Complex DNA Binding Kinetics of HIV-1 Nucleocapsid Proteins

by Jialin Li

B.S. in Physics, Nanjing University
M.S. in Physics, Northeastern University

A dissertation submitted to

the Faculty of
the College of Science of
Northeastern University
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

November 24, 2014

Dissertation directed by

Mark Williams
Professor of Physics
Abstract

The Gag polyprotein of human immunodeficiency virus type 1 (HIV-1) plays an important role in the retroviral replication and packaging. It is further processed into matrix (MA), capsid (CA) and nucleocapsid (NC). During the maturation process, Gag is cleaved into three domains, MA, CA and NC and NC is further cleaved from NCp15 to NCp9, and finally NCp7(1). HIV-1 NCp7 is a 55-amino acid polypeptide with two zinc fingers (ZF), each with the invariant CCHC motif and 15 basic residues. The NC protein is the primary region for the interaction of Gag with nucleic acids. NCp7 is a nucleic acid chaperone that can facilitate numerous nucleic acid rearrangements throughout the complex reverse transcription process. The matrix (MA) domain of Gag is also capable of nucleic acid binding. The chaperone activity of these NC proteins has not been fully understood. We use optical tweezers to stretch single λ-DNA molecules through the helix-to-coil transition in the presence of NCp15, NCp9 and NCp7 as a function of pulling rate and solution conditions. The results reveal multiple timescales in the kinetics of interaction among DNA and NC proteins. We also demonstrate that NCp15 exhibits significantly slower kinetics than NCp9 and NCp15, consistent with an important regulatory role for the p6 domain that is cleaved from NCp15 during viral maturation. Previous studies have shown that rapid kinetics between DNA and protein is a major component of protein’s chaperone. Here we show that there are additional, slower DNA binding modes that allow NC to remain electrostatically and nonspecifically associated with DNA. In addition, we show that the DNA binding kinetics are also regulated by the cleavage of the p6 domain from NCp15 as retroviral replication proceeds.
Acknowledgements

I would like to express my sincere gratitude to everyone who contributed to and supported the completion of my PhD study.

First, I would like to thank my advisor, Professor Mark Williams for his patient guidance during my whole doctoral study and research. His knowledge, kindness and support has inspired me a lot, especially when I was having a hard time. Second, I want to thank Dr. Micah McCauley. He made lots of lab work much easier and he is always willing to help me in everything in the lab, like laser alignments, DNA preparation, programming and so on. Third, I would like to thank Thaya who trained me to work in the lab. Fourth, I want to thank all the lab members to build such a friendly environment in the lab to study and work.

Moreover, I would like to thank all the faculty and staff at the department of physics for their assistance during my study and I also want to thank Professor Armen Stepanyants, Professor Nathan Israeloff and Professor Alain Karma for being my committee members.

Last, I want to express my sincere thanks to my family, especially my parents. Their encouragement and support smooth the way to get my PhD degree.
Table of Contents

Abstract ........................................................................................................................................ ii

Acknowledgements .................................................................................................................... iii

Table of Contents ....................................................................................................................... iv

List of Figures ............................................................................................................................ vi

List of Tables .............................................................................................................................. viii

Chapter 1 Introduction .................................................................................................................. 1
  1.1 DNA structure ....................................................................................................................... 1
  1.2 DNA transcription and reverse transcription ................................................................. 2
  1.3 DNA structure ....................................................................................................................... 3
  1.4 Force induced melting ......................................................................................................... 6
  1.5 Single molecule studies of Single-Stranded DNA binding (SSB) proteins, Actinomycin D (ActD), and High-Mobility Group B (HMGB) proteins ................................................................................................. 7

Chapter 2 Single molecule studies on the kinetics of HIV-1 NCp7 ........................................... 12
  2.1 HIV-1 retrovirus .................................................................................................................. 12
     2.1.1 HIV-1 retroviral life cycle and HIV-1 NC proteins ....................................................... 12
     2.1.2 NC chaperone activity and retroviral reverse transcription ........................................ 14
     2.1.3 HIV-1 NCp7 .................................................................................................................. 15
  2.2 Single molecule DNA force-extension curves in the presence of HIV-1 NCp7 ............... 16
  2.3 Quantifying the general characteristics of DNA stretching with HIV-1 NCp7 ............... 20
  2.4 Quantifying the interaction kinetics between HIV-1 NCp7 and DNA by varying the DNA pulling rate ................................................................................................................................................. 23
  2.5 Multiple stretching experiments on the same DNA molecule with NCp7 ....................... 26
  2.6 Discussion ............................................................................................................................ 31

