and deprotonation of the bases occur only if the DNA strands are dissociated. The protonation occurs at the NH\(_2\) amino group of A, G, and C bases with pK\(_a\) equal to 3.8, 2.9 and 4.6, respectively\(^5\)\(^5\), whereas the deprotonation sites are the –NH-CO- groups of the bases with pK\(_a\) between 9.5 and 10.5 in high salt.

Hysteresis observed in the release part of the DNA stretching cycle can also support the melting nature of the overstretching transition. If the overstretched form of DNA consists of two separate single strands, then increasing hysteresis at high pH can be explained by a growing negative charge on the ssDNA as they deprotonate at high pH.
On the contrary, the absence of hysteresis observed in the low pH can be explained by charge neutralization and reduction of the mutual repulsion between the two strands.

Results from stretching experiments show that the overstretching force decreases with increasing temperature and these results are consistent with thermal melting studies\textsuperscript{54,75}. However, to be able to compare stretching results with the results provided from thermal melting studies, we need to obtain the stability of dsDNA at any temperature $\Delta G(T)$. Predictions of temperature dependence of $F_{\text{overstretch}}$ can be calculated directly from force-extension curves of dsDNA and ssDNA. The total transition free energy can be written as $\Delta G(T) + \Delta G(F)$, where $\Delta G(T)$ and $\Delta G(F)$ are the temperature and force dependent melting free energies, respectively. The overstretching transition occurs when the total transition free energy is zero. This means that at melting temperature ($T_m$), $\Delta G(T)$ is zero, so no force is required for melting the DNA. Conversely, at lower temperatures, $\Delta G(T)$ is positive, so a force must be applied to the DNA to lower $\Delta G(F)$ until the total transition free energy is zero. Because $\Delta G(F)$ can be calculated directly from stretching experiments by Eq. (1.50) and $\Delta G(T) + \Delta G(F)$ is zero at $F_{\text{overstretch}}$, measurements of the force-extension curves of ssDNA and dsDNA provide a direct measurement of the stability of dsDNA at any temperature $\Delta G(F)$.

Studies of the ionic strength dependence of the DNA overstretching transition show that the decrease in $F_{\text{overstretch}}$ in low salt concentration is consistent with the predictions of polyelectrolyte theory. The salt dependent part of the helix-coil transition free energy is given by\textsuperscript{76}: 
\[ \Delta G = k_B T \left( \frac{1}{\xi_{ss}} - \frac{1}{\xi_{ds}} \right) \ln \left( \frac{I}{I_0} \right) \]  

(1.51)

where \( I \) is the solution ionic strength and \( \xi \) is the dimensionless linear charge density, with \( \xi = \frac{l_B}{h} \) and \( l_B = \frac{e^2}{\varepsilon k_B T} \). Here, \( e \) is the elementary charge, \( h \) is the length per unit charge, \( \varepsilon \) is the dielectric constant, and \( l_B \) is the Bjerrum length. The variation of the \( F_{\text{overstretch}} \) with \( I \) is given by\(^{68}\):

\[ \frac{\partial F_{\text{overstretch}}}{\partial \ln(I)} = \frac{k_B T}{l_B} v, \]

(1.52)

where

\[ v = \frac{h_{ss} - h_{ds}}{b_{ss} - b_{ds}}, \]

(1.53)

and \( v \) is the ratio of the differences in length per unit charge \( (h) \) and base pair \( (b) \) when DNA is overstretched. When the distance between two DNA stands is less than the Debye screening length, the two strands can be considered as a single strand with twice the charge density with \( v = 0.5 \). However, if the average distance between strands is greater than the Debye screening length, then \( v = 1.7 \). When one strand is stretched while the other one is relaxed, then \( v = 1.2 \). Because \( v \) obtained from stretching experiments is close to 0.5, it can be suggested that both DNA strands are stretched and remain close to each other during the overstretching transition.
1.10 Experimental Procedures

1.10.1 Instrument Setup

The dual-beam optical tweezers instrument used in this study consists of two counter-propagating diode lasers, each with 200 milliwatts of 830 nm light (JDS Uniphase, San Jose, CA). Passing through quarter wave plates, beams are circularly polarized and then convergently directed and focused to a small spot inside a liquid flow cell, using 60X, 1.0 numerical aperture water immersion microscope objectives (Nikon, Tokyo) forming the optical trap. As the beams leave the trap, the polarization, now horizontal, allows another beam-splitting cube to direct the light into a lateral effect photodiode detector, which determines the deflection of each beam to within a few microns, and outputs a voltage that is directly proportional to the force being exerted on the bead in the optical trap. The white light sources and charged-coupled device (CCD) cameras provide simultaneous images of the tip and the beams.

Figure 1.11. Schematic diagram of optical tweezers instrument. Passing through quarter wave plates, two counter-propagating near infra red beams are circularly polarized and then convergently directed and focused into a small spot inside a flow cell using a confocal pair of microscope objectives forming dual beam optical trap. As the beams leave the trap, the polarization, now horizontal, allows another beam-splitting cube to direct the light into a lateral effect photodiode detector, which determines the deflection of each beam to within a few microns, and outputs a voltage that is directly proportional to the force being exerted on the bead in the optical trap. The white light sources and charged-coupled device (CCD) cameras provide simultaneous images of the tip and the beams.
beam-splitting cube to direct the light into a lateral effect photodiode detector (UDT Sensors, Hawthorne, CA), which determines the deflection of each beam to within a few microns, and outputs a voltage that is directly proportional to the force being exerted on the bead in the optical trap. The white light sources and charged-coupled device (CCD) cameras provide simultaneous images of the tip and the beams. A diagram of the optical tweezers used in this research is shown in Figure 1.11, courtesy of Dr. Micah J. McCauley.

1.10.2 DNA Stretching

Bacteriophage λ-DNA (New England Biolabs), ~ 48,500 base pairs, was labeled with biotin on each 3' terminus and was repurified by extraction with phenol and chloroform and ethanol precipitation. To tether single DNA molecules, two 5-μm streptavidin coated polystyrene beads (Bangs Laboratories, Fishers, IN) were trapped in the optical trap and on the end of a glass micropipette (World Precision Instrument, Sarasota, FL). A very dilute solution containing biotin-labeled λ-DNA, typically in 10 mM HEPES pH 7.5, 100 mM [Na+] was run through the cell until one single DNA molecule was captured between the two beads. The flow cell, and thus the glass micropipette, may be moved using a feedback-compensated piezoelectric stage (Melles Griot), causing the single DNA molecule to stretch between two beads, resulting in a force-extension measurement, as described previously (see section 1.9.3). The position measurements were converted to a measurement of the molecular extension by correcting for the trap stiffness. After capturing a single DNA molecule in the tethering buffer, the molecule was stretched to
verify that the usual force-extension curve was obtained and that only a single molecule had been tethered.

To measure the effect of different proteins and chemicals, 4 to 5 cell volumes of a buffer solution containing a fixed amount of each was added to the experimental cell until the buffer surrounding the captured DNA molecule was completely exchanged. All measurements were performed at room temperature.

In order to calculate the pulling rate-dependence of the overstretching force reported in this work, the pipette was moved in different size steps of 5–250 nm at a rate of \( \sim 1 \) step per second and after each step the force was measured 100 times and averaged, thus averaging out contributions of thermal motion to the force measurement.

1.10.3 Protein Preparation and Purification

1.10.3.1 Phage T7 Gene 2.5 Protein

Wild type gp2.5 and gp2.5-Δ26C were purified from BL21(DE3)pLysS cells overexpressing histidine-tagged version of their genes as previously described\(^78\). Following the purification, the histidine tag was proteolytically cleaved using PreScission protease (GST-tagged, Amersham). The cleaved histidine tag and the protease were subsequently removed using nickel-NTA agarose (Qiagen) and GSTTrap\(^\text{TM}\)HP columns, respectively. The purified proteins were dialyzed against storage buffer (50 mM Tris-HCL, pH 7.5, 0.1 EDTA, 1 mM DTT, 50% glycerol) and stored at –20 °C. The storage buffer for gp2.5-Δ26C contained additional 150 mM NaCl. For experiments requiring high concentrations of gp2.5 the protein solution was concentrated using an Amicon Ultra centrifugal filter device (Millipore) with 10 kDa cut off.
1.10.3.2 α Subunit of E. coli Pol III

Plasmid Construction- Plasmids encoding E. coli full-length Pol III α subunit and Pol III \(^{1-917}\) with an N-terminal histidine tag (pET28a-α and pET28a- α\(^{917}\), respectively) were generous gifts from Prof. J. Kuriyan (UC-Berkeley). Plasmids encoding E. coli Pol III \(^{917-1160}\) (pAlpha917-1160) and E. coli Pol III \(^{978-1160}\) (pAlpha978-1160) were constructed from plasmid pET28a-α, as follows: NdeI restriction sites were introduced at residues 917 and 978 using the QuikChange site-directed mutagenesis kit (Stratagene) and the following primers with the respective reverse complements, respectively: 5′-GAAAGCGGAACATATGGGTCAGGCCGATATG-3′ and 5′-GAGATTTGAGCGTTATCATATGCGGCTAAGGCTGAAAGAC-3′. After digesting with NdeI, the plasmids were then ligated to yield the E. coli Pol III \(^{917-1160}\) and E. coli Pol III \(^{978-1160}\), with amino acid residues 917-1160 and 978-1160 of wild-type E. coli Pol III, respectively. The integrity of the constructs was confirmed by automated DNA sequencing analysis (MGH Core Facility or MIT Biopolymers Laboratory, Cambridge, MA).

Protein Preparation- Wild-type E. coli Pol III and E. coli Pol III \(^{1-917}\) were expressed in Tuner (Novagen), Tuner pGro7 (Takara), or BL21(DE3) pLysS competent cells, while E. coli Pol III \(^{917-1160}\) and E. coli Pol III \(^{978-1160}\) were expressed in Tuner pGro7. Competent cells were prepared by using the CaCl\(_2\) method. All proteins were expressed by inducing cultures at \(OD_{600} = 0.8\) with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) at 30 °C for 3-4 h. The cells were harvested, resuspended by sonication, and lysed by treatment with lysozyme and DNase in lysis buffer (50 mM HEPES pH 7.5, 250 mM NaCl, 10% Glycerol, and 2 mM β-mercaptoethanol). Clarified cell extract was loaded
onto 2 mL Ni-NTA-Agarose (Qiagen) and bound proteins were eluted with a step gradient at 0.01, 0.025, 0.1, 0.25, 0.5 M imidazole in the lysis buffer. Pure fractions identified via SDS-PAGE were concentrated in Vivaspin centrifugal concentrators (Vivascience) and the buffer exchanged to 50 mM HEPES pH 7.5, 250 mM NaCl, and 10% Glycerol to remove imidazole. For experiments conducted in low salt, the buffer was exchanged to 50 mM HEPES pH 7.5, 50 mM NaCl, 10% Glycerol by using Vivaspin centrifugal concentrators. Purified protein concentrations were determined by Bradford protein assay (Bio-Rad), and concentrated proteins were stored at -80 °C.

1.10.3.3 Phage T4 Gene 59 Protein

Wild type gp59 purified as described by Morrical et al. following overexpression in E. coli strain BL21(DE3)plysS. gp59R12A mutation was engineered by QuikChange. It was expressed by inducing cultures with isopropyl-β-d-thiogalactopyranoside (IPTG) at 37 °C. After extracting the cell pellet, sonication was used for lysis. Wild type gp59 or gp59R12A was purified from the cell lysate using the following three columns: cellulose phosphate, hydroxyapatite, and single strand DNA cellulose. After each column, purity was determined through SDS-PAGE electrophoresis. After purification, the protein was tested for nuclease contamination through a standard electrophoresis assay. The protein was stored at –20 °C in 20 mM Tris-HCl pH 7.4, 0.2 mM EDTA, 50 mM NaCl, 1mM β-mercaptoethanol, and 62% glycerol.

1.10.3.4 Phage T4 Gene 32 Protein

T4 gene 32 protein and its truncated form *I used in this study were prepared as described.
1.10.3.5 Glycerol

An aqueous solution of 40% glyoxal was obtained from Sigma-Aldrich.

1.11 Chapter Overview

DNA is the carrier of genetic information in every living cell. Our view of this genetic blueprint has been expanded considerably in recent years with a new understanding of the molecular mechanisms that regulate the multiple reactions required for the replication of a genome. It has a great importance to know how cells grow, divide and differentiate, because mistakes provoke alterations in proteins or their expressions that give rise to different diseases. However, there are still many unanswered questions regarding the understanding of the reactions catalyzed by the DNA replication proteins and the functional role of these proteins for the replication of a chromosome. The enzymes and other protein factors that carry out DNA replication in *E. coli* and in eukaryotic cells are analogous, suggesting that the biochemical mechanism of DNA replication is similar in all cells. *E. coli* and the bacteriophages that infect it have been used extensively in molecular biology research and provide excellent model systems for analyzing DNA replication. Our goal is to understand the protein-protein and protein-DNA interactions responsible for the coordination of events at the replication fork. In the studies presented here, we use single molecule force spectroscopy to quantify nucleic acid binding thermodynamics and kinetics. These measurements can often be extended to solution conditions not available to ensemble experiments, allowing us to obtain significant new insight into specific nucleic acid-protein interactions.
In chapter 2, we further test the validity of the force-induced melting model without relying on thermodynamic conclusions. The quantitative force-induced melting model is analogous to thermal melting, where double-stranded DNA is converted into single-stranded DNA. However, an alternative explanation suggests that the B-DNA transforms into another double-stranded form, termed S-DNA, that is 1.7 times longer, largely unwound, but preserves the interstrand base pairing. To determine the extent to which the DNA base pairs are exposed to solution during the transition, we have undertaken a study of λ-DNA overstretching in the presence of glyoxal, a chemical used previously to map DNA thermal melting. We show that if overstretching is a strand separation, then all of the force-melted base pairs should become glyoxal-modified, and therefore essentially permanently single-stranded.

In chapter 3, we show that intact phage T7 gene 2.5 protein and its deletion mutant lacking 26 C-terminal residues, gp2.5-Δ26C, lower the phage λ-DNA melting force as measured by single molecule force spectroscopy. We determine the equilibrium binding constants of these proteins to single-stranded DNA as a function of salt concentration not available in bulk binding studies. We also calculate the free energy of gp2.5 dimerization per dimer. Our results agree with previous binding studies, where available. We propose a model to explain the proteins’ structure-function relationship, and show that dimeric gp2.5 must dissociate prior to binding to single-stranded DNA, a dissociation that consists of a weak non-electrostatic and a strong electrostatic component.

In chapter 4, by studying the rate-dependent DNA melting force in the presence of gp2.5 and gp2.5-Δ26C we probe the kinetics and thermodynamics of gp2.5 binding to single- and double-stranded DNA. These force measurements allow us to determine the
protein binding site sizes on single- and double-stranded DNA, and the binding rate ($k_a$) of both proteins to single-stranded regions that is formed as the DNA molecule is stretched. We determine the equilibrium association constants of these proteins to dsDNA ($K_{ds}$) as a function of salt concentration. We show that the salt dependence of dsDNA binding is similar to that previously observed for ssDNA binding (chapter3), and the four orders of magnitude salt-independent difference between ssDNA and dsDNA binding is attributed to nonelectrostatic interactions involved only in ssDNA binding. The results support our model (chapter 3) in which salt-dependent dimerization interactions must be broken for DNA binding, and further suggest that these gp2.5 monomers, formed upon breaking a gp2.5 dimer, search dsDNA by 1D diffusion to find available binding sites on ssDNA.

In chapter 5, we show that the $\alpha$ subunit of the replicative DNA polymerase III of *E. coli* has an affinity for both double- and single-stranded DNA. Our data suggest that the portion of the protein that binds to double-stranded DNA stabilizes the DNA helix, as protein binding must be at least partially disrupted to melt DNA. In addition, the single-stranded DNA binding component appears to be passive as the protein does not facilitate melting, but binds instead to regions already separated by force therefore stabilizing the single-stranded form of DNA. By studying the constructs of segments of the $\alpha$ subunit we show that the N-terminal region is responsible for dsDNA stabilization, whereas the C-terminal region binds to melted DNA suggesting that this domain may interact with ssDNA created during the DNA replication process.

In chapter 6, we characterize the interactions of stretched DNA with gene 59 protein. We show that gp59 binds more strongly to double-stranded DNA and determine the
equilibrium binding constant to dsDNA as a function of salt concentration. We also examine the effects of gp59 on the DNA duplex stability in the presence and absence of gp32 and its fragment, *I. Our results demonstrate that gp59 is capable of strongly destabilizing both gp32-DNA and *I-DNA interactions. This result confirms a previous hypothesis that the local weakening of gp32-ssDNA binding is a requirement for helicase loading at that site.

In chapter 7, we formulate the conclusions and discuss the biological importance of our studies and possible future studies.
Chapter 2

Structural Evidence of Force-Induced DNA Melting

This chapter has been accepted for publication in Biophysical Journal (Shokri L., McCauley, M. J., Rouzina I., and Williams M. C.).

2.1 Background

Optical tweezers have been used extensively for studying the mechanical properties of single biomolecules, such as DNA, by allowing the molecule to be mechanically distorted under various conditions\textsuperscript{30,33,83}. As double-stranded DNA (dsDNA) that is torsionally unconstrained is stretched, its response appears elastic due to entropy. By further stretching the dsDNA, a sudden transition occurs, as the molecule is elongated to almost twice its contour length without a strong increase in force. The resulting plateau in the force-extension curve shows the cooperative overstretching transition. Single-stranded DNA, when stretched at this force, is about 1.7 times longer than the same strand wound in the B-DNA helix, and therefore ssDNA is favored by force near the transition. In addition, the mechanical work performed on the molecule within the cycle when it is stretched as dsDNA and relaxed as ssDNA is consistent with the free energy of DNA melting\textsuperscript{53,68}. Several other characteristics of the transition are consistent with the equilibrium melting nature of this transition, including its high reproducibility, its apparent independence of the pulling rate, and the small hysteresis observed upon DNA
relaxation in high salt, which increases in lower salt. In contrast, stretching beyond the plateau leads to an abrupt increase in force, until complete strand separation occurs above 120 pN, at a force which is dependent upon the pulling rate.\textsuperscript{66,67,84}

An alternative explanation for DNA overstretching hypothesizes a secondary structure transition from B-form dsDNA into a stretched form of dsDNA termed S-DNA\textsuperscript{31,66,84,85}, in which the dsDNA unwinds and forms a ladder-like structure with the base stacking largely disrupted, but most of the inter-strand hydrogen bonding preserved, as proposed by early computer simulation studies\textsuperscript{86-88}. In this model the final strand separation does not occur until after the overstretching plateau, where stretching becomes non-equilibrium\textsuperscript{86-89}. The S-DNA model is difficult to test because experimentally there is no information known about it except that the energetics of the B to S-DNA transition is expected to be very much the same as for B-DNA melting\textsuperscript{90}. Attempts were made to deduce the elasticity and structural parameters for S-DNA from the small part of the DNA extension curve at forces above the plateau\textsuperscript{47,91}, or from the residual apparent DNA strand winding remaining after the plateau. However, because this portion of the DNA stretching curve is typically non-equilibrium\textsuperscript{84}, these parameters cannot be considered to be equilibrium properties of a new dsDNA structure.