Chapter 3 DNA binding properties and kinetics of wild type and mutant HIV-1 Gag ................ 34
3.1 Background on wild type and mutant HIV-1 Gag .........................................................34
3.2 Interactions of HIV-1 Gag wild type with DNA ..........................................................35
3.3 Interactions of HIV-1 Gag mutants with DNA ...............................................................37
  3.3.1 CANA .........................................................................................................................37
  3.3.2 HIV-1 DZF1 and DZF1+2 .........................................................................................38
3.4 Discussion ......................................................................................................................39

Chapter 4 Distinct nucleic acid interaction properties of HIV-1 nucleocapsid protein precursor NCp15 explain defects in viral infectivity .................................................................41
  4.1 Background on HIV-1 NCp7, NCp9 and NCp15 ............................................................41
  4.2 Quantifying the general characteristics of DNA stretching with HIV-1 NCp7, NCp9 and NCp15 ................................................................................................................41
  4.3 Quantifying the interaction kinetics between HIV-1 NC proteins with DNA by pulling rate measurements ..........................................................................................45
  4.4 Discussion ....................................................................................................................49

Chapter 5 DNA binding properties of HIV Viral protein R (Vpr) ........................................52
  5.1 Background on HIV-1 Vpr ..........................................................................................52
  5.2 Single molecule studies of Vpr-DNA interactions .......................................................53
    5.2.1 Formation of loops on DNA in the presence of Vpr ..............................................54
    5.2.2 DNA compaction with Vpr ...................................................................................55
  5.3 Discussion ....................................................................................................................56

Chapter 6 Conclusions ......................................................................................................58
References .........................................................................................................................60
List of Figures

Figure 1.1 DNA structure in 2D plane ........................................................................... 2
Figure 1.2 Diagram of optical tweezers and flow cell ................................................ 4
Figure 1.3 DNA force-extension curves ...................................................................... 5
Figure 1.4 DNA stretching curves of T4 gp32 ................................................................. 8
Figure 2.1 Structure of HIV-1 virion and Gag .............................................................. 13
Figure 2.2 Life cycle of retrovirus .............................................................................. 14
Figure 2.3 Sequence of HIV-1 NCp7 ......................................................................... 15
Figure 2.4 DNA stretching curves of 20 nM NCp7 ......................................................... 17
Figure 2.5 dsDNA and ssDNA stretching curves with NCp7 ........................................ 19
Figure 2.6 Model of hysteresis area ratio calculation .................................................. 21
Figure 2.7 NCp7 concentration dependence of hysteresis ........................................... 22
Figure 2.8 NCp7 concentration dependence of slope ................................................ 23
Figure 2.9 NCp7 stretching curves under different pulling rates .................................. 25
Figure 2.10 Non-equilibrium NCp7 bound dependent on stretching time .................. 26
Figure 2.11 NCp7 stretching curves after multiple stretching and rinsing ................. 28
Figure 2.12 Slope and hysteresis changes of NCp7 after multiple stretching and rinsing ...... 30
Figure 2.13 Slope changes of NCp7 after rinsing competitor DNA ................................. 30
Figure 2.14 NCp7 stretching curves after Mg buffer rinsing ........................................ 31
Figure 3.1 HIV-1 Gag wild type and mutants

Figure 3.2 HIV-1 Gag and NCp7 stretching curves

Figure 3.3 Comparison of the binding between NCp7 and Gag

Figure 3.4 CANC stretching curves

Figure 3.5 DZF1 and DZF1+2 stretching curves

Figure 4.1 NCp7, NCp9 and NCp15 stretching curves

Figure 4.2 Slope of NC components dependent on concentration

Figure 4.3 Hysteresis of NC components dependent on concentration

Figure 4.4 NC components stretching curves under different pulling rates

Figure 4.5 Non-equilibrium NC components bound dependent on stretching time

Figure 4.6 Model of NCp15 binding to DNA

Figure 5.1 Functions of HIV-1 Vpr during viral lifecycle

Figure 5.2 Vpr stretching curves

Figure 5.3 DNA loop formation and histogram of loop formation forces

Figure 5.4 DNA compaction with Vpr by constant force measurements
List of Tables

Table 1 Hysteresis of Gag wild type and mutants in different concentration.........................40

Table 2 Kinetic parameters of NC proteins bound to DNA.......................................................49
Chapter 1