There were several modeling studies performed in an attempt to deduce the structure of S-DNA. The early modeling studies\textsuperscript{88,92} all suggested quite different S-DNA structures that were also dependent on the strand attachment (i.e. via one DNA strand at each end, for example via 5´5´, 3´3´, or 5´3´ ends). In addition, the calculated free energy required for reaching these structures was always about an order of magnitude higher than the experimentally measured mechanical work of overstretching. An additional problem with
the modeled S-DNA structures is that no explanation for the experimentally observed high cooperativity of the overstretching transition was found within the B-to-S modeled transformation\textsuperscript{88,92}. More recent modeling studies\textsuperscript{93-95} that were able to more adequately sample the large entropy of the melted DNA state show that force-induced melting is an energetically much more favorable process than unstacking while retaining hydrogen bonding.

The major open question of the FIM model of DNA overstretching is how forces much higher than the transition plateau force, $F_m$, can be supported by just a few base pairing interactions remaining by the end of the transition. In future work, we will show that this is a kinetic phenomenon, fully consistent with the removal of most of the base pairing interactions during the transition (M. J. McCauley, L. Shokri, I. Rouzina, and M. C. Williams, in preparation). The other open question also concerns the kinetics of FIM. Indeed, a recent modeling study of this process\textsuperscript{96} suggests that the melting plateau is expected to depend significantly on the pulling rate, while experimentally it does not. As will be shown elsewhere, FIM is indeed expected to appear quasi-equilibrium, i.e. pulling rate-independent, given that large DNA fragments melt cooperatively within the continuous DNA duplex, rather than one bp after another at the duplex/single strand boundary.

The FIM model has quantitatively predicted the dependence of the overstretching transition on the solution conditions such as pH, temperature, ionic strength, and the presence of DNA binding ligands\textsuperscript{53,68}. These predictions have been tested experimentally and the results agree with the expected force-induced melting behavior. Experiments monitoring the pH and temperature dependence of the transition reveal that the base pairs
are broken as the DNA is stretched\textsuperscript{54,55}, while salt-dependent measurements are in agreement with thermal melting studies and show that the two strands remain close and stretch together\textsuperscript{50}. This implies that melting occurs primarily within the internal domains rather than from the free ends. Also, experiments in the presence of ssDNA binding proteins or DNA binding drugs resemble those obtained in thermal melting experiments, quantitatively consistent with the FIM model\textsuperscript{56,57,59-61,63,64,97}. Although there is substantial thermodynamic evidence in favor of a FIM model and no predictive and quantitative model for S-DNA has appeared, it has been suggested that S-DNA remains a possibility or that the model is under debate, although it may exist only as a metastable state or under specific solution conditions\textsuperscript{30,47,91,96,98}.

In this study we attempt to determine the extent to which the DNA base pairs are exposed to solution during the transition in order to further test the validity of the FIM model without relying on thermodynamic conclusions. To accomplish this, we have undertaken a study of λ-DNA overstretching in the presence of glyoxal, a chemical that forms a stable DNA adduct with exposed guanine residues. It introduces an additional ring to the G base, thereby sterically preventing GC base pair (bp) reannealing\textsuperscript{99}. Glyoxal has been used in the past to map DNA thermal melting\textsuperscript{100-102}. It is the use of glyoxal that first allowed researchers to prove that DNA melting proceeds via steps in which individual ∼100-500 bp segments melt out cooperatively, resulting in the peaks in differential DNA melting profiles (i.e. in the \(df/dT\) vs \(T\) curves, where \(f\) is the fraction of DNA bp melted). More specifically, when the DNA was incubated with glyoxal at some fixed temperature close to but lower than the average melting temperature, \(T_m\), only certain AT rich fragments of DNA were permanently fixed in the single stranded state.
with glyoxal\textsuperscript{100,102}. Determining the location of these sites either by single strand digestion\textsuperscript{102,103} or electron microscopy\textsuperscript{101,104,105} lead to the identification of these low-melting sites with the sites predicted by direct calculations for the particular DNA sequence\textsuperscript{101}. In this work we use the same approach in order to quantify the fraction of DNA base pairs that are exposed as DNA is overstretched. In contrast to conventional thermal melting experiments, where the fraction of DNA bp melted is quantified at each temperature by ultraviolet spectroscopy, in our single molecule study we can fix the fractional DNA melting by fixing the molecular end-

**Figure 2.1.** Schematic diagram illustrating the expected effect of glyoxal on DNA stretching and relaxation for the force-induced melting model and the S-DNA model. a) Model for DNA force-induced melting, in which the DNA base stacking and base pairing interactions are progressively broken as DNA is overstretched. b) Expected effect of glyoxal on DNA, in which exposed base pairs are modified by glyoxal, represented by spheres attached to bases in the diagram. Note that while we have shown all exposed bases modified for simplicity, it is likely that primarily guanosine bases are modified under the conditions used here. c) In the force-induced melting model, at forces below the overstretching force, shown here at approximately 30 pN, bases that were previously modified at the overstretching force remain modified, and upon relaxation remain single-stranded, resulting in a longer DNA molecule at that force relative to the length of a fully double-stranded molecule. d) In the S-DNA model, DNA base pairs are expected to remain hydrogen-bonded, but to unstuck, as shown here. e) Because S-DNA remains base-paired, glyoxal would not be expected to modify overstretched DNA in this model, except perhaps for some frayed DNA at the end of the molecule. Upon relaxation to 30 pN, base pairs previously converted to S-DNA revert to the normal B-form double helix. The length of DNA at 30 pN should be essentially the same as that observed without glyoxal treatment, if the S-DNA model is correct.
to-end extension at some given position within the FIM transition range. If the
overstretching transition is indeed FIM, then the fractional bp melting, $f$, at any extension
during this transition is defined by the relative closeness of the molecular extension to
either dsDNA or ssDNA length, as illustrated schematically in Figure 2.1. Therefore, we
expect that glyoxal exposure of DNA stretched to the fractional extension $f$ into the
plateau should result in the permanent melting of the fraction $f$ of DNA bp. The latter
effect should be observed in subsequent DNA stretches, in which the molecule is
expected to behave as a linear combination of ds $(1-f)$ and ss $(f)$ fractions of DNA. In
contrast, if the S-DNA model holds, the molecular extension of the DNA should change
very little after glyoxal treatment, as illustrated in Figure 2.1.

Below we show that indeed the majority of the elongated DNA base pairs become
permanently single stranded under the action of glyoxal. Since the glyoxal modification
is very selective for the melted DNA bases, this result demonstrates the high
accessibility of the bases in their overstretched state. It therefore argues in favor of
ssDNA nature of the overstretched DNA and against base paired S-DNA. In order to
find optimum conditions for the selective modification of the force-melted bases we
studied this reaction as a function of glyoxal concentration, time of exposure and over the
range of solution pH values between 7.5 and 9.7, and at two different salt conditions of
100 mM and 5 mM [Na$^+$]. Interestingly, we found that ssDNA fixation at 100 mM [Na$^+$]
requires high glyoxal concentrations of ~0.5 M and long glyoxal exposures of ~30 min.
The reaction with glyoxal appeared to be much more efficient in low salt, while the pH
increase had only a minor facilitating effect on this reaction. Below we discuss the
possible reasons for these salt-dependent differences in the apparent fractional DNA melting.

2.2 Direct Evidence of the Melting Nature of the Overstretching Transition

Presented in Figure 2.2 are the force vs extension curves for a typical individual $\lambda$-DNA molecule obtained at 100 mM [Na$^+$], pH 7.5 in the absence (black solid line) and presence (red solid line) of 0.5 M glyoxal. The DNA stretching and relaxation are performed in steps of 100 nm, and the force measurements at each extension are collected and averaged over 1 second thus, the processes of DNA stretching or relaxation are relatively fast and take about one to two minutes. Addition of glyoxal to solution has only a minor effect on DNA stretching curves obtained at these pulling rates. This result shows that no guanine residues have reacted with glyoxal to provide permanently melted bases on this time scale.

In order to observe the much slower permanent fixation of the melted state of DNA by glyoxal due to its binding to unpaired G bases\textsuperscript{107}, we held DNA at some particular extension within the DNA structural transition (indicated in Figure 2.2.a by red arrow) for different amounts of time, and then relaxed it. We expect that if complete fixation of melted regions were achieved, then upon relaxation and subsequent stretches the DNA molecule would exhibit stretching behavior similar to ssDNA up to the fixation point, followed by the shortened transition plateau, corresponding to FIM of the remainder of the dsDNA molecule. Indeed exposures for about 20-30 min in solution of 0.5 M glyoxal lead to significant permanent melting of dsDNA. No permanent melting of any fraction
of dsDNA was ever observed for the stretching forces below plateau, independent of the solution conditions, or the length of the glyoxal exposure time. This result is in agreement with the selective fixation by glyoxal of open DNA regions \(^{103,104,106-108}\). It provides direct evidence that the transition force does indeed separate the two DNA strands, including the loss of Watson-Crick hydrogen bonding.

However, a small fraction of the presumably force-melted bp does not become permanently fixed by glyoxal. This can be concluded based on the fact that when the long-time glyoxal-exposed DNA is gradually relaxed from its fixed position, it does not

\[\text{Figure 2.2: Effect of glyoxal on force-extension data.} \]

*Figure 2.2:* Effect of glyoxal on force-extension data. a) Stretching–relaxation curves in the absence (black) and presence of 0.5 M glyoxal (color). After introducing glyoxal into solution, the DNA molecule was stretched to a fixed position (solid red). The translation stage was held at the corresponding extension (indicated by arrow), and after 30 minutes the molecule was relax.

- Solid line: stretch
- Dashed line: relax

### 100 mM Na⁺, pH 7.5

0.5 M glyoxal

Data is taken in 10 mM HEPES pH 7.5, 100 mM [Na⁺] (95 mM NaCl and 5 mM NaOH). DNA stretching-relaxation curves in the absence of glyoxal are shown in solid and dashed black lines, respectively. The solid lines in pink and cyan are DNA models for dsDNA and ssDNA, respectively.
immediately follow a hybrid ssDNA-dsDNA force-extension curve at that extension, but
first traces back some part of the plateau, and then starts to decrease as for a slightly
shorter ssDNA-dsDNA hybrid complex. Interestingly, all subsequent rapid stretch-release
DNA cycles (shown by the blue solid and dashed lines, respectively) retrace this first
relaxation curve. This result simply indicates that no significant further glyoxal binding
occurs during subsequent rapid stretches of the same molecule.

Presented in Figure 2.2.b are DNA stretching curves obtained under the same solution
conditions as in Figure 2.2.a in the absence (black) and presence of 0.5 M glyoxal (color).
The green, red and blue-colored curves correspond, respectively, to DNA stretched
previously and held there for 30 min to 0.25, 0.50 and 0.75 fractional extension into the
plateau region. We were unable to obtain significant data by holding stretched DNA at
the end of the transition, i.e. at \( f > 0.75 \), due to breaking of the single DNA molecules.
Presumably, complete strand separation does occur under those conditions during our
long-time extension. This observation is consistent with the hypothesis that only small
fraction of DNA base pairs supports the molecule at these extensions.

We see that, as expected, the apparent fraction of permanently glyoxal-modified
DNA, \( f_a \), does increase with \( f \), the expected fractional DNA melting by force. In order to
quantify \( f_a \) for each stretching condition we fitted the pre-plateau portion of the curve to
the linear combination of pure ss and ds curves with respect to \( f_a \):

\[
b = (1 - f_a) \cdot b_{ds} + f_a \cdot b_{ss}.
\]  

(2.1)

Here, \( b(F) \), \( b_{ds}(F) \) and \( b_{ss}(F) \) are extensions per bp for DNA after permanent partial
melting, as well as for the pure ss- and dsDNA as a function of force, \( F \). The elastic
behavior of pure dsDNA was extensively characterized, and shown to be well-described by the worm-like chain (WLC) model:

$$b_{ds}(F) = b_{ds}^{\text{max}} \left[ 1 - \frac{1}{2} \left( \frac{k_B T}{FP_{ds} K_{ds}} \right)^{1/2} + \frac{F}{K_{ds}} \right],$$

(2.2)

where the extension per base pair, $b_{ds}^{\text{max}}$, is the contour length of dsDNA, while $K_{ds}$ is the elastic modulus, and $P_{ds}$ is the persistence length of dsDNA. The ssDNA elasticity is also well characterized, and shown to be well-fitted by the extensible freely jointed chain (FJC) model as:

$$b_{ss}(F) = b_{ss}^{\text{max}} \left[ \coth \left( \frac{2P_{ss} F}{k_B T} \right) - \frac{1}{2} \frac{k_B T}{P_{ss} F} \right] \left[ 1 + \frac{F}{K_{ss}} \right],$$

(2.3)

where, $b_{ss}^{\text{max}}$ is the contour length per base pair in ssDNA, and $K_{ss}$ and $P_{ss}$ are the elastic modulus and the persistence length of ssDNA, respectively. The value of $f_a$ was estimated by finding the best linear fit to Eq. (2.1) in the minimum mean-square sense. To perform this fit, we used previously measured values for $b_{ds}$, $P_{ds}$, and $K_{ds}$ as $0.34 \pm 0.001$ nm, $48 \pm 2$ nm, and $1200 \pm 200$ nm, respectively. Although, there are known dependencies of these parameters upon ionic strength and pH, changes in these values do not affect fitting to Eq. (2.1) significantly. On the other hand, the FJC parameters of ssDNA, $b_{ss}$, $P_{ss}$, and $K_{ss}$, may become quite different upon glyoxal binding. Therefore, we first fitted DNA stretching curves for which the length of permanently glyoxal-melted DNA was known to Eq. (2.1) with respect to FJC parameters of Eq. (2.3). Values for ssDNA have been found previously ($b_{ss} = 0.56$ nm, $P_{ss} = 0.75$ nm and $K_{ss} = 800$ pN). These values for $b_{ss}$ and $K_{ss}$ are held constant, though the value for $P_{ss}$ has been altered to $0.9 \pm 0.1$ nm, which is still characteristic of unmodified ssDNA, and likely reflects less hairpin
formation at small extensions than in reference\textsuperscript{31}. This result is in agreement with the observations of the authors of reference\textsuperscript{103}, who found that the glyoxal-bound ssDNA on electron microphotographs were indistinguishable within error from pure thermally melted ssDNA. We then used these elastic ssDNA parameters obtained at 100 mM [Na\textsuperscript{+}] and pH 7.5 to fit the apparent fractional DNA melting for all other solution conditions. As discussed in\textsuperscript{50,55} we do not expect changes in ssDNA parameters under the range of solution conditions studied to significantly affect our conclusions.

<table>
<thead>
<tr>
<th>Fractional DNA melting by force, $f$</th>
<th>Permanently glyoxal-modified ssDNA fraction, $f_a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>pH 7.5 100 mM [Na\textsuperscript{+}] 0.5 M Glyoxal</td>
<td>pH 8.5 100 mM [Na\textsuperscript{+}] 0.5 M Glyoxal</td>
</tr>
<tr>
<td>0.25 ± 0.01</td>
<td>0.22 ± 0.03</td>
</tr>
<tr>
<td>0.5 ± 0.01</td>
<td>0.33 ± 0.01</td>
</tr>
<tr>
<td>0.75 ± 0.01</td>
<td>0.44 ± 0.05</td>
</tr>
</tbody>
</table>

\textbf{Table 2.1.} Permanently glyoxal-modified ssDNA fraction, $f_a$. Data are reported as mean ± standard error for three or more molecules.

The resultant fractions of permanently melted DNA, $f_a$, fitted for every explored $f$ are summarized in Table 2.1. The DNA force-extension curves used for these fits obtained by averaging multiple stretching curves for at least three individual DNAs for each fixation length are presented by dashed lines in Figure 2.3. The four panels of Figure 2.3 represent results obtained at four different solution conditions, as indicated in the Figure. We see that at low $f = 0.25$ practically all elongated bp become permanently single stranded upon glyoxal exposure under all solution conditions tested, i.e. $f_a \approx f$. As expected, $f_a$ grows as $f$ increases to 0.5 and 0.75. However, in the latter cases, $f_a$ was only
0.33 ± 0.01 and 0.44 ± 0.05, respectively (see Table 2.1). In other words, about 70% and 60% of elongated bp become permanently single stranded in each case.

Interestingly, in our experiments the apparent fractional DNA melting $f_a$ was never observed to exceed its fractional elongation $f$. This result is consistent with the high selectivity of glyoxal fixation of the bases. In the case of thermal DNA melting, this selectivity was shown to be due to the heterogeneity of DNA bp stability. Thus, at some temperatures below average $T_m$ certain AT rich segments are predominantly opened, and fixed by glyoxal, while the rest are still closed$^{101,104,105}$. It is the extremely slow rate of the glyoxal/guanine covalent bonding that makes this reaction practically non-observable for hydrogen bonded base pairs$^{107}$. Indeed, formaldehyde, which is analogous to glyoxal, but a much faster ssDNA fixation agent, was shown to be able to unwind any DNA duplexes even at room temperature$^{26,110-113}$. The increasing temperature exponentially facilitates the rate of glyoxal-promoted unwinding of DNA duplex$^{103}$. Therefore, selective fixation of the low-melting segments requires very specific glyoxal exposure conditions$^{100,101,105}$. In the case of force-induced melting the force becomes an equivalent of the temperature, while the fractional extension into the plateau region, $f$, is an equivalent of fractional DNA melting$^{53}$. Our glyoxal exposure experiments are conducted under the condition of the fixed fractional DNA melting, rather than fixed force. This is why even in the homogeneous DNA only the force-melted base pairs are expected to be fixed. Indeed, fixation in the ss state of the additional bases would lead to the abrupt drop of the force, which in turns, should lead to the exponential slowing of the further guanine fixation by glyoxal.
Gradual thermal opening of the subsequently more stable regions in λ-DNA upon temperature increase was previously monitored by glyoxal fixation and mapped on the molecule by ethidium fluorescence\textsuperscript{103}. At both 0.17 and 0.3 fractional DNA melting, most of the glyoxal-fixed regions were several hundred bp long, and appeared to be rather close to the molecule’s center. A smaller melted fragment was also found at one of the ends of the molecule. The FIM of λ-DNA in our experiments, most likely, happens within the same regions, as DNA is progressively stretched into the transition, in contrast to the conclusions of reference\textsuperscript{96}, in which it was proposed that DNA can be melted by force exclusively from its free ends.

2.3 Effect of Solution Conditions on the $f_a$

In the attempt to understand why the apparent DNA melting remains smaller than the fractional DNA extension, we used the longer glyoxal exposure times. However, the time of our experiments are limited by the drift of the optical tweezers instrument to a few hours. During this time no further increase in apparent binding was determined. We further tested the effect of the glyoxal reaction rate on $f_a$ by varying the solution pH from 7.5 to 9.7. It is known that in this pH range the rate of glyoxal/G binding increases by about 10-fold, while the binding remains essentially irreversible\textsuperscript{107}. Quantification of the permanently melted DNA fraction at 100 mM [Na\textsuperscript+\textsuperscript{] and three different pH conditions is summarized in Table 2.1. The data presented in Table 2.1 and Figure 2.3 (a-c) suggest that $f_a$ for a given $f$ grows with pH very insignificantly over this pH range.
Figure 2.3. Mapping DNA force-induced melting in the presence of glyoxal. Stretching (solid line) and relaxation (dash line) curves are shown in the absence (black) and presence of glyoxal (color). In color are the stretch and relaxation curves after DNA has been overstretched for some time (30 minutes for 100 mM [Na+] and 10 minutes for 5 mM [Na^+] in the presence of glyoxal at the corresponding extensions (indicated by arrows). For each relaxed curve we averaged the relaxation curves of three or more molecules, and error bars are determined from the standard error. We used only the stretches of DNA molecules with the least number of nicks by excluding all the stretches from DNA molecules that broke at some point during the overstretching transition. Data is taken in a) 10 mM HEPES pH 7.5, 100 mM [Na^+] (95 mM NaCl and 5 mM NaOH), and 0.5 M glyoxal; b) 10 mM HEPES pH 8.5, 100 mM [Na^+] and 0.5 M glyoxal; c) 10 mM HEPES pH 9.7, 100 mM [Na^+] and 0.5 M glyoxal; d) 10 mM HEPES pH 7.5, 5 mM [Na^+] (5 mM NaOH) and 0.2 M glyoxal.
In contrast, \( f_a \) is always significantly higher and very close to \( f \) in 5 mM monovalent salt solution. In addition, higher \( f_a \) in low salt can be achieved at lower glyoxal concentrations of 0.2 M, and with shorter exposure times of \( \sim 10 \) min, instead of 30 min, (compare the average stretching curves in Figure 2.3.a and d. This more extensive reaction with glyoxal in lower salt can hardly be explained by faster glyoxal binding. Indeed, glyoxal binding to G bases is known to be salt-independent in terms of both its kinetics and equilibrium binding constant\(^{107,108}\), as expected for a neutral and nonpolar molecule. Therefore, the reason for the apparent more extensive glyoxal interaction with DNA observed in lower salt should be searched for in the difference in DNA behavior, not in its glyoxal binding.

Based on these data, it is likely that the low vs. high salt situations are different not by the number of guanine residues modified with glyoxal, but rather by the fact that in high salt some modified force-melted ssDNA regions can still partially re-anneal upon DNA relaxation. Indeed, the rate of G/glyoxal reaction is, most likely, still much slower\(^{107}\) than our exposure time, such that only a fraction of all exposed G residues become modified under all of the solution conditions explored in this work. As was shown in reference\(^{99}\), re-annealing of partially glyoxal-fixed DNA sequences is similar to re-annealing of the slightly mismatched ssDNA strands. Such re-annealing occurs at temperatures lower than the \( T_m \) of the perfectly complementary strands, which depends on the fraction of modified or mismatched bases. In low salt, however, these imperfect duplexes become unstable, with the apparent \( T_m \) below room temperature. Indeed, low salt leads to a higher energetic penalty for the ds/ss boundaries\(^{18}\), as well as to the lower stability of every bp\(^{114}\). Consistent with this interpretation, at 100 mM [Na\(^+\)] the residual melting force plateau
after glyoxal exposure becomes slightly lower than the original $F_m$ in the absence of glyoxal. At the same time, no residual plateau is observed in the DNA relaxation curve after glyoxal exposure in 5 mM [Na$^+$], consistent with all force-exposed bp remaining permanently single stranded (see also Table 2.1).

A secondary effect that could also contribute to the low and high salt differences in $f_a$ is the following. DNA melting partially occurs from its free ends by peeling off of the strand that is not under tension. In high salt (100 mM [Na$^+$]) this strand will, most likely, form its own intra-strand secondary structure in the form of hairpins, in which G residues will be protected. Even though the G residues in its complementary DNA strand will be permanently modified upon glyoxal exposure, such protection, will, quite likely, lead to more DNA regions being able to re-anneal upon DNA relaxation, thereby leading to a smaller apparent fraction of permanently glyoxal-modified DNA. In contrast, in low salt most of the ssDNA hairpins will become unstable and unfold (for the reasons mentioned above), leading to modification of the G residues in these regions. This phenomenon can also explain why a smaller fraction of all bp melted by stretching become permanently modified by glyoxal as the DNA is stretched further into the plateau (see Table 2.1). Indeed, while the least stable AT-rich regions melt first upon DNA stretching in the middle of $\lambda$-DNA, further DNA stretching leads to more melting from the ends, and therefore more hairpin-protected bp, and less permanent melting. These features are specific for the $\lambda$-DNA sequence, and therefore will likely vary with the DNA sequence studied.
2.4 Conclusions

In summary, in this study, we have shown that a majority of the DNA base pairs elongated by force during the overstretching transition become permanently single-stranded when exposed for a sufficient time to glyoxal. Since glyoxal is known to covalently bind only solution-exposed guanine bases\textsuperscript{104,106-108}, this result provides direct evidence that the elongated base pairs have lost their Watson-Crick hydrogen bonding, i.e. that they are melted. According to all modeling studies, most of the inter strand base pairing is expected to be preserved in S-DNA\textsuperscript{86-89}. Indeed, the length of the overstretched DNA is close to the length of the stretched single DNA strand. This implies that independently of the details of the proposed S-DNA structure, most of the stacking interactions in it should be destroyed, while the double stranded structure should rely on the inter-strand hydrogen bonding. Thus our results strongly argue in favor of the melting nature of the overstretching transition in DNA, and against a double-stranded form of overstretched DNA, i.e. S-DNA.

In addition we have shown that just as for the thermal melting, in force-induced melting the glyoxal is an appropriate agent to bind and permanently fix exclusively the ss regions, while not unwinding the double stranded regions of DNA. This result is reinforced by the specific setup of our experiment in which the fractional DNA melting by force is controlled (as opposed to the control of the force or temperature). Such set-up ensures that the permanent fixation is achieved for only the equilibrium force-melted base pairs, since any additional glyoxal binding leads to the abrupt drop of the force associated with the exponential drop in already extremely slow G/glyoxal reaction rate. This makes our approach for the studies of force-melting of DNA with glyoxal reliable and
quantitative. We have also shown that this selective glyoxal fixation of force-melted DNA remains valid in the broad range of solution pH between 7.5 and 9.7, and for the half to a few hours of glyoxal exposure time at room temperature and ~0.5 M glyoxal concentrations. According to this study, the apparent glyoxal-fixed fraction of DNA is a better measure of fractional force-melted DNA in low (5 mM [Na⁺]) rather than in high (100 mM [Na⁺]) salt, due to the partial re-annealing of the modified DNA strands in the latter case.
Chapter 3

Salt Dependent Binding of Phage T7 Gene 2.5 Protein to Single-stranded DNA

*Parts of this chapter are taken from a previously published manuscript in Journal of Biological Chemistry*\(^6\) (Shokri L., Marintcheva B., Richardson C. C., Rouzina I., and Williams M. C.).

### 3.1 Background

Genomic replication is a complex process comprised of numerous interconnected steps, each of which exhibits a high degree of complexity and involves many proteins properly assembled at the DNA replication fork. The replication system of bacteriophage T7 is similar to those of higher organisms. T7 has been used extensively in molecular biology research and provides excellent model system for analyzing the DNA replication. T7 phage encodes most of its own replication proteins, therefore bypassing the host replication machinery.

The genetic material of phage T7 is a linear dsDNA with about 39,930 nucleotides per strand\(^{115}\). T7 replication can be initiated at the origin and proceeds bidirectionally very fast due to its efficient and economical replication system. The T7 DNA replication system shown in Figure 3.1 requires few proteins: The T7 gene 5 protein (gp5), a DNA polymerase (80 kDa) that forms a tight 1:1 complex with its processivity factor, *E. coli*
Figure 3.1. The bacteriophage T7 replication fork. The bacteriophage T7 replisome consists of the hexameric T7 helicase-primase (gp4), two copies of the T7 polymerase (gp5 complexed with its processivity factor, *E. coli* thioredoxin (trx)), and the T7 ssDNA-binding protein (gp2.5) that coats the transiently exposed ssDNA in the replication loop. Based on T7 replication forks figure from http://richardson.med.harvard.edu

Thioredoxin functions as a clamp to facilitate polymerase processivity. It swings across the DNA-binding groove and encircles the primer-template exiting the polymerase.

The T7 gene 2.5 protein (gp2.5), encoded by gene 2.5 of the bacteriophage T7, is a single stranded binding protein containing 232 residues\textsuperscript{120}. It facilitates the annealing of complementary strands of DNA more efficiently than T4 phage gp32 and *E. coli* SSB\textsuperscript{124-126}. It physically interacts with both T7 DNA polymerase and the T7 helicase-primase\textsuperscript{121,127-130} and plays multiple roles in T7 DNA replication and recombination\textsuperscript{121,126,131-136} and is considered to be critical for establishing coordinated leading and lagging strand DNA synthesis\textsuperscript{137-139}. In the absence of DNA, gp2.5 forms a
stable homodimer in solution. Its nucleic acid and protein interactive properties are strongly dependent on the domain structure of the protein.\textsuperscript{120,127-129,140-142}

The presence of an acidic C terminus is a common feature of all prokaryotic and mitochondrial ssDNA-binding proteins, which have been shown to modulate the DNA binding properties of the proteins and mediate protein-protein interactions.\textsuperscript{143} The crystal structure of gp2.5\textsuperscript{144} reveals a core that consists of a conserved OB-fold (oligosaccharide/oligonucleotide-binding fold) that is well adapted for interactions with ssDNA. The OB fold consists of invariant aromatic residues surrounded by positively charged amino acids. While aromatic residues stack with the DNA bases, the positively charged amino acids contact the phosphate backbone.\textsuperscript{129,145-149}

**Figure 3.2.** A proposed model for the functional role of the phage T7 gp2.5 acidic C terminus. The C-terminal tail of gp2.5 interacts with the DNA-binding core of the protein and functions as a biological switch effectively protecting the positively charged DNA-binding cleft from binding to random negatively charged surfaces, thus coordinating the ssDNA binding, protein–protein interactions, and multiple reactions occurring at the replication fork. For clarity, the figure represents only one monomer. Based on Fig. 6 from Marintcheva B. \textit{et al}.\textsuperscript{146}

Gene 2.5 protein has a highly acidic C-terminal tail.\textsuperscript{150} The C-terminus contains 21 amino acids of which 15 have acidic side chain, separated by non-polar amino acids. It ends with an aromatic residue Phenylalanine (F) that is essential for the function of gp2.5 and can add another negative charge due to its free carboxyl group. C-terminus is required for dimer
formation and for interactions with T7 DNA polymerase and the helicase-primase\textsuperscript{127,129,146,151-153}. It is shown\textsuperscript{151} that C-terminus can bind to the positively charged subdomain of the T7 gene 5 DNA polymerase, which is a site of binding of the thioredoxin and the C-terminus of gene 4 helicase-primase. Models have been proposed in which the C-terminal tail of wild type protein interacts with the DNA-binding core of the protein\textsuperscript{78,144,146}, and functions as a biological switch effectively protecting the positively charged DNA-binding cleft from binding to random negatively charged surfaces, thus coordinating the ssDNA binding, protein–protein interactions, and multiple reactions occurring at the replication fork (Figure 3.2). It is shown that the T7 helicase-primase also has an acidic C-terminal tail with a functionally important C-terminal aromatic residue\textsuperscript{154}. This common feature can be considered as additional evidence that they are part of a regulatory mechanism involving competition for binding to the same site\textsuperscript{146}. A genetically altered protein, gp2.5-Δ26C, lacks the C-terminal 26 residues. gp2.5-Δ26C binds ssDNA more tightly than does the full-length protein\textsuperscript{120}.

3.2 T7 gp2.5 Exhibits Behavior Similar to T4 gp32

Previous works of studying the interactions of bacteriophage T4 gene 32 protein (gp32) to ssDNA determine the association constant of gp32 to ssDNA over a range of salt concentrations, including physiological conditions\textsuperscript{59-61}. It is found that the association constant of wild type gp32 exhibited very little salt dependence below 200 mM [Na\textsuperscript{+}], in contrast to bulk experiments, which had previously shown a very strong salt dependence in high salt. A C-terminal truncation of gp32, denoted *I, does not show salt-independent binding at low salt. It was proposed that the lack of salt dependence to wild type gp32-
ssDNA binding in low salt was due to a conformational change involving the acidic C-terminal domain of gp32, which may be required for gp32-ssDNA binding\textsuperscript{75,97}, and because this conformational change begins to strongly alter gp32 binding near physiological salt concentrations, the C-terminal domain of gp32 may act as a regulatory switch.

The differences in complexity between these two systems (see section 3.1 and 6.1) suggest the possibility of more sophisticated coordination between the DNA binding properties of the single-stranded binding protein from T7, gp2.5, and the protein-protein interactions with other T7 replication proteins. Therefore, we conducted a single molecule study of salt-dependent ssDNA binding by the single-stranded DNA-binding protein from bacteriophage T7 in order to examine the differences between the DNA binding activities of T4 gp32 and T7 gp2.5.

We utilized DNA stretching to investigate the effects of full-length gp2.5 as well as gp2.5-Δ26C on DNA duplex stability and melting. In accord with the known preferential binding to ssDNA, both proteins reduce the melting force. However, the differences between force-extension curves during stretching and relaxation, (i.e. stretching hysteresis), in the presence of both proteins indicate that the system does not reach equilibrium on the time scale of our experiment. The technical limitations of our instrument made it difficult to pull slowly enough to reproduce the complete equilibrium DNA stretching curve. Therefore, we measured the equilibrium force at the midpoint of the DNA melting transition, at the extension corresponding to the melting of half of the base pairs. To do this, we rapidly stretched the dsDNA to a fixed end-to-end extension halfway between the dsDNA and ssDNA contour lengths and monitored the force, which
converged to equilibrium at this given extension, on the time scale of several minutes. We then calculated the equilibrium binding constants of the protein to ssDNA from the protein concentration dependence of the measured equilibrium DNA melting force\textsuperscript{53,60,68}. We found that the binding affinity of the full-length gp2.5 and its C-terminal truncation for exposed ssDNA regions were salt-dependent and differed by several orders of magnitude in low salt. To explain this difference in binding affinity, we developed a model in which a salt-dependent dimerization interaction must be broken in order for gp2.5 to bind to DNA. We then quantified this binding interaction as a function of salt concentration; herein we discuss its relevance to gp2.5 interactions with other T7 replication proteins.

### 3.3 Quantifying ssDNA Binding Affinity

To probe the effect of these proteins on duplex DNA destabilization, we measured the force-extension curve of \(\lambda\)-DNA in the presence and absence of gp2.5 and gp2.5-\(\Delta26C\), over a range of salt and protein concentrations. Typical results of our measurements in 50 mM [Na\(^+\)] buffer are shown in Figure 3.3. In the absence of protein, the DNA melting force is independent of pulling rate\textsuperscript{59} and shows very little hysteresis. gp2.5 (7 \(\mu\)M) and gp2.5-\(\Delta26C\) (300 nM) significantly lower the melting force relative to that observed in the absence of protein. However, higher concentrations of gp2.5 compared with gp2.5-\(\Delta26C\) are required to observe a significant reduction in the melting force. Although this behavior is expected for a protein that binds preferentially to ssDNA, this is the first observation of natural DNA denaturation in the presence of gp2.5. When the DNA was relaxed, the relaxation curves did not match the stretching curves, as shown by the dashed
Figure 3.3. Single-stranded DNA binding affinity of gp2.5 and gp2.5-Δ26C. λ-DNA stretching (solid line)-relaxation (dashed line) curves in 10 mM HEPES pH 7.5, 50 mM [Na+] (45 mM NaCl, 5 mM NaOH) in the absence of protein (black) and in the presence of 7 µM gp2.5 (red) and 300 nM gp2.5-Δ26C (blue).

The considerable hysteresis shows that protein dissociation from exposed regions of ssDNA and subsequent DNA reannealing upon relaxation is slower than the 4-min duration of a stretching and relaxation cycle. In addition, lower rates of DNA stretching resulted in lower melting forces, signifying that the protein-ssDNA association and dissociation were slow on the time scale of the force-induced DNA melting.

Ideally, if we were able to pull slowly enough, we should be able to measure the complete equilibrium DNA stretching curve in the presence of the "slow" protein. However, our optical tweezers instrument does not allow pulling slower than 5-10 nm/s because of a position drift on longer time scales. Instead, we could measure the equilibrium force at the midpoint of the DNA melting transition. To this end, we rapidly stretched the dsDNA to a fixed end-to-end extension halfway between the dsDNA and ssDNA contour length and monitored the force, which converges to an equilibrium value, over a time scale of several minutes. This approach was used previously for equilibrium studies of T4 gene 32 protein interactions with DNA. Typical results obtained in 25 mM [Na+] buffer are given in Figure 3.4.
After rapid stretching to the transition midpoint, the force decreased exponentially with time, representing additional DNA melting due to protein binding. The time-dependent force data can be fit to the following relation,

\[
F(t)_{\text{stretch}} = F_m + (F_k - F_m) \exp\left(-\frac{t}{\tau_{\text{melt}}}\right),
\]

where \(F_m\) is the equilibrium force obtained at long times, \(F_k\) is the initial melting force (the subscript \(k\) indicates that this kinetically determined force likely depends on pulling rate), and the constant melt is the time for DNA melting due to protein binding. To measure the equilibrium force during the DNA renaturation, the DNA molecule was first stretched through the DNA melting transition and then relaxed quickly to the midpoint of the transition, and the change in force in the presence of each protein was monitored for at least 15 min. The observed force increased exponentially, representing DNA reannealing due to protein dissociation. The time-dependent force data upon relaxation were fit to an expression analogous to Eq. (3.1). The fact that both denaturation and renaturation
experiments converge to the same force, denoted by $F_m$, clearly demonstrates that the melting force obtained in the experiment outlined in Figure 3.4 represents the equilibrium DNA melting force.