Introduction

1.1 DNA structure

DNA (Deoxyribonucleic acid) is a molecule that encodes the genetic instructions used in the development and functioning of all known living organisms and many viruses. DNA molecules consist of two long nucleotides: each is composed of nucleobases, adenine (A), thymine (T), guanine (G) and cytosine (C), as well as a backbone made of alternating deoxyribose and phosphate groups shown in figure 1.1a. The carbon atoms on deoxyribose ring can be labeled from 1 to 5. In the 1950s, the double helix DNA structure was postulated based on crystallography data (2). In the structure, known as B form, it has a diameter of 2 nm, base pair separation of 0.34 nm, and repeats every 10 base pair. Two DNA strands are linked together by base pairing by its nucleobases, forming right-handed double helix.
Figure 1.1 DNA structure shown in 2D plane (3). Double stranded DNA linked by hydrogen bonds (dash lines). Carbon atoms in deoxyribose ring of the top two nucleotides are labeled from 1 to 5. Four ends of DNA strands are indicated with 3’ or 5’. Nucleobases are indicated by letter symbols.

The phosphate groups that form the DNA back bone are highly negatively charged, which makes the DNA strands highly negatively charged. Cations in solution screen the repulsion force between two DNA strands in the DNA. Therefore, the salt concentration affects the stability of the DNA duplex, with high salt concentrations favoring duplex formation.

1.2 DNA transcription and reverse transcription
Before DNA transcription starts, RNA polymerase binds to DNA and unwinds the DNA double helix to create a DNA bubble so RNA polymerase can access single-stranded DNA. During transcription, a DNA sequence is read by the RNA polymerase, which produces a complementary, anti-parallel RNA strand, where all thymine (T) nucleotides are replaced by RNA complement uracil (U). DNA transcription can happen in different position of one gene at the same time to produce many messenger RNA (mRNA), which will then be used to create proteins by translation.

A retrovirus stores its nucleic acid in the form of an RNA genome and then integrates its genetic information into the host cell by reverse transcription to create a DNA copy from its genomic RNA sequence. The host cell then transcribes and translates the viral genes along with its own genes, producing the proteins to assemble new copies of the virus. Details of the human immunodeficiency type 1 (HIV-1) retrovirus and the process of reverse transcription will be introduced in the next chapter.

1.3 Optical tweezers

In the 1970s Arthur Ashkin first found optical scattering and gradient forces on micron scaled particles and later, this technique was put into use to trap small particles, which is now commonly called optical tweezers (4-6). Optical trapping nowadays provides a method for studying molecular motors in the order of 1 to 100 piconewton (pN) in force resolution and nanometer (nm) in spatial resolution (7).

In our work, we use dual beam laser optical tweezers shown in Figure 1.2. The focal point of both beams can trap polystyrene bead and another bead can be trapped by the glass pipette attached to the flow cell. Then, we flow bacteriophage lambda DNA of 48500 base pairs, 16.5 micron in contour length. Biotins on both 3’ end of the DNA molecule can form strong bonds with both beads and the DNA is free to rotate known as torsionally relaxed.
Figure 1.2 (a) Schematic diagram of a dual beam optical tweezers instrument for stretching single DNA molecules. Two laser beams are overlapped to form an optical trap to trap one bead, and the other bead is attached to a glass micropipette. (b) Schematic diagram of the flow cell used in optical tweezers.

Figure 1.3 shows the typical stretching curves of DNA in 100 mM Na\textsuperscript{+}, 10 mM Hepes, pH=7.5 buffer. During the stretching process, the DNA is extended to its B-form contour length, which is about 0.34 nm/bp and the force is below 30 pN. When stretching further, the force increases dramatically. In this region, the double stranded DNA (dsDNA) can be fitted into Worm-Like Chain (WLC) model, which represents the elasticity of a rigid polymer, described by the equation:

$$
  b_{ds}(F) = b_{ds}^{\text{max}} \left[ 1 - \frac{1}{2} \left( \frac{k_B T}{F P_{ds}} \right)^{\frac{1}{2}} + \frac{F}{K_{ds}} \right] \quad \text{(eq 1.1)}
$$