As illustrated in Figure 3.5, increasing amounts of either protein results in a lower DNA melting force, reflecting progressive protein-induced duplex destabilization. Below, we use these data to determine the protein-DNA association constants of both proteins. The shift in the DNA melting temperature due to protein binding can be related to the protein-DNA binding constants to dsDNA and ssDNA as follows\(^\text{111}\) (see section 1.7):

$$\frac{1}{T_m^0} - \frac{1}{T_m} = \frac{k_B}{\Delta H} \ln \left[ \frac{(1 + K_{ds}C)^{n_{ds}}}{(1 + K_{ss}C)^{n_{ss}}} \right], \quad (3.2)$$

where $n_{ss}$, $n_{ds}$, $K_{ss}$, and $K_{ds}$ are the binding site size in nucleotides and binding constants of the protein to ssDNA and dsDNA, respectively, and $C$ is the bulk protein concentration. The change in DNA melting temperature and force due to duplex destabilization by protein can be related via general thermodynamics by the relation:

$$\Delta G_{\text{destabilization}} = \left( F_m - F_m^0 \right) \Delta x = \left( T_m - T_m^0 \right) \Delta S, \quad (3.3)$$

where $\Delta S$ and $\Delta x$ are the difference in entropy and extension per base pair between the protein-bound ssDNA and dsDNA, $T_m^0$, $T_m$, $F_m^0$, and $F_m$ are the melting temperature and force in the absence and presence of protein, respectively, and $\Delta G_{\text{destabilization}}$ is the change in DNA melting free energy per base pair induced by protein binding. Combining Eqs. (3.2) and (3.3), using the relationship $\Delta H = \Delta S T_m^0$, and taking into the account that gp2.5 is a single-stranded binding protein, i.e. that $K_{ss} >> K_{ds}\text{120}$, we can find the shift in melting force due to protein binding as follows.
By fitting the melting force as a function of concentration to Eq. (3.4), we obtain a measurement of $K_{ss}$ under the desired solution conditions. Below, we use this to examine the salt dependence of gp2.5 and gp2.5-$\Delta$26C association to ssDNA.

Figure 3.5. Salt-dependant equilibrium melting force. a) The measured DNA equilibrium melting force as a function of protein concentration, $F_m(C)$, for gp2.5. Measurements are shown in 5 mM [Na$^+$] (filled green diamond), 25 mM [Na$^+$] (filled blue square), and 50 mM [Na$^+$] (filled red circle). b) The measured DNA equilibrium melting force as a function of protein concentration, $F_m(C)$, for gp2.5-$\Delta$26C. Measurements are shown in 25 mM [Na$^+$] (filled blue square), 50 M [Na$^+$] (filled red circle), 75 mM [Na$^+$] (filled violet triangle), and 100 mM [Na$^+$] (filled lime-green diamond). Lines were fitted to data using Eq. (3.4) and $\chi^2$ analysis. Each data point was obtained by averaging over three or more measurements, and the error bars were determined from the S.D. of those measurements.

3.4 Salt-Dependant Binding of T7 gp2.5 and T7 gp2.5-$\Delta$26C to ssDNA

Fitting the $F_m(C)$ data presented in Figure 3.5 to Eq.(3.4), we found $K_{ss}$ for each protein at different salt concentrations. To perform this fit, we used our measured difference in length per base pair $\Delta x$ between the dsDNA and ssDNA from the stretching curves for
Figure 3.6. Salt-dependant equilibrium binding constant to ssDNA ($K_{ss}$). The measured dependence of logarithm of the binding constant ($K_{ss}$) to ssDNA as a function of logarithm of salt concentration for gp2.5 (red diamond) and gp2.5-Δ26C (blue square) and the linear fit to the data for gp2.5 (red line) and gp2.5-Δ26C (blue line). The brown circle represents binding of gp2.5 to poly(dT) DNA at 50 mM NaCl using a fluorescence-based study by Kim et al.\textsuperscript{120}, and the navy triangle represents binding of gp2.5-Δ26C to 70-base oligonucleotides at 50 mM KCl using an electrophoretic mobility shift by Hyland et al.\textsuperscript{140}.

As shown in Figure 3.6, the C-terminal truncation of gp2.5 binds ssDNA about 100-fold stronger than the wild type gp2.5 in 100 mM [Na\textsuperscript{+}] salt, in accord with previous data\textsuperscript{117,140}. However, previous data were limited to a single measurement of $K_{ss} = 1.2 \times 10^6$ M\textsuperscript{-1} at 50 mM NaCl for gp2.5 by fluorescence quenching\textsuperscript{120} compared with our measured value of $(3.5 \pm 0.6) \times 10^5$ M\textsuperscript{-1} at the same salt concentration. Similarly,
previous data were limited to a single measurement of $K_{ss} = 3 \times 10^7$ M$^{-1}$ at 50 mM KCl for gp2.5-Δ26C by electrophoretic mobility shift assay\textsuperscript{140}, which compares well with our measured value of $(7.8 \pm 0.9) \times 10^7$ M$^{-1}$ in 50 mM [Na$^+$] solution. For the electrophoretic mobility shift assay, we used the value obtained for binding to 70-base oligonucleotides as determined from the "slow mobility" lane of the gel. For this lane, the measured wild type gp2.5 value for $K_{ss}$ agrees well with that measured in the earlier fluorescence quenching assay\textsuperscript{120,140}.

Our measurements show that the ssDNA association constant of gp2.5-Δ26C depends much more strongly on salt than does that of wild type gp2.5, such that the stronger binding of the truncated protein becomes exaggerated in lower salt. Specifically, the total number of ions released upon protein binding, given by $n_{Na^+} = -d \log (K) / d \log ([Na^+]^+)$, is equal to $n_{Na^+} = 1.1 \pm 0.1$ for wild type and $2.5 \pm 0.2$ for the deletion mutant. Given that the deletion mutant, gp2.5-Δ26C, differs from the wild type protein, gp2.5, only by the absence of the last 26 C-terminal residues, it is surprising that the two binding constants differ so strongly.

3.5 Conclusions

Using single molecule DNA FIM in the presence of gp2.5 and gp2.5-Δ26C, we have determined the binding affinity of both proteins as a function of salt concentration. We find that in low salt, the binding affinity of gp2.5-Δ26C for ssDNA exceeds that of gp2.5 by 2-3 orders of magnitude. In addition, the binding affinities of both proteins are salt-dependent, and the salt dependence of gp2.5-Δ26C binding to DNA is much stronger. Because gp2.5-Δ26C lacks only the C terminus of gp2.5, this result shows that the
presence of the C terminus on gp2.5 reduces its affinity for DNA. The C terminus of
gp2.5 is flexible and may extend away from the β-barrel of the OB-fold into solution.

**Figure 3.7.** Schematic diagram of the model for gp2.5 dimer formation, which must be
broken prior to DNA binding. a) For gp2.5-Δ26C, the DNA binding site is always available
for DNA binding because of the lack of the C-terminal tail. b) C-terminal tails stabilize the
gp2.5 dimer by their "domain swap," i.e. binding to the basic regions of its gp2.5 partner. The
gp2.5 cationic binding site for its partner's C-terminal tail is also its ssDNA binding site.
Therefore, gp2.5 is monomeric in its DNA-bound state and dimeric in its unbound state. c, a
conformational change in gp2.5 involving dimer dissociation is required prior to gp2.5
nucleic acid binding.

However, acidic and aromatic residues of the C-terminal tail make it an ideal mimic of
ssDNA. Therefore, one possible mechanism for this reduction in binding affinity is the
electrostatic binding of the C terminus to the DNA binding site. If this were the case, the
C terminus would compete with ssDNA for binding to the cationic DNA binding site,
which would result in weaker binding to DNA in lower salt relative to a protein lacking
the C terminus. Such a salt-dependent DNA binding regulation mechanism was recently
reported for bacteriophage T4 gene 32 protein, the ssDNA-binding protein for T460,97.
However, gp2.5 is known to form a dimer in solution, whereas gp2.5-Δ26C has not been
observed to form dimers in solution; thus any model for gp2.5 binding must take this
dimerization interaction into account.
Based on the lack of dimerization for gp2.5-Δ26C in solution, it has been suggested previously that the C-terminal segments stabilize the dimer form of gp2.5 by a domain swapping interaction\textsuperscript{155} across the dimer interface\textsuperscript{142,144}, as illustrated in Figure 3.7.b. This model resembles the model described above for electrostatic regulation of DNA binding by the C terminus, but in this case DNA binding requires dimer dissociation by which the C terminus is removed from the DNA binding site of each protein. This interaction is more complex than that suggested for T4 gene 32 protein, because the dimer formation likely involves more interactions than just the C termini binding to the DNA binding sites. In particular, there is an additional dimer interface that contributes to the overall dimerization interaction and that also must be disrupted for DNA binding by gp2.5 to occur\textsuperscript{78}. In the case of T4 gp32, a C-terminal truncation, *I, showed a strong salt dependence at low salt, yielding a value of $n_{Na^+} = 3$, similar to that observed here for gp2.5-Δ26C. In contrast, wild type T4 gp32 showed very little salt dependence, with $n_{Na^+} = 0$. This is similar to that observed for wild type gp2.5. However, in the case of T4 gp32, the association constants for both *I and wild type gp32 converged at 200 mM NaCl. This suggests that above 200 mM NaCl, the C-terminal domain of gp32 spends most of its time in solution. In turn, this result reveals that the binding of the gp32 C-terminal domain to its DNA binding site, which regulates its DNA binding, is an entirely electrostatic effect that is screened at salt concentrations above 200 mM NaCl. In contrast, the results on gp2.5 and gp2.5-Δ26C reported here show that the inhibition of binding by the gp2.5 C-terminal tail persists even at 1 M NaCl. Therefore, the interactions that are responsible for the difference in gp2.5 and gp2.5-Δ26C binding must have a nonelectrostatic component, in contrast to the interactions that determine the
difference between T4 gp32 and *I binding. The additional requirement for breaking the gp2.5 dimer interface, discussed below, is sufficient to explain this difference between the salt dependence of T4 gp32-DNA binding relative to the DNA binding of T7 gp2.5.

Although our data alone cannot prove the above model of gp2.5 dimerization and ssDNA binding, our results are fully consistent with the model and provide further support for it. There are two likely scenarios for gp2.5 binding. One possibility is that the C terminus of one monomer is removed from the DNA binding site and exposed to solution, whereas the other monomer binds DNA. A second possibility is that the dimer dissociates, with both C-terminal tails removed from the DNA binding pocket of the other protein. Below, we used our data on gp2.5 and gp2.5-Δ26C binding to ssDNA over a broad range of salt concentrations to quantitatively probe the salt-dependent gp2.5 interaction that regulates gp2.5 binding to DNA. We show that this interaction consists of two components: electrostatic, representing the release of the C-terminal tail, and non-

Figure 3.8. The free energy of dimerization measured as a function of the logarithm of salt concentration. Using the binding constants of gp2.5 and gp2.5-Δ26C to ssDNA in 25 and 50 mM [Na+] buffer, we determined the values of \( \Delta G_{\text{dimer}} \) directly. By extrapolating the salt-dependent data for each protein separately, we were able to calculate \( \Delta G_{\text{dimer}} \) in other salt concentrations. The arrow shows the value of the non-electrostatic component of \( \Delta G_{\text{dimer}} \).
electrostatic, which likely represents dissociation at the dimer interface. A schematic diagram of the model describing these interactions is shown in Figure 3.7. The fact that we see a significant nonelectrostatic component to the interaction energy that regulates gp2.5 binding supports a model in which a breaking of the dimer interface is required for gp2.5 binding. However, without additional evidence we cannot completely exclude nonelectrostatic interactions between the C-terminal tail and the DNA binding site, which might allow for very weak gp2.5 dimerization when bound to DNA.

If the DNA binding of gp2.5 differs from gp2.5-Δ26C binding only because of a requirement of dimer dissociation, and the dimer interaction is in pre-equilibrium to DNA binding, then the binding affinity of gp2.5 compared with that of gp2.5-Δ26C is reduced by the probability of dimer dissociation, \( P_{\text{dimer}} \),

\[
K_{ss}^{\text{gp2.5}} = K_{ss}^{\text{gp2.5-Δ26C}} \times P_{\text{dimer}},
\]

where \( K_{ss}^{\text{gp2.5}} \) and \( K_{ss}^{\text{gp2.5-Δ26C}} \) are the association constants of gp2.5 and gp2.5-Δ26C to ssDNA, respectively, and equation as follows:

\[
P_{\text{dimer}} = \frac{e^{\Delta G_{\text{dimer}}/k_B T}}{e^{\Delta G_{\text{dimer}}/k_B T} + 1}.
\]

Here \( \Delta G_{\text{dimer}} \) is the attractive (negative) free energy of gp2.5 dimerization per dimer, which can be obtained by solving Eqs. (3.5) and (3.6).

\[
\Delta G_{\text{dimer}} = -k_B T \ln \left( \frac{K_{ss}^{\text{gp2.5-Δ26C}}}{K_{ss}^{\text{gp2.5}}} \right).
\]

The gp2.5 dimerization free energy as a function of salt in the range of 5 to 100 mM [Na\(^+\)], obtained according to Eq. (3.7) using our measured binding constants for both proteins, is presented in Figure 3.8. We find that \( \Delta G_{\text{dimer}} \) is strongly salt-dependent, in agreement with the significant contribution to gp2.5 dimerization from the
electrostatically driven C-terminal tail (CTT) swap between gp2.5 monomers. Indeed, the 
electrostatic similarity between the CTT and the region of ssDNA that binds gp2.5 
implies that gp2.5 dimerization should be almost as salt dependent as binding of gp2.5- 
Δ26C to ssDNA. An analogous salt-dependent CTT unbinding from the ssDNA binding 
groove of T4 gene 32 protein was described recently\textsuperscript{60,97}. Quantitatively, this means that 
one might expect \( d \Delta G_{\text{dimer}} / d \ln ([Na^+]^-) \) to be similar in magnitude and opposite in sign 
to \( d \ln \left(K_{\alpha \alpha}^{\text{gp2.5-Δ26C}}\right) / d \ln ([Na^+]^-) \). Experimentally, we find that \( d \Delta G_{\text{dimer}} / d \ln ([Na^+]^-) \)
\( \cong 1.4 \pm 0.1 \) and \( -d \ln \left(K_{\alpha \alpha}^{\text{gp2.5-Δ26C}}\right) / d \ln ([Na^+]^-) \) \( \cong 2.5 \pm 0.1 \). This result suggests that the 
CTT is highly charged but somewhat less charged than the section of ssDNA that binds 
gp2.5.

In ~1 M salt, all electrostatic interactions are significantly screened by salt ions, and 
the remaining interactions reflect their non-electrostatic salt-independent component\textsuperscript{22}. 
Therefore, extrapolating \( d \Delta G_{\text{dimer}} ([Na^+]^-) \) to 1 M [Na\(^+\)], we obtained an estimate of the 
non-electrostatic component of the gp2.5 dimerization free energy \( \Delta G_{\text{dimer}}\cong (-1.2 \pm 0.8) 
k_BT = 0.7 \pm 0.4 \) kcal/mol. This weak dimer interaction explains why gp2.5-Δ26C does not 
form dimers in solution but crystallizes as a dimer\textsuperscript{144}. The CTT swap makes gp2.5 
dimerization progressively stronger in lower salt. However, even at the physiological salt 
concentration of ~100 mM [Na\(^+\)], the dimerization free energy is moderate, at 
approximately -5 k_B T and -3 kcal/mol, and should not completely preclude thermal dimer 
dissociation, which is required for gp2.5 binding to ssDNA.

In a previous study, Rezende \textit{et al.}\textsuperscript{78} examined gp2.5, gp2.5-Δ26C, and several 
mutants at the dimerization interface by gel filtration in 50 mM KPO\(_4\) pH 7.0, 0.1 mM
EDTA, 0.1 mM dithiothreitol, and 10% glycerol buffer at a single protein concentration. Several salt concentrations were examined: 150, 250, and 500 mM NaCl. gp2.5-Δ26C was reported to be a monomer at all salt concentrations, and gp2.5 was reported to be a dimer at 150 and 250 mM and a monomer at 500 mM NaCl. The mutants at the dimerization interface were reported to be dimers at 150 mM NaCl and monomers at 250 and 500 mM NaCl. These results are consistent with our model of gp2.5 dimerization in which most of the dimerization free energy comes from the electrostatic interactions of the anionic C-terminal tails of the protein with the cationic ssDNA binding site on the dimerization partner. Although dimerization becomes weaker with increasing salt concentration, it also requires interactions at the dimer interface. The fact that the gel filtration study suggests that dimer dissociation occurs at 0.5 M NaCl, in contrast to our prediction of a marginally stable dimer at 1 M NaCl, is likely because of the lower protein concentration used in the gel filtration study.

According to previous studies\textsuperscript{121,127-129} the CTT of gp2.5 is known to interact with several components of the T7 replication fork. Therefore, it is possible that these interactions regulate CTT-induced gp2.5 dimerization, thereby controlling gp2.5 binding to ssDNA at the replication fork. Such control may have an important regulatory role, by preventing the extensive gp2.5 binding and accompanying melting of double-stranded DNA that is not involved in replication. Alternatively, exposure of the C terminus upon gp2.5 binding to DNA could itself regulate the activities of other proteins at the replication fork. Further studies of gp2.5 interactions with DNA and other replication proteins are needed to distinguish between these possibilities and elucidate the critical interactions between components of the model T7 DNA replication machinery.
In summary, we have quantitatively determined the affinity of gp2.5 and gp2.5-Δ26C binding to DNA as a function of salt concentration. We find that gp2.5 binding is regulated by electrostatic interactions involving the C terminus, in addition to a weak non-electrostatic binding component. We have presented a model in which protein dimerization involving the C terminus regulates DNA binding by gp2.5. Our model suggests that, although the dimerization interaction is primarily electrostatic, there is a weak non-electrostatic component, which likely represents interactions at the dimer interface.
Chapter 4

Kinetics and Thermodynamics of Salt-Dependent T7 gene 2.5 Protein Binding to Single- and Double-stranded DNA

Single Molecule Stretching Measurements are done by Leila Shokri at Department of Physics, Northeastern University. Protein Purification is done by Boriana Marintcheva at Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School. This work is done in collaboration with Prof. Charles C. Richardson form the Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, and Prof. Ioulia Rouzina from the Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota.

4.1 Background

In our previous work\textsuperscript{62}, we utilized DNA stretching to study the effect of both proteins on DNA duplex stability and melting. Both proteins were observed to lower the DNA melting force. The observed decrease in the DNA melting force indicates that the binding ligand destabilizes the DNA helix\textsuperscript{51,53,68}. To quantify this helix-destabilization, we previously determined the equilibrium DNA melting force in the presence of protein at very long times (\textasciitilde 20 minutes)\textsuperscript{62}. The equilibrium melting force was then used to determine equilibrium binding constants of these proteins to ssDNA as a function of salt concentration. We observed a several order of magnitude difference between the salt
dependent binding affinity of full length gp2.5 and its genetically altered mutant in low salt. Our previous results suggest that, while gp2.5 binding to ssDNA is regulated by strong electrostatic interactions involving the C-terminus, there is also a weak non-electrostatic interaction at the dimer interface. We developed a model in which a dimeric gp2.5 must dissociate in order to bind to ssDNA\textsuperscript{62}. According to our model, the dimer dissociation consists of a weak non-electrostatic and a strong electrostatic component. Recently, Marintcheva et al. showed that the gp2.5 CTT competes for the same binding surface as ssDNA, consistent with the results from our single molecule measurements\textsuperscript{146}.