Typical measured values in this model gives contour length per base pair $b_{ds}^{\text{max}}=0.34$nm, persistence length $P_{ds}=48$ nm, stretch modulus $K_{ds}=1200$ pN, though they are known dependent on salt concentration and pH (8-10). At a force of 60pN, a phase transition from dsDNA to an overstretched form of DNA occurs (11-13). In this region, we see a force plateau from 60 pN to 65 pN about 10 microns in extension length (14). After the transition, dsDNA is transformed into single stranded DNA (ssDNA) and the stretching curve goes up dramatically again.
Correspondingly, the ssDNA stretching curve can be fitted into Freely-Jointed Chain (FJC) model, described by the equation:

\[ b_{ss}(F) = b_{ss}^{\max} \left[ \coth \left( \frac{2FP_{ss}}{k_BT} \right) - \frac{k_BT}{2FP_{ds}} \right] \left[ 1 + \frac{F}{K_{ss}} \right] \] (eq 1.2)

This model gives \( b_{ss}^{\max} = 0.56 \text{nm} \), \( P_{ss} = 0.75 \text{ nm} \) and \( K_{ss} = 720 \text{ pN} \). When the transition is at equilibrium, the area between the stretching and release curve gives the DNA melting free energy \((8,13,15-18)\).
Figure 1.3 DNA force-extension curves: Solid circles represent the actual stretching and release curve of the DNA only. The purple curve represents the theoretical stretching curve of dsDNA, based on WLC (eq 1.1). The green curve represents the theoretical stretching curve of ssDNA, based on FJC (eq 1.2)(19).

Once we have proteins flow in solution, which bind to DNA, the stretching curve will change. This will allow us to study the thermodynamics and kinetics of those proteins how they interact with nucleic acids during DNA transcription and replication.

1.4 Force induced melting

Previous studies have suggested that, when DNA is stretched to high force, it will melt because the force increases the length of the DNA, converting dsDNA into ssDNA and disrupting both base pairing and base stacking(16,17,20,21). The major concern is that when the transition reaches to the contour length of ssDNA, the two strands do not come apart completely. Another model to explain that is the B-form dsDNA is converted into a new form of dsDNA called “S-DNA”, where the strands unwind and base-pairing remains but not stacking(22).

While the structure of DNA is still unclear in the overstretching transition, some experiments use glyoxal (23,24), intercalating dyes, and single stranded binding proteins (SSBs) (18,20,25-30) conclude that DNA overstretching is force-induced melting of dsDNA into ssDNA under most physiological conditions. Glyoxal (C2H2O2) is a small molecule that binds irreversibly to DNA with slow kinetics (23,24). In the experiment, the DNA molecule is held at fixed extensions for 30 min in the presence of glyoxal, which is the timescale required for DNA binding. When initially stretching DNA in the absence of glyoxal, it shows an expected contour length for dsDNA, 0.34nm/bp and the DNA release curve is almost identical to the stretching. After treatment with glyoxal, it exhibited a much longer contour length because the DNA molecule had become partially ssDNA and partially dsDNA. In addition, subsequent stretches also showed an additional increase in contour length. This evidence shows that DNA overstretching is force-induced melting of dsDNA into ssDNA under the solution conditions probed in these experiments (31).

Other single molecule studies have visualized the nature of the structure change with both optical tweezers and fluorescence imaging (20). A DNA molecule is stretched in the presence of YOYO,
a fluorescent dye which only intercalates into the paired bases of dsDNA. Therefore, the length of DNA to which YOYO binds equals to the fraction of dsDNA. Furthermore, the experiments on Human mitochondrial SSB which only binds and wraps relaxed ssDNA and the SSB Replication Protein A, which only binds ssDNA, also provide evidence to support the force-induced melting model. Although this study shows that DNA melted by peeling from the ends, it is likely that in the presence of binding ligands melting also occurs in the middle of the DNA molecule, as is the case for thermal melting (21,32,33).

The most recent study combining fluorescence microscopy and optical trapping has demonstrated that there is a balance between the three structures of overstretched DNA: peeled ssDNA, base pair breaking (yielding melting bubbles) and formation of S-DNA are governed by both DNA topology and local DNA stability (25). This comprehensive study shows at low ionic strength (50 mM NaCl), melting bubbles form when peeling is prohibited; at high ionic strength (150 mM NaCl + 20 mM MgCl2), melting bubble formation is disfavored and S-DNA is favored; and in topologically open DNA, peeling is suppressed with increasing ionic strength. An accompanying study in the same journal also showed melting bubbles appearing at low ionic strength (34).