In the current study, we determine the effect of pulling rate on the DNA melting force in the presence of full length gp2.5 and its deletion mutant. The strong hysteresis observed upon stretching and relaxing the DNA molecule suggests that this experiment does not occur in equilibrium. It has previously been shown that measurements of a non-equilibrium rate-dependent DNA melting force can be used to investigate the kinetics of the process responsible for DNA melting\textsuperscript{59}. To investigate whether or not this method can be applied to the current system, we stretched and relaxed single dsDNA molecules in the presence of both proteins at a constant pulling rate. From this data, we determine the association rate, $k_a$, for gp2.5 binding to ssDNA.

Our measurements of gp2.5 association kinetics show that the gp2.5-ssDNA association rate is enhanced by 1D sliding of the protein along dsDNA, which in turn suggests that the rate of DNA melting is regulated by protein binding to dsDNA. Based on a simple model for 1D diffusion of the protein along dsDNA, we determine the equilibrium association constant of both proteins to dsDNA, $K_{ds}$, as a function of salt concentration for the first time. Both proteins show a salt dependence to dsDNA binding.
that parallels that observed previously for ssDNA binding. The equilibrium association constant of gp2.5 and gp2.5-Δ26C to dsDNA found in this work is almost four orders of magnitude lower than the measured equilibrium association constant to ssDNA. Taken together, these results suggest that the salt-dependent dimerization mechanism previously proposed for ssDNA binding is also responsible for regulating gp2.5 binding to dsDNA. In addition, gp2.5 binding to ssDNA is enhanced by significant nonelectrostatic binding interactions that result in increased affinity of the protein for ssDNA relative to dsDNA at all salt concentrations.

4.2 Rate dependence of DNA melting force in the presence of gp2.5

We used DNA stretching to probe the effect of gp2.5 and gp2.5-Δ26C on the DNA melting force as a function of pulling rate. Stretching curves for a single λ-DNA molecule in the absence or presence of gp2.5 and gp2.5-Δ26C are shown in Figure 4.1.a and b, respectively. In both cases, the presence of the protein reduces the DNA melting force. However, to observe considerable reduction in the overstretching force, higher concentrations of gp2.5 compared to gp2.5-Δ26C are required. As the dsDNA molecule is pulled at different rates of \( \nu = 5-250 \) nm/s, the molecule extends to the B-form contour length and then begins to melt at the particular force \( F_k(\nu) \), (where the subscript \( k \) indicates that this kinetically determined force is likely to depend on pulling rate \( \nu \)). In the absence of protein, the DNA melting force is independent of the pulling rate and shows very little hysteresis. However, in the presence of both gp2.5 and gp2.5-Δ26C, the melting force is significantly lowered and moreover depends on the pulling rate. The hysteresis observed in the release part of the stretching cycle clearly demonstrates the
non-equilibrium nature of the DNA melting by gp2.5 and its deletion mutant. The observed non-equilibrium DNA melting force is determined by the rate of protein binding to ssDNA during duplex melting. Therefore, this force is different from the equilibrium DNA melting force that was studied in our previous work. Our work here is analogous to dsDNA thermal melting studies. However, in this case the melting force varies linearly with the natural logarithm of the pulling rate, while in thermal melting studies the DNA melting temperature varies linearly with the natural logarithm of the heating rate.

At any given pulling rate $\nu$, the effect of gp2.5-$\Delta 26C$ on $F_a(\nu)$ is larger than the effect of gp2.5. The unwinding forces decrease as the amount of either protein increases, reflecting progressive protein-induced duplex destabilization. Over the entire range of

Figure 4.1. Rate-dependant DNA stretching cycle. a) Stretching (solid line) – relaxation (dashed line) curves in 10mM HEPES pH 7.5, 50 mM [Na$^+$] (45 mM NaCl and 5 mM NaOH) in the absence of protein (black) at a pulling rate of 250 nm/s and in the presence of 30 $\mu$M gp2.5 at pulling rates of 250 nm/s (red), 100 nm/s (green), 25 nm/s (blue), and 5 nm/s (light blue). b) Stretching (solid line) – relaxation (dashed line) curves in 10 mM HEPES pH 7.5, 100 mM [Na$^+$] (95 mM NaCl and 5 mM NaOH) in the absence of protein (black) at a pulling rate of 250 nm/s and in the presence of 530 nM gp2.5-$\Delta 26C$ at pulling rates of 250 nm/s (red), 100 nm/s (green), 25 nm/s (blue), and 5 nm/s (light blue).
pulling rates, the non-equilibrium DNA melting force $F_k(\nu)$ is always smaller than the equilibrium DNA melting force in the absence of protein, $F_m^0$, and is larger than or equal to the equilibrium melting force in the presence of protein, $F_m^p$, that was previously studied in the same solution conditions.$^{62}$

In previous studies of the bacteriophage T4 gene 32 protein, a protein that dissociates very slowly from ssDNA and has a strong ssDNA binding cooperativity$^{59,60}$, the non-equilibrium DNA melting force $F_k(\nu)$ was found to be larger than the equilibrium melting force in the presence of protein, $F_m^p$. This was true even for the lowest possible pulling rate of $\nu = 5$nm/s and for all concentrations of gp32. In that case, it was not possible to pull slowly enough to observe the saturation of $F_k(\nu)$ to the equilibrium melting force $F_m^p$ as the pulling rate $\nu$ was decreased. In contrast, full length gp2.5 and its deletion mutant both exhibit much faster binding kinetics, such that it is possible to pull slowly

![Figure 4.2: Measured non-equilibrium DNA melting force as a function of pulling rate. Data are shown in the absence of protein (black diamond), in the presence of 10 $\mu$M gp2.5 (red square), 20 $\mu$M gp2.5 (green triangle), 30 $\mu$M gp2.5 (blue circle), 230 nM gp2.5-$\Delta$26C (pink square), 300 nM gp2.5-$\Delta$26C (light green triangle), and 460 nM gp2.5-$\Delta$26C (cyan circle). Linear fits are shown as continuous lines. Each data point is obtained by averaging 3 or more measurements and error bars are determined from the standard error. Data is taken in 10 mM HEPES pH 7.5, 50 mM [Na$^+$] (45 mM NaCl and 5 mM NaOH).](image-url)
enough (5-25 nm/s) to observe force saturation (Figure 4.1). This is likely due to gp2.5’s non-cooperative or weakly cooperative binding to ssDNA.

Below, we use our experimental results to determine the association kinetics of gp2.5 and gp2.5-Δ26C binding to ssDNA. As discussed in previous work\textsuperscript{59,60}, rate-dependent DNA melting occurs when the pulling rate is equal to the rate at which thermal fluctuations cause a certain number of base pairs to open, and these exposed regions of ssDNA are captured by protein binding,

\[ v = N_b n_{ss} \Delta x \frac{k_a}{S n_{ss}} , \quad (4.1) \]

where \( n_{ss} \) is the binding site size in nucleotides and \( \Delta x \) is the increment in length per base-pair of protein-bound ssDNA relative to dsDNA. Thus, \( n_{ss} \Delta x \) is the length released upon a single protein binding event. \( N_b \) is the number of helix/coil boundaries in the

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure4.3.png}
\caption{Protein association rate as a function of protein concentration. a) gp2.5 association rate in 5 mM salt (red diamond), 25 mM salt (green triangle), 50 mM salt (blue circle). b) gp2.5-Δ26C association rate in 25 mM salt (green triangle), 50 mM salt (blue circle), 100 mM salt (brown square). Lines are fit to the data using McGhee & von Hipple isotherm. Dashed lines show the three dimensional (3D) diffusion limit.}
\end{figure}
Figure 4.4. Salt-dependant equilibrium binding constant to ssDNA and dsDNA. The measured dependence of logarithm of the equilibrium association constants to ssDNA ($K_{ss}$) and dsDNA ($K_{ds}$) as a function of logarithm of salt concentration for gp2.5 (red square and red triangle, respectively) and gp2.5-$\Delta^{26}C$ (blue circle and blue diamond, respectively). The linear fits to the data are shown for both proteins in continuous and dashed lines. The open pink square represents binding of gp2.5 to poly(dT) DNA at 50 mM NaCl using a fluorescence-based study by Kim et al.\textsuperscript{120}, and the open cyan circle represents binding of gp2.5-$\Delta^{26}C$ to 70-base oligonucleotides at 50 mM KCl using an electrophoretic mobility shift by Hyland et al.\textsuperscript{140}. The error in measurements are small, roughly the size of the symbols. ssDNA binding results are taken from the previous DNA stretching measurements, Shokri. et al.\textsuperscript{62}

DNA molecule. As long as the melting force is significantly lower than $F_m^0$, and protein binding is slow relative to base-pair opening fluctuations in the middle of the DNA molecule, DNA melting occurs primarily from the ends of the molecule, so $N_b \approx 2$ (A. Hanke, L. Shokri, I. Rouzina, and M. C. Williams, in preparation). Here $k_a$ is the rate of the single protein finding its contiguous binding site, which is created by the melting of $n_{ss}$ base-pairs at the boundary between the dsDNA and protein-covered ssDNA.

The statistical weight $s = e^{-\Delta G/k_bT}$ is the DNA base-pair stability. Thus, the factor $k_a n_s^s$ in Eq. (4.1) is the rate at which the melting of $n_{ss}$ base pairs occurs along with subsequent protein binding. Because we apply a force to the ends of the DNA molecule, the stability of each base pair becomes a function of the force. For our melting forces, which are always greater than 20 pN, we
can use the linear approximation $\Delta G = \Delta G^0 - F \Delta x$ \cite{59}. Here, $\Delta G^0$ is the extrapolated free energy of the DNA helix-coil transition per base-pair in the absence of force and protein. Note that in Ref.\cite{59} we defined $\Delta G^{nc}$ after a small correction to $\Delta G^0$, which represented an additional equilibrium free energy change related to rapid, noncooperative DNA destabilization by protein. However, there is no evidence that gp2.5 exhibits such a rapid binding equilibrium state, so we do not consider such a state in the model presented here. By substituting the DNA base-pair stability and the linear approximation of the DNA melting free energy into Eq. (4.1), and solving for $F_k(\nu)$, we obtain:

$$F_k(\nu) = F_m^0 + \frac{k_B T}{n_s \Delta x} \ln \left( \frac{k_v}{k_a} \right),$$

where

$$k_v = \frac{V}{N_b n_s \Delta x}. \tag{4.3}$$

It follows from Eq. (4.2) that for a fast enough protein association, when $k_a > k_v$, protein binding decreases the non-equilibrium DNA unwinding force, i.e., $F_k(\nu) < F_m^0$. As the pulling rate becomes faster, i.e., as $k_v \equiv k_a$, the unwinding force reaches the equilibrium DNA melting force in the absence of protein, i.e., $F_k \equiv F_m^0$. At this fast pulling rate, the protein has no time to affect the force-induced DNA melting. As the melting rate decreases, the transition force decreases until at $k_v \equiv k_d$ it becomes equal to the equilibrium DNA melting force in the presence of protein, i.e., $F_k(\nu) \equiv F_m^p$. At this point, the melting force $F_k(\nu)$ saturates at $F_m^p$ and stops changing as a function of pulling rate $\nu$. We measured the DNA melting force $F_k(\nu)$ as a function of $\nu$ in the
presence of both proteins over a range of salt concentration of 5-50 mM [Na\(^+\)] for gp2.5 and 25-100 mM [Na\(^+\)] for gp2.5-Δ26C. Eq. (4.2) predicts that the DNA melting force \( F_k(\nu) \) exhibits a linear dependence on \( \ln(\nu) \).

Measurements of \( F_k(\nu) \) vs. \( \ln(\nu) \) in 50 mM [Na\(^+\)] in the absence and presence of both proteins are shown in Figure 4.2. In the absence of protein, the DNA melting force depends weakly on the pulling rate. In contrast, in the presence of both proteins, the observed linear dependence is consistent with the prediction of Eq. (4.2).

Because \( \Delta x \), the extension per base pair, is known from the stretching curves, we could obtain the number of nucleotides of ssDNA that bind to the full length gp2.5 or its deletion mutant from the slope of our measured \( F_k(\nu) \) vs. \( \ln(\nu) \).

We used the stretching curves for ssDNA in the presence of gp2.5-Δ26C in a buffer containing 5 mM [Na\(^+\)] to determine \( \Delta x \), which is not expected to depend significantly on ionic strength\(^{68}\). From the slopes of the lines shown in Fig. 4.2, we found \( n_{ss} \) to be 7 ±
1 for gp2.5 and 6 ± 1 for gp2.5-Δ26C, which is in agreement with previous measurements of these quantities\textsuperscript{120}.

We obtained the measurements of \( k_a \) for each \( F_k(\nu) \) vs. \( \ln(\nu) \) data set by considering the condition \( k_\nu \equiv k_a \) or \( F_k(\nu) \equiv F_m^0 \). These measurements of \( k_a \) are shown in Figure 4.3.a and b. Here, \( k_a \) is a function of the protein concentration \( C \), and exceeds the 3D diffusion limit (see section 1.8) for almost all conditions studied with gp2.5-Δ26C, given by

\[
\text{k}_{\text{diff}} = 4\pi DR = \frac{2k_B T}{3\eta} \approx 10^9 \text{ M}^{-1}\text{s}^{-1},
\]

where \( R \) is the protein size, estimated as 1 nm, \( D = \frac{k_B T}{6\pi \eta R} \) is the 3D diffusion coefficient, and \( \eta \) is the solution viscosity\textsuperscript{59,60,158}. The 3D diffusion limit is not exceeded for gp2.5 under these conditions. While there are several possible models describing protein translocation along nucleic acids, such as inter-segment transfer, hopping and sliding\textsuperscript{29}, because the DNA molecule is straightened out by force and the protein binding kinetics are fast\textsuperscript{59}, sliding is likely the best model to explain this rate enhancement. Therefore, we assume that the proteins bind to dsDNA non-cooperatively and weakly, such that they can slide on dsDNA (while they are still bound to dsDNA) until they can find a specific binding site between dsDNA and the protein-bound ssDNA. It has been shown that for the condition \( k_d < k_a \), proteins find the binding site primarily via facilitated 1D motion along the DNA molecule, which means that most of the time proteins find their binding sites before dissociating from the dsDNA\textsuperscript{59,159,160}. In this case, the rate of binding to ssDNA is given by\textsuperscript{59}:

\[
k_{a,1D} = \left( \frac{2\theta}{n_d} \right)^2 \cdot k_s, \tag{4.4}
\]
where $k_s = 10^7 \text{ s}^{-1}$, is the conventional 1D sliding rate on dsDNA\textsuperscript{161}, $n_{ds}$ is the protein binding site size on dsDNA in nucleotides, and $\Theta$ is the fraction of dsDNA bases bound by protein described by the McGhee \& von Hippel isotherm\textsuperscript{26} (see section 1.6):

$$\Theta = K_{ds} \cdot n_{ds} \cdot C \frac{(1-\Theta)^{n_{ds}}}{\left(1 - \Theta \Theta/n_{ds}\right)^{n_{ds}-1}}.$$ \hspace{1cm} (4.5)

Here, $K_{ds}$ is the non-equilibrium association constant for protein binding to dsDNA. Fits of $k_s(C)$ for gp2.5 for $5 \text{mM}<[Na^+]<50 \text{mM}$ and for gp2.5-Δ26C for $25 \text{mM}<[Na^+]<100 \text{mM}$ are shown in Figure 4.3.a and b. The main fitting parameters are the equilibrium non-cooperative binding constants to dsDNA ($K_{ds}$) and binding site size to dsDNA ($n_{ds}$). We obtained a measurement of $n_{ds}$ and $K_{ds}$ as a function of salt concentration for the same solution conditions used previously to determine $K_{ds}$\textsuperscript{62}. As expected, $n_{ds}$ did not vary significantly with salt concentration and from our fitting, we found $n_{ds} = 5 \pm 1$ for gp2.5 and $n_{ds} = 7 \pm 2$ for gp2.5-Δ26C. These values for $n_{ds}$ are very close to the values found above for $n_{ss}$. Although the 3D diffusion limit is not exceeded for g2.5, this model still fits the data well for that protein. This suggests that the 3D diffusion limit is in fact exceeded for gp2.5, but the concentration used to calculate this rate would need to be replaced by the effective concentration of monomers available for DNA binding, which is significantly reduced by the dimerization interaction described in our earlier work\textsuperscript{62}.

The $K_{ss}$ measurement for both proteins as a function of salt concentration are shown in Figure 4.4, and are compared with previous measurements of $K_{ss}$\textsuperscript{62,120,140}. The
association constant for gp2.5-Δ26C binding to dsDNA is 100 times greater than that of gp2.5 to dsDNA in low salt. From Eqs. (4.4) and (4.5), we find that \( k_a \propto K_{ds}^{-2} \).

Therefore, for gp2.5-Δ26C the rate of finding ssDNA by 1D diffusion is \( 10^4 \) times greater compared to gp2.5. This shows that dsDNA binding is likely a significant factor in the regulation of gp2.5 activity \( \textit{in vivo} \). This dsDNA binding is itself regulated by dimer dissociation, discussed below. gp2.5 binding to dsDNA shows very little salt dependence, with a \( \log (K) / \log ([Na^+]^{1/2}) \) slope of approximately 0.3, while gp2.5-Δ26C has a stronger salt dependence with a slope of 1.4. The salt dependences of both proteins are similar to that which was observed for binding to ssDNA\(^6^2\), such that the inhibition of dsDNA binding by the gp2.5 C-terminal tail (CTT) persists even at 1 M NaCl. This suggests that the interactions responsible for the difference in gp2.5 and gp2.5-Δ26C binding to both ssDNA and dsDNA have a non-electrostatic component. The current results provide additional evidence supporting our model of dimer dissociation\(^6^2\), and provide additional quantitative information on the free energy of the gp2.5 dimerization per dimer (\( \Delta G_{\text{dimer}}^{ss,ds} \)).

The gp2.5 dimerization free energy can be determined directly from our experimental data of \( K_{ss,ds}^{gp2.5} \) and \( K_{ss,ds}^{gp2.5-\Delta26C} \) in the range of 5 to 100 mM \([Na^+]\) as follows\(^6^2\):

\[
\Delta G_{\text{dimer}}^{ss,ds} = -k_B T \cdot \ln \left( \frac{K_{ss,ds}^{gp2.5-\Delta26C}}{K_{ss,ds}^{gp2.5}} - 1 \right).
\] (4.6)

The results are presented in Figure 4.5. The measured dimerization free energy shows a strong salt dependence, which is in agreement with the significant contribution to this free energy of electrostatic interactions of the gp2.5 DNA binding domain\(^6^2\). By extrapolating the free energy of the gp2.5 dimerization per dimer to 1 M \([Na^+]\), we
obtained an estimate of the nonelectrostatic component of the gp2.5 dimerization free energy equal to \((-1.2 \pm 0.4) k_B T = (0.7 \pm 0.2) \text{kcal/mol}\), which is in excellent agreement with the value we previously obtained by studying the gp2.5 binding to ssDNA using a different method\textsuperscript{62}. For comparison, our results for ssDNA binding from previous work are also presented in Fig. 5.

### 4.3 Conclusions

In this manuscript, we use rate-dependent measurements of DNA force-induced melting to determine the kinetics and thermodynamics of T7 gene 2.5 protein’s interactions with both dsDNA and ssDNA. By using a simple model in which bound proteins are captured as ssDNA is exposed due to increased thermal base pair opening under an applied force, we determine the rate at which proteins bind to exposed ssDNA as a function of the rate at which the DNA is stretched. By performing these rate measurements as a function of protein concentration, we determine the concentration dependence of the protein binding rate. We find that this dependence is not linear in protein concentration, and that it exceeds the 3D diffusion rate expected for this reaction. We therefore conclude that the rate at which individual gp2.5 and gp2.5-Δ26C proteins find their ssDNA binding sites is due primarily to 1D diffusion along dsDNA. The concentration dependence of the binding rate is consistent with this model. By applying the model for 1D diffusion, we determine the equilibrium association constant for gp2.5 and gp2.5-Δ26C binding to dsDNA. We then repeat this measurement over a range of salt concentrations, allowing us to determine $K_{d}$ as a function of salt concentration.
By combining two types of DNA stretching measurements, we have determined the equilibrium association constants of gp2.5 and gp2.5-Δ26C to both ssDNA and dsDNA as a function of salt concentration, as shown in Fig. 4. We observe that gp2.5-Δ26C shows stronger binding to both ssDNA and dsDNA at all salt concentrations below 1 M, and that gp2.5-Δ26C binding to DNA is orders of magnitude stronger in low salt. These measurements support our previously proposed model for salt-dependent regulation of gp2.5 binding to DNA by the requirement for salt-dependent dimer dissociation prior to DNA binding by the protein. The observed salt dependence of this interaction is consistent with the fact that it is mediated by a highly acidic C-terminal domain that gp2.5-Δ26C lacks.

The present study sheds light on several new aspects of gp2.5-DNA interactions. First, we find that both ssDNA and dsDNA binding have very similar salt dependence, which suggests that the same dimer dissociation mechanism is required for binding to both dsDNA and ssDNA. We also find that, while the salt dependences of both types of binding are similar, ssDNA binding is stronger than dsDNA binding by a factor of $(7 \pm 2) \times 10^3$, averaged over the measured salt concentrations. The ratio of binding constant to ssDNA and dsDNA $(K_{ss}/K_{ds})$ for both proteins shows only a very weak dependence on salt concentration. The ratio of these binding affinities represents the salt-independent difference between gp2.5’s dsDNA and ssDNA binding affinity, which is similar for both gp2.5 and gp2.5-Δ26C. This difference in binding affinities is likely due to nonelectrostatic interactions that can only be formed with ssDNA. We can use this information to calculate the nonpolyelectrolyte part of the protein-ssDNA binding free energy, $\Delta G_{ned} = k_B T \ln(K_{ss}/K_{ds})$, which is $8.5 \pm 0.4$ $k_B T = 4.9 \pm 0.2$ kcal/mol. This is
comparable to the nonpolyelectrolyte ssDNA binding free energy measured for cooperative binding of T4 gene 32 protein\textsuperscript{97}.

Finally, it is useful to compare the properties of T7 gp2.5 to those previously measured for T4 gp32. There are several important similarities between gp2.5 and gp32. Both proteins have a highly acidic C-terminal region that regulates its dsDNA and ssDNA binding, and this regulation is strongly electrostatic and salt-dependent. Both wild type proteins exhibit only very weakly salt-dependent binding in low salt. Removal of the acidic domain reveals a truncated protein that exhibits much stronger dsDNA and ssDNA binding. In both cases, the kinetics of ssDNA binding by the protein is strongly enhanced by noncooperative binding to ssDNA and facilitated by fast 1D sliding. There are also several important differences between the two proteins. The most obvious difference is that gp2.5 forms a dimer in solution through its C-terminal tail, and the dimerization interaction is responsible for the salt-dependent regulation of the DNA binding by the acidic C-terminal tail. In contrast, gp32 binding is regulated by self-association of the C-terminal domain to the protein monomer. On the other hand, the resulting salt-dependent behavior of both proteins is similar. One other major difference between gp32 and gp2.5 is that gp32 is known to bind highly cooperatively, with a cooperativity parameter of at least $\omega = 1000$, such that gp32 binding next to other gp32 monomers is enhanced by a factor of 1000\textsuperscript{162}. Because gp32 binding to dsDNA is noncooperative, its cooperative binding to ssDNA enhances its ssDNA binding relative to that of dsDNA by the same factor of 1000. Therefore, based on our previous measurements of gp32 binding to dsDNA, we found that the intrinsic binding of gp32 to ssDNA was greater than that to dsDNA by only a factor of $10^{60}$. It is then somewhat surprising that in this study we find
gp2.5 binding to ssDNA to be on the order of $10^3 - 10^4$ times greater than its binding to
dsDNA.

If gp2.5 binding is indeed only weakly cooperative, then all of the additional binding
affinity to ssDNA relative to dsDNA must come from interactions between gp2.5 and
ssDNA. It is possible that the additional intrinsic gp2.5 binding affinity relative to gp32 is
due to the presence of additional DNA-interactive residues in the accessible DNA
binding site. Alternatively, the binding sites of gp32 and gp2.5 may have similar numbers
of interacting residues, but some of these interactive residues may only be available upon
cooperative binding, perhaps requiring a conformational change to make them accessible.
In any case, the ssDNA-specific interactions that we have quantified here are likely
critical for gp2.5’s functions at the T7 replication fork. In addition, our results also show
that dsDNA binding by gp2.5 is weak but important because it likely facilitates rapid
binding of gp2.5 to exposed ssDNA that is created at the T7 replication fork.
Chapter 5

Double- and Single-stranded DNA Binding of the Catalytic α Subunit of E. coli Replicative DNA Polymerase III

Protein Purification is done by Leila Shokri at DNA Damage Recognition and Tolerance Laboratory at Department of Chemistry and Chemical Biology, Northeastern University. Single Molecule Stretching Measurements are done by Dr. Micah J. McCauley at Department of Physics, Northeastern University, and Computer simulations are done by Dr. Česlovas Venclovas at Laboratory of Bioinformatics, Institute of Biotechnology, Vilnius, Lithuania. This work is done in collaboration with Prof. Penny J. Beuning from the Department of Chemistry and Chemical Biology, Northeastern University.

5.1 Background

Cellular DNA replicases are asymmetric dimers that efficiently and accurately copy millions of base pairs of DNA in every cell cycle. The E. coli replicase DNA polymerase III is a complex of 10 subunits that tightly coordinate leading and lagging strand synthesis\(^{163}\). The polymerase subunit of this complex is the 1160-residue α protein, encoded by the dnaE gene\(^{163}\). E. coli Pol III contains a separate 3′-5′ exonuclease proofreading protein, ε, that together with the θ subunit is tightly bound to the α subunit\(^{163}\). The β processivity clamp encircles DNA, binds the Pol III α subunit, and facilitates highly processive DNA replication by Pol III\(^{163}\). The clamp loader is a dynamic
complex that includes combinations of the $\gamma$, $\delta$, $\delta'$, $\tau$, $\chi$, and $\psi$ subunits and has roles in loading the $\beta$ clamp onto DNA, dimerization of polymerase core, and polymerase recycling on the lagging strand$^{163-165}$.

DNA polymerases are classified as members of the A, B, C, D, X, and Y families$^{166}$. *E. coli* Pol III $\alpha$ protein is a member of the C family$^{166,167}$. In many Gram negative bacteria, including *E. coli*, DNA pol III is responsible for genomic DNA replication$^{163}$. Protein sequence analysis indicated that *E. coli* Pol III $\alpha$ subunit has little or no detectable sequence similarity to eukaryotic or archaeal polymerases of known structure$^{168}$. A recent crystal structure of residues 1-917 of *E. coli* $\alpha$ reveals that the protein adopts a right-hand shape typical of DNA polymerases$^{79}$. The *Thermus aquaticus* full-length Pol III crystal structure shows that this DNA polymerase is similar overall to that of *E. coli*$^{169}$. One of the major differences between the structures of Pol III and those of many other DNA polymerases is that the fingers domain, which binds the incoming nucleotide, is much larger in Pol III than in most polymerases$^{79}$. Three conserved acidic residues in the palm domain are responsible for polymerase activity. The palm domain of Pol III is similar to that of X family member Pol $\beta$$^{79,158}$. The thumb, fingers, and palm domains form a deep cleft that could bind DNA, although the only available structure of *E. coli* Pol III lacks DNA substrates$^{79}$. The extreme N-terminus of Pol III harbors a Polymerase and Histidinol Phosphatase (PHP) domain, which may be involved in pyrophosphatase or exonuclease activity$^{79}$. The PHP domain of *Thermus thermophilus* has been shown to exhibit Zn$^{2+}$-dependent 3'-5' exonuclease activity$^{170}$, however the metal ion-binding residues are not conserved in the *E. coli* PHP domain, suggesting a different function for this domain in *E. coli* Pol III$^{169}$.  

97
*E. coli* Pol III α subunit has at least two distinct domains that may be important for binding DNA. A ‘helix-hairpin-helix’ (HhH) motif initially predicted in the α subunit by sequence analysis\(^{171}\), is a widespread motif involved in non-sequence-specific binding of either ds- or ss-DNA. Crystal structures of the Pol III α subunits of *E. coli*\(^{79}\) and *T. aquaticus*\(^{169}\) confirm the presence of the HhH motif slightly N-terminal to the internal β-processivity clamp binding motif (Figure 5.1). In the *E. coli* structure this motif is annotated as part of the ‘fingers’ domain (the ‘little finger’)\(^{79}\), and in that of *T. aquaticus* as part of the β clamp binding domain\(^{169}\). However, the HhH motif in both structures may be considered as part of a distinct \((\text{HhH})_2\) domain formed by two consecutively duplicated HhH motifs\(^{172}\). \((\text{HhH})_2\) domains are present in a majority of HhH-containing proteins and provide a symmetric way of binding to dsDNA, as in DNA polymerase β\(^{172}\). HhH domains are also known to mediate protein-protein interactions\(^{173}\). The second putative DNA binding domain in the α subunit is an oligonucleotide/oligosaccharide binding (OB) fold domain\(^{174,175}\). The OB-fold domain is located near the C-terminus\(^{79,168,169,176}\), which is not present in the *E. coli* structure. OB-fold domains are functionally diverse, as they are involved in binding oligonucleotides, oligosaccharides, or metal ions, and also in mediating protein-protein interactions\(^{176,177}\).

To investigate DNA-binding properties of the *E. coli* Pol III α subunit and its individual regions, we first modeled the three-dimensional structure of the OB-fold domain and found that it is consistent with ssDNA binding. Next, we used single molecule DNA stretching experiments to show that the full-length *E. coli* Pol III α subunit binds both double-stranded DNA and single-stranded DNA. Further, we show that the dsDNA binding domain maps to the N-terminal 917 residues that harbor the
(HhH)$_2$ domain, whereas the C-terminal 182 residues including the OB-fold domain bind pre-formed ssDNA without actively melting the DNA. Therefore, this domain may interact with ssDNA created by another process during DNA replication, such as with the template strand or ssDNA created during proofreading.

5.2 Homology Modeling of the OB-fold Domain of *E. coli* Pol III α Subunit

Known structures related to the putative OB-fold domain of *E. coli* Pol III α subunit were identified by searching with PSI-BLAST$^{178}$ in the non-redundant protein sequence database taken from NCBI (ftp://ftp.ncbi.nih.gov/blast/db/). A number of matching OB-fold domains could be reliably detected including anticodon binding domains of tRNA synthetases, ssDNA binding proteins and the RecG “wedge” domain. However, these relatives share at most only ~20% identical residues with the OB-fold domain of *E. coli* Pol III α subunit. At this level of sequence similarity (often referred to as the “twilight zone”) the accuracy of a homology model to a large degree is determined by the correctness of underlying sequence-structure alignment$^{179}$. Therefore, modeling of the OB-fold domain was performed using a procedure in which the alignment is refined iteratively by building and assessing 3D models corresponding to different alignment variants. Briefly, reliable alignment regions are first delineated with the PSI-BLAST-ISS tool$^{180}$ and only the alignment in unreliable regions is adjusted. Models corresponding to different alignment variants are then built automatically with Modeller$^{181}$ and assessed both with Prosa2003$^{182}$ and visually. Iterations are repeated until Prosa2003 energy scores cannot be improved any further and the visual assessment reveals no significant
flaws in the modeled structure. Previously, this procedure proved to be successful both in a blind-mode test and in detecting sequence-structure mapping errors in OB-fold domain crystal structures during large scale analysis.

Figure 5.1: Catalytic α subunit of E. coli replicative DNA polymerase III. (top) Domain architecture of α subunit; PHP - Polymerase and Histidinol Phosphatase domain, polymerase core denotes palm, thumb and fingers domains, HhH – helix-hairpin-helix DNA-binding motif, β-binding motif – internal canonical β-binding motif, OB domain – predicted OB-fold domain. Below, three different fragments of α subunit used in the study are represented by thick lines with corresponding domains mapped onto them. (bottom) A multiple sequence alignment and a molecular model of the OB-fold domain encompassing residues 978-1078 of the α subunit. Aligned sequences include other bacterial homologs of E. coli pol III α subunit (DPO3A_ECOLI) and structural templates used in modeling. Conserved residues shared by more than half of aligned sequences are highlighted in blue (identical) and green (similar). Secondary structure corresponding to the OB model is displayed above the alignment. ECOLI, Escherichia coli; HAEIN, Haemophilus influenzae; NEIMA, Neisseria meningitides; MYCTU, Mycobacterium tuberculosis; DEIRA, Deinococcus radiodurans; AQUAE, Aquifex aeolicus; STAAS, Staphylococcus aureus; BACHD, Bacillus halodurans.

Since the use of multiple templates may also help to improve the accuracy of the model, three structural templates belonging to different OB-fold families were used:
anticodon binding domain of *E. coli* Aspartyl-tRNA synthetase (PDB code: 1C0A)\(^{185}\), the RecG “wedge” domain from *Thermotoga maritima* (1GM5)\(^{186}\) terminal domain of human Replication Protein A 70 KDa subunit (1FGU)\(^{187}\). More recently, as the X-ray structure of Pol III α subunit from *Thermus aquaticus* became available (2HPI)\(^{169}\) its OB-fold domain was added as the evolutionarily closest structural template. In addition to corroborating the accuracy of the conserved β-barrel core of the model, it enabled us to better represent the conformation of the flanking regions. Side chains in the final 3D model were positioned with SCWRL\(^{188}\).

A favorable Prosa2003 energy Z-score (-8.0) suggests that the model (Figure 5.1) is an accurate representation of the *E. coli* Pol III α subunit OB-fold domain structure. For comparison, three out of four OB-fold domains used as structural templates display worse Z-scores (1C0A: -8.2; 1GM5: -7.7; 1FGU: -7.5 and 2HPI: -7.3). Note that the poor Z-score for the OB-fold domain of *T. aquaticus* Pol III α subunit (2HPI) is due to a large number of missing residues in loops.

### 5.3 Pol III α Subunit Binds to Force-melted DNA

A sequence of extension and relaxation curves for DNA in the presence of 100 nM full-length Pol III α subunit is shown in Figure 5.2.a Upon the addition of 100 nM α in a 100 mM Na\(^+\) solution, the extension curve is largely unaffected. The slight deviations at low stretching forces are due to protein-protein and protein-DNA aggregation. The protein concentration of 100 nM was chosen to minimize this aggregation. After force-induced melting, the melted DNA does not appear to re-anneal on the ~5 minute time scale of relaxation. Subsequent extension data shows that some re-annealing has occurred while
the DNA was relaxed, though a significant fraction has remained single stranded. After a few cycles, the extension and relaxation data fully overlap, suggesting that the protein binds over a timescale of ~30 minutes.

To determine the fraction of DNA bound to protein, (x), a series of constant extension experiments was performed. Protein was introduced into solution, the translation stage held at a fixed position for some time (30 minutes, unless noted), and then relaxed. For relatively low fixed extensions, (< ½ of the length of the overstretching plateau), a

**Figure 5.2.** DNA stretching cycle in the presence of 100 nM full-length α protein. a) In this data series, the initial cycle (blue) shows a slight increase in the melting plateau, and reveals substantial hysteresis. Only a fraction of the melted DNA returns to the double helical form, as protein bound to single stranded DNA inhibits annealing over the timescales of these experiments. Subsequent cycles (from green to red) show that additional regions become stabilized by protein binding, until an apparent equilibrium length of ~½ of the double strand is melted. Strand separation occurs in the last cycle at ~140 pN. b) A sequence of constant extension experiments in the presence of 100 nM full-length protein. DNA is stretched to the extension marked by each arrow, held for 30 minutes, and then relaxed. Relaxation data is shown as open circles. As DNA is held to greater extensions, correspondingly larger fractions of the double helix are destabilized through protein binding to the single stranded regions, though some unbinding is evident upon relaxation. Solid lines show the best fits to this data, according to Equation 1, and yield destabilized fractions of 0.00 ± 0.01 (violet), 0.14 ± 0.02 (blue), 0.24 ± 0.02 (green), 0.40 ± 0.02 (yellow) and 0.42 ± 0.03 (red). Though the fixed extensions are progressively increased, the melted fraction saturates at ~½, as noted above.
fraction of the double helix is permanently melted, as shown in Figure 5.2.b. This fraction corresponds closely to the length that is melted by force. The value of the fraction \( x \) may be found by fitting the observed contour length \( b \) to:

\[
b(x) = b_{ds} (1-x) + b_{ss} x,
\]

(5.1)
a linear combination of the contour length of double stranded \( b_{ds} \) and single stranded \( b_{ss} \) DNA. Results for these fits are shown in Figure 5.2.b, where the fits are confined to data below 30 pN, to minimize unbinding effects described below. Only a maximum of \(~\frac{1}{2}\) of the overstretching plateau disappears, even when the extension is held near the end of the plateau. Thus, significant melting of dsDNA must be necessary to observe regions of stable ssDNA. For longer extensions, unbinding is also noted, but is only complete upon full DNA relaxation \( F \leq 1 \text{ pN} \) for at least 20 minutes (data not shown).

Proteins that actively bind to ssDNA generally lower the observed melting force as the protein helps to stabilize ssDNA relative to dsDNA. The fact that full length \( \alpha \) does not lower the melting force, but inhibits re-annealing up to the extension length of the stage indicates that this protein passively binds to regions that have been melted by force, and that it is unable to stabilize DNA that is melted during short-lived thermal fluctuations. Initially, partial binding inhibits re-annealing of unbound regions, and after \(~30 \text{ minutes}\) binding to the ssDNA appears complete. As these melted regions appear relatively stable, dissociation from ssDNA must be relatively slow (of the order of an hour). Furthermore, the fact that only one half of the melting transition is affected indicates that the protein cannot saturate the ssDNA lattice.
5.4 Pol III $\alpha$ Subunit Stabilizes dsDNA

Although the slight increase in the DNA melting force observed in high salt suggests dsDNA stabilization, this effect is more obvious in low salt. Figure 5.3 shows dsDNA stretched in a solution of 100 mM Na$^+$ and 10 mM Na$^+$, respectively. In the absence of protein, the expected destabilization of the dsDNA with decreasing salt concentration is observed$^{50}$. To quantify this effect, we measure the melting force at the approximate midpoint of the transition (averaging the observed force over the range 0.44 – 0.46 nm per base pair). The melting force is observed to decrease from 62.6 ± 1.5 pN in 100 mM Na$^+$ to 52 ± 3 pN in 10 mM Na$^+$, comparing well with previously measured values (the higher uncertainty in the force is typical for low salt measurements)$^{50}$. Upon the addition of full-length $\alpha$ protein,
the melting force increases to $66 \pm 2$ pN. Thus, full-length protein stabilizes dsDNA, but this effect is seen only in low salt concentrations. Higher protein concentrations do not appear to significantly further stabilize the DNA.

5.5 N-Terminal Construct α1-917 Binds to dsDNA

In order to isolate the dsDNA and ssDNA binding sites, we obtained fragments of the α protein and performed stretching experiments in the presence of these protein fragments. The N-terminal construct, residues 1-917, showed no evidence of binding to ssDNA (Figure 5.4). In standard buffer conditions (100 mM Na\textsuperscript{+}), there is little evidence of DNA binding. Data collected in lower salt (10 mM Na\textsuperscript{+}), reveals dsDNA stabilization, though

\begin{figure}[h]
\begin{center}
\includegraphics[width=\textwidth]{figure5.png}
\end{center}
\caption{DNA binding of N-terminal construct (α1-917). To provide a reference transition force, DNA is first extended/relaxed in 100 mM [Na\textsuperscript{+}] (black), melting at 62.6 ± 1.5 pN. a) No change is evident within experimental error for 100 nM protein in 100 mM [Na\textsuperscript{+}] or at 500 nM protein (data not shown). b) When the buffer is exchanged for 10 mM [Na\textsuperscript{+}], DNA melts at a lower force of 57 ± 2 pN and relaxes into the helical form with substantial hysteresis (purple). As protein is added to the lower salt solution, DNA stabilization is observed at higher protein concentrations (the melting force is restored to 66 ± 2 pN), no binding to the single strand is observed within the uncertainty of the experiment, though some hysteresis remains (blue, green).}
\end{figure}

5.5 N-Terminal Construct α1-917 Binds to dsDNA

In order to isolate the dsDNA and ssDNA binding sites, we obtained fragments of the α protein and performed stretching experiments in the presence of these protein fragments. The N-terminal construct, residues 1-917, showed no evidence of binding to ssDNA (Figure 5.4). In standard buffer conditions (100 mM Na\textsuperscript{+}), there is little evidence of DNA binding. Data collected in lower salt (10 mM Na\textsuperscript{+}), reveals dsDNA stabilization, though
higher concentrations of protein are required relative to the full-length protein (Figure 4b). Although the dsDNA binding activity of the α1-917 fragment is weaker than that of the full-length α protein, this fragment clearly contains a dsDNA binding site. It is likely that the tandem HhH motifs are responsible for dsDNA binding. Moreover, our observations allow us to rule out ssDNA binding by the HhH motifs or other regions of the N-terminus of α.

5.6 C-Terminal Constructs α 917-1160 and α 978-1160 Bind to ssDNA

C-terminal protein fragments consisted of the residues 917-1160 and 978-1160, which includes the putative OB-fold region (residues 978-1078). Force-extension experiments with these proteins indicate that there is little stabilization of dsDNA, even in low salt.

### Figure 5.5: DNA binding of C-terminal α constructs. A series of extension and relaxation cycles (in the order: blue, green, red) show significant binding to the single strand, similar to that of full-length protein. Upon relaxation in the presence of protein, a significant fraction remains single stranded as bound protein inhibits formation of the double helix. Slight dsDNA stabilization is evident as well. Both a) a longer construct (α917-1160) and b) a shorter construct (α978-1160) show definite binding to the single strand. The rough appearance of the extension data is evidence of aggregation and of protein unbinding from the single strand.
However, binding to ssDNA is evident, as shown for the fragment composed of the residues 978-1160 in Figure 5.4. Strong aggregation effects also appear in the extension curves, though this may also be due to a binding mode in which ssDNA is wrapped around the protein upon binding. As the protein binds into the ‘bubbles’ of melted DNA, the single strand may be wrapped around the protein. The wrapped ssDNA is shorter than naked ssDNA or dsDNA and this may cause an increase in the measured tension upon subsequent extension curves. If these proteins (or groups of proteins) are unbound, then the DNA will show a sudden decrease in force upon extension, as shown in the jagged force-extension data in Figure 5.5. As the DNA is relaxed, the protein binds again, to be dissociated upon a subsequent stretching cycle. Though this binding mode may be unique to the isolated C-terminal constructs of the protein, it is clear that a ssDNA binding site is contained within this fragment. Our modeling, in agreement with other observations\textsuperscript{168}, suggests that the OB-fold of α is most similar to a subclass of OB-folds that includes archaeal/eukaryotic ssDNA binding proteins and anticodon binding domains rather than to eubacterial SSBs. Notably, the OB-fold domain of α has fewer aromatic residues lining the putative ssDNA binding path than either the tRNA anticodon-binding domain of tRNA synthetases or ssDNA-binding OB-fold domains. This suggests relatively weaker binding of ssDNA by the OB-fold domain of α, which is in agreement with our observations.

5.7 Conclusions

By using single molecule force spectroscopy, we have demonstrated that the full-length Pol III α subunit binds to both dsDNA and ssDNA. Based on modeling and structural\textsuperscript{169}
studies that located an OB-fold within the C-terminus, we isolated the putative dsDNA and ssDNA binding sites by constructing fragments that contain only the dsDNA binding site (α1-917) or only the ssDNA binding site (α917-1160 and α978-1160). While the existence of a dsDNA binding site on the polymerase is not unexpected, the role of the ssDNA binding site is unknown, but may be required for processes involving ssDNA during replication, such as binding template DNA or ssDNA during proofreading. In previous studies, a chimeric RB69 DNA polymerase fused to SSB exhibited higher processivity and a better ability to polymerize through pause sites in the template that are attributed to secondary structure compared to the wild-type DNA polymerase, without altering the fidelity of the polymerase. The location and orientation of the OB-fold domain in the *T. aquaticus* Pol III structure suggest that the OB-fold may be able to undergo a domain rearrangement in order to bind downstream template DNA. Moreover, it has been shown that in the case of *E. coli* Klenow fragment, four base pairs are unpaired in order to allow the terminal nucleotide of the primer to enter the exonuclease proofreading site. The results from previous studies leave open the possibility that the OB-fold domain of α may be involved in binding ssDNA generated during proofreading.

The characteristics of ssDNA binding by both the full-length protein and C-terminal binding fragment differ substantially from that of classical SSB proteins such as *E. coli* SSB, T4 gene 32 protein (gp32), and T7 gene 2.5 (gp2.5). In particular, all three of these SSB proteins are capable of actively destabilizing dsDNA. For example, both T4 gp32 and T7 gp2.5 lower the DNA melting force in real time as the DNA is being stretched, such that the observed melting transition force is lowered in the presence of the
protein\textsuperscript{60,62}. This indicates that these SSB proteins are capable of binding to and stabilizing ssDNA that is created by transient thermal fluctuations that induce short-lived bubbles of melted DNA. In contrast, ssDNA binding by the $\alpha$ subunit occurs only after ssDNA has been fully melted by force. This unusual behavior may be functionally important, as ssDNA binding will likely only occur after other replication processes create ssDNA.
Chapter 6

Interaction of the Bacteriophage T4 Gene 59 Protein with DNA in the Presence of Gene 32 Protein

Single Molecule Stretching Measurements are done by Leila Shokri at Department of Physics, Northeastern University. Protein Purification is done by Amy M. Branagan at Departments of Biochemistry and Microbiology and Molecular Genetics, University of Vermont College of Medicine. This work is done in collaboration with Prof. Scott W. Morrical from the Departments of Biochemistry and Microbiology and Molecular Genetics, University of Vermont College of Medicine.

6.1 Background

DNA replication is a multistage reaction that requires the recruitment of enzymes and proteins that collectively form the replisome and vary in complexity depending on the organism. The complex replication machinery of T4 bacteriophage uses two modes of replication initiation. The first mode is an origin-dependant replication that occurs early in infection while the second mode is a recombination-dependant replication (RDR) and occurs at later times. These two modes are linked functionally and many of the phage replication proteins are responsible for both modes of DNA synthesis\textsuperscript{191,192}.

The replisome is composed of three units, the primosome (a complex of the helicase (gp41) bound to the primase (gp61) and the helicase loading protein (gp59)), the
holoenzyme (a DNA polymerase (gp43), a sliding clamp (gp45), and a clamp loader (gp44/62)), and the single-stranded binding protein (gp32). These phage-encoded proteins work together to form the minimal complex designed for the rapid production of multiple copies of the phage genome that permits efficient leading and lagging DNA synthesis and replication fork movement. The T4 DNA Polymerase (gp43) contains both 5'-3' polymerase and 3'-5' exonuclease activities and works together with single-stranded DNA binding protein (gp32) to catalyze DNA synthesis on both the leading and lagging strand of the replication fork. The processivity of the polymerase is greatly increased by the interaction of this enzyme with its accessory proteins (the processivity clamp (gp45) and the ATP-dependent clamp-loader (gp44/62)) along with the single-stranded DNA binding protein (gp32).

The ATP-dependent helicase (gp41) opens the template DNA ahead of the replication fork. The primase (gp61) synthesizes a short RNA primer in a template-directed polymerization that initiates the synthesis of Okazaki fragments by the polymerase (gp43) in lagging strand DNA synthesis. The fourth accessory protein (gp59) stimulates the activities of the replicative DNA helicase by promoting the assembly of gp41 onto single-stranded DNA that is covered with cooperatively bound gp32. Additionally, 5'-3' nuclease RNase H and ligase enzymes (found in nearly all organisms) are also required to remove primers and seal the nick between adjacent Okazaki fragments, respectively.

The second mode of replication leading to a rapid acceleration of DNA synthesis requires phage-encoded DNA recombination proteins, the strand exchange protein (UvsX), its accessory protein (UvsY), the putative exonuclease (gp46/47), and topoisomerase (gp39/52/60) in addition to the major origin-dependant replication proteins.
mentioned above\textsuperscript{191}. In this mode, origins of DNA replication in the infecting phage DNA initiate homologous recombination. The resulting two forks traverse the genome leaving each 3′ end of the lagging strand uncompleted. These single stranded ends can invade homologous sequences on the same or another DNA molecule, and, in the presence of phage replication machinery, can serve as a primer for leading-strand synthesis in the rightward direction. This process is self perpetuating when the fork reaches the genome end and these repeated cycles result in very long phage DNAs that are subsequently cleaved to form unit-length molecules\textsuperscript{197}.

Because the same simple replication machinery is responsible for both modes of phage DNA synthesis, the mechanism of loading the phage-encoded helices (gp41) must be versatile. The recruitment of gp41 to the lagging strand is limited by its low affinity for single-stranded DNA and by saturation of the lagging strand by gp32. Single-stranded DNA binding proteins compete with the helicase and inhibit its loading on single-stranded DNA binding sites due to their high affinity to ssDNA and their abundance. Therefore, an accessory protein is needed for efficient loading of helicase. Bacteriophage T4 gene 59 helicase loading protein is an important component of the coupled DNA replication and recombination systems. In the recombination-dependant pathway, lagging strand synthesis is obstructed in the absence of gp59 due to blocking the access of gp41 by the combination of phage-encoded gp32, UvsX, and UvsY\textsuperscript{197,198}.

The full length 34 kDa gp32 (301 amino acids) has been studied extensively\textsuperscript{124,199,200}. It binds cooperatively to ssDNA and destabilizes the secondary structure of the DNA\textsuperscript{201}. T4-encoded gp32 contains three domains: A domain, B domain, and core. The acidic, C-terminal A domain (residues 254-301) is the site of protein-protein interactions involved
in binding other T4 DNA replication and recombination proteins. The N-terminal B
domain (residues 1-21) is responsible for gp32-gp32 interactions resulting in cooperative
single-stranded DNA binding. The central core domain (residues 22-253) contains the
single-stranded DNA binding site consisting of a zinc-stabilized
oligonucleotide/oligosaccharide binding (OB) fold\textsuperscript{149,202,203}.

The basic 26 kDa monomeric gene 59 protein is a mediator protein that loads the
gp41 helicase onto gp32-saturated ssDNA molecules. It has a high affinity for gp41 and
gp32 as well as DNA and has been shown to bind forked DNA substrates, single-stranded
DNA, and duplex DNA. However, it has a higher affinity for forked DNA\textsuperscript{204}. It has been
shown that gp59 binds to single-stranded polynucleotides, including native ssDNA, with
a binding size of nine to ten nucleotide per monomer and exhibits at least two distinct
ssDNA-binding modes dependant on salt and/or lattice effects. These modes are
distinguished by their cooperativities of binding, a high-cooperativity mode at low salt,
and a low-cooperativity mode at higher salt\textsuperscript{205}. The crystal structure of gp59 has revealed
an $\alpha$-helical two-domain fold, a C- and an N-terminal domain. It has been proposed that
duplex DNA binds to the N-terminal domain, the lagging strand traverses a groove
between the two terminal domains, and the leading strand binds to the bottom surface of
the C-terminal domain\textsuperscript{206,207}. It also can form complexes with gp41 through the C-
terminus of gp41 and with gp32 through its acidic C-terminal A domain (residues 254-
301) and it has been suggested that it may exist as an oligomer in a complex with gp32
and gp41 (1:1 stoichiometry based on subunits)\textsuperscript{206,208}. It has also been demonstrated that
Cys-42 of gp59 interacts with Cys-166 of the core domain of gp32, lowering the affinity
of gp32 for single-stranded DNA, which may be a key factor in the removal of gp32 for helicase loading\textsuperscript{206,209}.

To further explore the mechanism of gp59 function, we conducted single-molecule force spectroscopy studies of gp59-DNA interactions, where we employed our well studied method of force-induced melting. We characterized the interactions of wild type protein with stretched DNA and then compared its behavior to that of its site-directed mutant (gp59\textsubscript{R12A}) with defects in dsDNA binding. From our stretching data in the absence of a fork, we found that gp59 binds more strongly to double-stranded DNA and we measured the equilibrium binding constant to dsDNA as a function of salt concentration. The present result appears to contrast with ensemble studies\textsuperscript{209} indicating that gp59 binds tightly and with high cooperativity to single-stranded DNA. We also examined the effects of gp59 on the DNA duplex stability in the presence and absence of gp32 and *I (gp32-A)\textsuperscript{61,82}, a gp32 truncated form in which the C-terminal A-domain has been removed by selective proteolysis. Our results demonstrate that gp59 is capable of strongly destabilizing both gp32-DNA and *I-DNA interactions. This result confirms a previous hypothesis that the local weakening of gp32-ssDNA binding is a requirement for helicase loading at that site\textsuperscript{81,210}.

6.2 T4 Gene 59 Protein Interactions with DNA

To determine the binding activity of gp59 and its site-directed mutant (gp59\textsubscript{R12A}), we used DNA stretching, where we measured the force-extension curve of double-stranded λ-DNA in the presence of gp59 and gp59\textsubscript{R12A} over a range of salt and protein concentration. Typical results of our measurements in 50 mM [Na\textsuperscript{+}] are shown Figure
6.1. The elevations in DNA force-induced melting in the presence of gp59 (200nM) and gp59R12A (200nM) indicate that both proteins bind preferentially to double-stranded DNA at low salt. However, gp59R12A shows reduced affinity for dsDNA compared to the wild type protein. Protein binding to dsDNA is expected to raise the DNA melting force because additional force is required to remove the protein before melting occurs\textsuperscript{63,64}, while the protein binding to ssDNA lowers the DNA melting force\textsuperscript{59-62}. The deviation at low stretching forces is due to protein-protein and protein-DNA aggregation.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure6_1.png}
\caption{a) Stretching-relaxation curves for \(\lambda\)-DNA in 10 mM Hepes, pH 7.5, 50 mM Na\textsuperscript{+} in the absence of protein (black) and in the presence of 200 nM gp59 (dark purple) and gp59R12A (olive). b) Enlargement of melting transition for 200 nM gp59 (dark purple), 100 nM gp59 (pink), and 25 nM gp59 (light purple). c) Enlargement of melting transition for 200 nM*I (olive), 100 nM*I (green), and 25 nM*I (light green).}
\end{figure}
In higher salt concentration (100mM [Na⁺]) the DNA melting force is increased slightly, about 1pN, in the presence of both proteins. As is shown in Figure 6.1.a, in the absence of protein and in the presence of gp59R12A, very little hysteresis is observed. However, the relaxation curve (solid line) does not match the stretching curve (dashed line) in the presence of the wild type protein. The observed hysteresis demonstrates that gp59 dissociates from dsDNA during DNA force-induced melting, further supporting the preferential binding of gp59 to dsDNA. The lack of hysteresis in the presence of gp59R12A can be relevant to the protein’s lower affinity for dsDNA. As illustrated in Figure 6.1.b and c, the increasing amounts of either protein results in an increase in the DNA melting force. Below, we use these data and employ the theory of force-induced melting of single DNA molecules to determine the protein-DNA association constants of both proteins.

In this model (see section 1.7) the free energy variations, \( \delta G \), due to changes in the force, \( \delta F \), are equivalent to variations due to changes in temperature, \( \delta T \),

\[
\delta G = \delta T \cdot \Delta S = \delta F \cdot \Delta x, \tag{6.1}
\]

where, \( \Delta S \) and \( \Delta x \) are the entropy and length changes per base upon DNA melting, respectively. The equilibrium melting force, \( F_m^0 \), changes in the presence of protein to \( F_m \), and this change is analogous to changes in the DNA thermal melting temperature due to protein binding. It has previously been shown\(^{111} \) that the equilibrium melting temperature, \( T_m^0 \), changes upon protein binding to a new observed melting temperature, \( T_m \), and it can be related to the protein-DNA binding constants to dsDNA and ssDNA as follows:
where, \( n_{ss}, n_{ds}, K_{ss}, \) and \( K_{ds} \) are the protein’s binding site size in nucleotides and binding constants to ssDNA and dsDNA, respectively. \( C \) is the bulk protein concentration. The approximation is valid for low protein concentrations and according to Eq. (6.1). The deviation in the DNA melting force will be:

\[
F_m \approx F_m^0 + \frac{k_B T C}{\Delta x} \left( \frac{K_{ds}}{n_{ds}} - \frac{2K_{ss}}{n_{ss}} \right).
\]

(6.3)

By fitting the melting force as a function of protein concentration to Eq. (6.3), we could find \( K_{ds} \) at different salt concentrations. In our fits we took into account the assumption that in this experiment gp59 binds more strongly to dsDNA, i.e. that \( K_{ds} >> K_{ss} \). Also, the hysteresis observed upon relaxation is quite small, allowing us to consider the measured melting force as an equilibrium quantity, which will leave us with some inaccuracy within the measurement error. Furthermore, we used previously measured parameter \( n_{ss} \) equal to 9205 and the salt-dependant value of DNA melting in the absence of the protein, \( F_m^0 \), from the work of Wenner et al. **50**. Our measured \( K_{ds} \) for gp59 and gp59\(_{R12A}\) are shown in Table 6.1.

<table>
<thead>
<tr>
<th>[Na(^+)] (M)</th>
<th>( K_{ds}^{gp59} \omega ) (M(^{-1}))</th>
<th>( K_{ds}^{gp59_{R12A}} \omega ) (M(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.001</td>
<td>-</td>
<td>((5.44 \pm 0.83) \times 10^8)</td>
</tr>
<tr>
<td>0.025</td>
<td>((1.28 \pm 0.54) \times 10^8)</td>
<td>((1.04 \pm 0.39) \times 10^8)</td>
</tr>
<tr>
<td>0.05</td>
<td>((3.83 \pm 0.71) \times 10^7)</td>
<td>((1.06 \pm 0.19) \times 10^7)</td>
</tr>
<tr>
<td>0.1</td>
<td>((2.40 \pm 0.15) \times 10^6)</td>
<td>-</td>
</tr>
</tbody>
</table>

**Table 6.1.** Equilibrium parameter of gp59 and gp59\(_{R12A}\) binding constant to dsDNA obtained from stretching studies using Eq.(6.3). Values are measured as mean ± standard error.
The increase in the DNA melting force observed in this study suggests that gp59 has a higher affinity to dsDNA than ssDNA, which is in contrast to the ensemble studies that show preferential binding to ssDNA\textsuperscript{205}. However, it was shown\textsuperscript{209} that gp59-ssDNA interactions can induce local changes in ssDNA structure. These changes put strain on interactions between neighboring gp32 monomers, and are responsible for the destabilization of gp32-ssDNA interactions. This situation resembles T4 UvsY recombination protein, which also destabilizes gp32-ssDNA in a manner independent of protein-protein interactions\textsuperscript{211} by wrapping of ssDNA around a UvsY hexamer. This suggests that the same wrapping mechanism occurs in gp59-ssDNA interactions. However, in our DNA stretching experiment, where wrapping of ssDNA is expected to be much weaker, gp59 prefers to bind more strongly to dsDNA, likely due to its inability to wrap stretched ssDNA at high forces.

6.3 Destabilization of gp32-ssDNA Interactions by gp59

Single-stranded binding proteins inhibit the annealing of ssDNA. This behavior can be revealed in our stretching experiment by a reduction in the DNA melting force and by the strong hysteresis upon DNA relaxation\textsuperscript{59-62}. Here we use DNA stretching to determine the effects of gp59 on the duplex DNA destabilization capabilities of gp32 and its proteolytic fragment *I (gp32-A). Figure 6.2 illustrates the typical results of our measured force-extension curve of double-stranded λ-DNA in 100 mM Na\textsuperscript{+}.

In the absence of gp59, gp32 reduces the melting force. However, in the presence of the gp32-gp59 complex, the non-equilibrium melting force increases and the hysteresis decreases, indicating a significant reduction in the helix-destabilization activities of gp32.
This is in agreement with a previous hypothesis in which the local weakening of gp32-ssDNA binding by gp59 is proposed to be essential for efficient helicase loading at that site. The measured gp59-ssDNA association constants are strongly salt dependent due to the multiple ionic contacts between protein and the phosphate residues. The observed decrease in the hysteresis is likely a result of these strong electrostatic gp59-DNA interactions. Therefore, the annealing capability of gp59 can be a key factor in the removal of gp32 from ssDNA.

Our observation that gp59 destabilizes *I-ssDNA complexes further supports a previous model in which the core domain of gp32 interacts with gp59 through Cys-166 of gp32 and Cys-42 of gp59. This indicates that the negative effect of gp59 on the helix-destabilization capabilities of gp32 can be due to a direct interaction between gp59 and the ssDNA-binding domain of gp32, which could destabilize the gp32-ssDNA interaction and can result in displacement of gp32 from the DNA upon helicase loading.

Figure 6.2. Stretching-relaxation curves for λ-DNA in 10 mM HEPES pH 7.5, 100 mM Na+ in the absence of protein (black) and in the presence of a) 350 nM gp32 (red), 350 nM gp32 and 10 nM gp59 (blue), b) 100 nM *I (red), 100 nM *I and 10 nM gp59 (blue).
The considerable hysteresis observed upon relaxation in the presence of gp32, *I (gp32-A), or the mixture of gp32 proteins and gp59 shows that the measured melting force is not an equilibrium quantity. Nevertheless, it is possible to determine the equilibrium melting force. To do this, the DNA-protein complex was stretched to the middle of the overstretching plateau. At this point, while the force is still non-equilibrium ($F'$), the DNA extension is held constant, and the measured force is allowed to approach equilibrium ($F_\text{m}$).

Typical measurements in 100mM [Na$^+$] are given in Figure 6.3. $F_\text{m}$, can then be used to determine the protein binding constant. We were able to measure the change in the equilibrium melting force as a function of protein concentration in 100mM [Na$^+$]. By analogy with thermal melting studies, the dependence of the equilibrium melting force ($F_\text{m}$) on protein concentration is given by:

**Figure 6.3.** Time dependence of the DNA melting force at constant position in the absence of protein (dark blue) and in the presence of a) 350 nM gp32 (black), 350 nM gp32 and 5 nM gp59 (red), 350 nM gp32 and 10 nM gp59 (blue), 350 nM gp32 and 20 nM gp59 (green), b) 100 nM *I (black), 100 nM *I and 10 nM gp59 (pink), 100 nM *I and 20 nM gp59 (light blue), 100 nM *I and 30 nM gp59 (light green).
\[ F_m = F_m^0 - \frac{2k_B T}{n_{ss} \Delta x} \ln (1 + K_{ss} C), \]  
\[ \text{(6.4)} \]

where, \( F_m^0 \) is the equilibrium melting force in the absence of protein, \( n_{ss} \) and \( K_{ss} \) are the protein’s binding site size in nucleotides and binding constants to ssDNA, respectively. \( C \) is the bulk protein concentration and \( \Delta x \) is the change in length between dsDNA and protein-bound ssDNA. Here, we have assumed that the binding to ssDNA is much greater than the binding to dsDNA.

The effects of gp32 and *I (gp32-A) on duplex DNA destabilization have been studied extensively and their equilibrium binding constants to ssDNA (\( K_{ss,\omega} \)) have been measured in different salt concentrations\(^{60,61}\). In this work we used the same method to study the helix-destabilization capabilities of gp32 and *I (gp32-A) in the presence of gp59. By measuring \( F_m \) for a range of protein concentrations in 100 mM \([\text{Na}^+]\) and fitting it to Eq. (6.4), we could obtain the binding constant to ssDNA, \( K_{ss,\omega} \). The obtained values of \( K_{ss,\omega} \) for gp32 and *I (gp32-A) decrease with increasing levels of gp59, showing that gp59 is able to co-occupy ssDNA in the presence of gp32 and its truncated mutant. The averaged

\[ \text{Figure 6.4.} \text{ Time dependence of the DNA melting force at constant position in the absence of protein (dark blue) and in the presence of a) 200 nM gp32 (black), 200 nM gp32 and 50 nM gp59 (red), 200 nM gp32 and 100 nM gp59 (green).} \]
value of $K_{\text{ss},\omega}$ for 350 nM gp32 and 100 nM *I (gp32-A) in the presence of gp59 are $(1.72 \pm 0.56) \times 10^7$ and $(3.25 \pm 0.9) \times 10^8$, respectively. We used the previously measured $\Delta \chi$ and salt-independent parameter $n_{\text{sr}}$ for either protein \textsuperscript{214,215}. This assumption does not affect our results, because the binding site size of gp32 and gp59 \textsuperscript{205} are similar. The negative effect of gp59 on the single-stranded DNA binding affinity of *I (gp32-A) shows that gp59-gp32 interactions do not occur only through the C-terminal sequence, and the ssDNA binding domain of gp32 (core domain) can be also responsible for this interaction. The negative effect of gp59 is the same for gp32 and *I (gp32-A), suggesting that the interaction between the core domain and gp59 is sufficient to destabilize gp32-ssDNA binding \textsuperscript{206,210}. Another interesting result is shown in Figure 6.4 where, in 100 mM $[\text{Na}^+]$ high amount of gp59 (50-100 nM) can compete with 200 nM gp32 in binding to ssDNA ($K_{\text{ss},\omega} = 1.13 \pm 0.3 \times 10^7$), resulting in a lower equilibrium melting force than just gp32 alone. This result demonstrates that high amount of gp59 can interact strongly with ssDNA in high salt when gp32 shows less affinity for ssDNA \textsuperscript{60,61}. This is despite the fact that gp59-ssDNA binding is weakened by stretching. However, DNA melting partially occurs from its free ends by peeling off of the strand that is not under tension. This means that gp59 can wrap single-stranded regions created at the ends of the DNA molecule. In 50 mM $[\text{Na}^+]$, 100 nM gp59 in the presence of 200 nM gp32 shows strong dsDNA binding resulting in an increase in DNA melting force. This can be due to the high affinity of gp59 to dsDNA in low salt ($10^7$-$10^8$ M\textsuperscript{-1}). In low salt, only very high amount of gp32 (about 800 nM) could destabilize gp59-dsDNA binding (data not shown). The value of $K_{\text{ss},\omega}$ in 100mM $[\text{Na}^+]$ obtained in this study is one order of the magnitude higher than the previously reported fluorescent study of gp59-poly(d\textepsilon A) binding at low salt \textsuperscript{205}. 


This can be due to the fact that gp59 exhibits higher affinities for ssDNA (random sequence) than for etheno-modified polynucleotides\textsuperscript{205}.

6.4 Conclusions

In this chapter, we showed that gp59 binds strongly to dsDNA in a highly salt dependant manner. The observed strong binding of gp59 to dsDNA rather than ssDNA is in contrast to previous ensemble studies\textsuperscript{204,205,216}. However, it has been shown\textsuperscript{209,217} that the cooperative binding of gp59 to ssDNA could introduce kinks or wraps in the ssDNA, similar to the binding of the UvsY hexamer to ssDNA\textsuperscript{211}. Because wrapping of ssDNA is expected to be much weaker in our stretching experiment, the fact that we do not see a strong interaction between gp59 and ssDNA provides further support for this model. gp59\textsubscript{R12A}, a site-directed mutant with defects in dsDNA binding, showed reduced affinity for dsDNA compared to the wild type protein. A crosslinking study\textsuperscript{206} has established that gp59 can interact with the core domain of gp32. This result is further buttressed by our results showing the negative effect of gp59 on *I (gp32-A)-ssDNA interaction. Also, because of the similar negative effects on the gp32-ssDNA and *I-ssDNA, we can suggest that the interaction between the core domain and gp59 is sufficient to destabilize gp32-ssDNA binding. The reduction of gp32-ssDNA affinity and/or cooperativity within a tripartite complex containing gp59 and gp32 simultaneously co-occupying ssDNA may facilitate the gp59-dependant helicase loading\textsuperscript{81,195,218}. Our results demonstrate that gp59 is capable of competing with gp32 for binding to ssDNA. We showed that gp59-ssDNA interactions are sufficient for destabilizing the gp32-ssDNA interactions and the annealing capability of gp59 can be a key factor in the removal of gp32 from ssDNA.
Chapter 7

Conclusions

Decades of research on the molecular machinery responsible for the coordination of events at the replication fork of *E. coli* and its bacteriophages T4 and T7 have been made considerable progress in understanding the elaborate process of genomic replication. However, many questions concerning the central roles of individual proteins in catalyzing the reactions at replication fork still remain to be uncovered. In this study, we performed single molecule force measurements using dual beam optical tweezers instrument to quantify the thermodynamics and kinetics of the protein-protein and protein-nucleic acid interactions that help us to better understand the function of these proteins in the cell. Also, we further investigated the nature of the overstretching transition in DNA stretching experiments to obviate the controversy in interpretation of this transition.

Our studies of the DNA overstretched transition in the presence of glyoxal brought new evidence in favor of the melting nature of the overstretching transition in DNA, and against a double-stranded form of overstretched DNA, i.e. S-DNA. We showed that a majority of the DNA base pairs elongated by force during the overstretching transition become permanently single-stranded when exposed to glyoxal for a sufficient time. Our result provides quantitative evidence that DNA overstretching is accompanied by disruption of the DNA helical structure, including loss of Watson-Crick hydrogen bonding. We showed that glyoxal fixation of force-melted DNA remains valid in the broad range of solution pH between 7.5 and 9.7, and has a better result in low salt than in
high salt, which is due to the partial re-annealing of the modified DNA strands in the latter case.

We measured the single DNA molecule stability in the presence of bacteriophages T7 gp2.5 and its deletion mutant lacking 26 C-terminal residues, gp2.5-Δ26C. Gene 2.5 protein is the single-stranded DNA binding protein of bacteriophages T7 and has essential roles in DNA replication and recombination in phage-infected cells. We have quantitatively determined $K_{ss}$, the association constant of gp2.5 and gp2.5-Δ26C to single-stranded DNA, over a range of salt concentration not available to bulk binding studies. From the rate-dependant kinetic melting of DNA, we measured binding site size to single- and double-stranded DNA, the weak but biologically significant double-stranded association constant ($K_{ds}$), and the rate of proteins finding the fluctuationally melted DNA binding sites ($k_a$). This nonequilibrium rate of cluster growth is analogous to the growth of clusters that occurs at a replication fork as the helicase unwinds the double helix. Our results show that dsDNA binding by gp2.5 is weak but important because it presumably facilitates rapid binding of gp2.5 to exposed ssDNA that is created at the T7 replication fork. The truncated protein which differs from the wild type only by the absence of the last 26 C-terminal residues exhibits more affinity to ssDNA and dsDNA than does the full length protein. To explain this difference we presented a model in which protein dimerization involving the C-terminus regulates DNA binding by gp2.5. Our model suggests that, although the dimerization interaction is primarily electrostatic, there is a weak non-electrostatic component, which likely represents interactions at the dimer interface. We also calculated the free energy of gp2.5 dimerization per dimer. Our results from the similar salt dependent ssDNA and dsDNA binding suggest that the same
dimer dissociation mechanism is required for binding to both dsDNA and ssDNA. The measurements of protein concentration dependence of $k_a$ showed that gp2.5 monomers, formed upon breaking a gp2.5 dimer, search dsDNA by 1D diffusion to find available binding sites on ssDNA.

More studies on the gp2.5 fragments and mutants help better understanding of the regulation of gp2.5 helix-destabilizing activities and allow characterizing the specific amino acid residues which have central role in gp2.5-DNA interaction. Also, more studies on the effect of gp2.5 on the DNA duplex stability in the presence of other replication proteins can shed light on the functional role of this protein as a key component in coordinating reactions at origins and replication forks.

The multisubunit enzyme DNA polymerase III holoenzyme (pol III) is responsible for duplicating the *Escherichia coli* chromosomal DNA. It is composed of three functionally distinct subunits: $\alpha$ (polymerase), $\varepsilon$ (proofreading), and $\theta$ (stability). Among these, the $\alpha$ subunit of pol III is the catalytic subunit responsible for the actual DNA polymerase reaction. We predicted that the C-terminal region of pol III (residues 979 to 1081) possesses substantial homology to proteins of the so-called OB-fold family. By purifying fragments of the protein that encompass the putative OB-fold domain, and then characterizing those fragments for their DNA binding activity using single molecule DNA stretching experiments, we tested the hypothesis that this region can bind single-stranded DNA (ssDNA). We showed that the characteristics of ssDNA binding by both the full length protein and the C-terminal fragment differ substantially from that of classical single-stranded binding proteins in which it binds only to ssDNA regions that have been fully melted by force. Also, our results imply that the $\alpha$ subunit has an affinity
for double-stranded DNA. Constructs of segments of the α subunit show that the N-terminal region is responsible for dsDNA stabilization, while the C-terminal region binds to melted DNA. These results suggest that the C-terminal domain may interact with ssDNA created during the DNA replication process.

There is one aromatic residue (Phe 1031), positioned precisely as in other ssDNA binding proteins (SSB, RPA), that should contribute to ssDNA binding. Future studies on mutant F1031A which decreases ssDNA binding can test this hypothesis. Also, further studies on other mutants (I1016F, R1025F, and R1025Y) which improve the ssDNA binding activities will allow characterizing the specific amino acids responsible for interactions with single-stranded DNA. Mutating I1016 should make it more similar to the OB-domain of tRNA-synthetases, and exonuclease VII, while R1025 mutant should be more similar to bacterial SSB proteins. Once the molecular details of the single-stranded DNA binding activity is better understood, studies can be extended to analyze the role of this OB-fold domain in DNA replication as well as its interaction(s) with other DNA binding proteins, including those involved in DNA replication and repair.

To further understand the coordination of events in the replication fork of bacteriophages T4, we further investigated the helix-destabilization in the presence of gp59, a helices loading protein, and gp59 R12A, a site-directed mutant with defects in dsDNA binding. We showed that both proteins bind strongly to dsDNA in a highly salt dependant manner. However, gp59 R12A shows reduced affinity for dsDNA compared to the wild type protein. From these results we obtained the apparent equilibrium binding constants of both proteins to double-stranded DNA over a range of salt concentration. Ensemble studies have shown that gp59 can interact with the core domain of gp32. We
investigated this behavior by studying the helix destabilization activity of the T4 single-stranded DNA binding protein (gp32) and its truncated mutant (*I) in the presence of gp59. Our measurements show that gp59 destabilizes the interactions between gp32 and single-stranded DNA. This effect is very strong despite the fact that gp59 binding to ssDNA is weakened by stretching. Also, because of the similar negative effects of gp59 on the gp32-ssDNA and *I-ssDNA, we suggested that the interaction between the core domain and gp59 is sufficient for destabilizing the gp32-ssDNA binding. The reduction of gp32-ssDNA affinity and/or cooperativity within a tripartite complex containing gp59 and gp32 simultaneously co-occupying ssDNA may facilitate the gp59-dependant helicase loading. Our results indicate that gp59 is capable of competing with gp32 for binding to ssDNA. We showed that gp59-ssDNA interactions are sufficient for destabilizing the gp32-ssDNA interactions and the annealing capability of gp59 can be a key factor in removal of gp32 from ssDNA. Also, we showed that in physiological salt and pH condition where gp32 shows less affinity of ssDNA, high amounts of gp59 can compete with gp32 in binding to ssDNA resulting in a lower equilibrium melting force than just gp32 alone.

Studies can be extended to other gp59 N-domain mutants that are found to be defective in duplex binding and result in a weaker helicase loading activity. Bacteriophage gp59 is an important component of the coupled DNA replication and recombination systems. In the recombination-dependant pathway, lagging strand synthesis is obstructed in the absence of gp59 due to blocking the access of gp41 by the combination of phage-encoded gp32, UvsX, and UvsY. Further studies on the effects of gp59 and its mutants on the DNA binding affinities of other recombination proteins like
UvsX, and UvsY will help deriving a quantitative model of protein-DNA interactions that occurs within the Helicase Loading Complex.
References