1.5 Single molecule studies of Single-Stranded DNA binding (SSB) proteins, Actinomycin D (ActD), and High-Mobility Group B (HMGB) proteins

**SSB proteins**

In vivo, during the processes like DNA replication and transcription, the stable DNA helix duplex structure must be melted, leaving ssDNA exposed and vulnerable. SSB proteins can protect ssDNA from nucleases, chemical degradation, and transient secondary structure formation by binding to exposed ssDNA as it is formed (33). SSB proteins play important roles in the life cycles of viruses and cells. For example, bacteriophage T4 gp32 represents relative simple model systems that has been fully studied by single molecule method. It is known that T4 gp32 has some NA chaperone properties, by which it can optimize the transcription and reverse
transcription (18,26,35). The major reason that T4 gp32 is able to facilitate transcription is by preventing the formation of template ssDNA and RNA secondary structures, allowing for faster polymerization of the complementary strand by the polymerase (36). HIV-1 NCp7 also facilitates polymerase activity, both by destabilizing secondary structures and through NA aggregation (37).

Figure 1.4 shows the force extension curves of DNA in the presence of gp32 and its CTD truncation mutant gp32 *I at a pulling rate of 100 nm/s (18,38). The DNA stretching curves with gp32 show very little effect as if the protein does not have enough time to bind. Also, the release curves indicate the DNA remains single stranded and bound by the protein. When the stretch and release cycle is performed at a slower rate, the stretching and release curve will become closer, and eventually converge to reach equilibrium. However, the on and off rate for gp32 is so slow that the equilibrium curve is unobtainable in our experiments.

Figure 1.4 T4 gp32 stretching curves: DNA stretching (solid) and release (dash) curves in the absence of protein (black) and in the presence of gp32 (red) and its mutant gp32 *I (blue)(18).
To quantify the kinetics of interaction between gp32 and DNA, we fix the DNA length near the melting transition midpoint and watch relaxation of the force over time (27,28,35,38). This procedure allows us to obtain the equilibrium melting force and the protein’s on and off rates. The slow off time for gp32 dissociation from ssDNA, on the order of several minutes, is due to cooperative binding, leading to its preferential dissociation only from the end of protein bound clusters. Furthermore, measuring the dependence of the melting force on pulling rate allows us to estimate the protein’s ssDNA on rate under the conditions when all of the available ssDNA is already bound by polymerized gp32, while new ssDNA sites for gp32 appear upon stretching the DNA. The measured on rate dependence on protein concentration suggests that gp32 finds its new ssDNA site by sliding along the dsDNA. Therefore we can determine the gp32 equilibrium binding constant to dsDNA. Finally, all of these properties of gp32 are studied as a function of monovalent salt concentration for wt gp32 and the CTD truncation mutant *I. The results enable us to build a model of the electrostatic regulation of gp32-DNA binding by this protein’s anionic CTD. The model is consistent with the result of all other studies of this protein, but also provides extensive new information (39).

**Small molecule binding to DNA**

ActD is a small molecule with strong antibiotic and anticancer activity (40,41). ActD can intercalate between dsDNA base pairs, bind to ssDNA and can even “hemi-intercalate” between the bases of single DNA strand (42-44). However, it was not previously clear what properties of ActD lead to its anti-cancer activity. Is its binding to ssDNA important or is it its ability to intercalate that gives it anti-cancer activity? To resolve this controversy, the Williams lab developed a method to quantify ActD’s equilibrium and kinetics as a function of stretching force. They measures the on rates while flowing ActD in and measure the off rates while washing ActD away. While holding the DNA at a constant force to measure the change in DNA extension, fits of the time-dependent extension yield the dissociation constant $K_d$. Extrapolation of the force dependence of ActD binding also yield the equilibrium dissociation constant in the absence of force (45).
We show that binding from melted DNA is rapid. However, the on rate obtained with 50 nM ActD in the absence of force is very slow (~$10^{-5}$ s$^{-1}$). The destabilization of dsDNA by force exponentially facilitates the extremely low ActD-dsDNA on and off rates. To summarize, while ActD is an intercalator, its likely binding mode is approached from a state of two parallel strands. These results support a model in which the biologically active mode of ActD binding is to pre-melted dsDNA, as found in transcription bubbles (46), thereby inhibiting RNA synthesis. The ActD study provides a new method for using force-dependent measurements to determine DNA structural dynamics upon protein binding. These measurements can be extrapolated to zero force to determine the binding parameters in the absence of force.

**HMGB eukaryotic nuclear DNA binding proteins**

High-mobility group (HMG) proteins are abundant nuclear proteins that support a range of cellular functions, such as DNA repair and cellular signaling (47-49). The HMGB subgroup is characterized by non-sequence-specific DNA binding. HMGB proteins bind to the DNA minor groove strongly, bending the DNA and enhancing the flexibility of the DNA double helix (50,51). In the single molecule stretching study of HMGB protein Nhp6A, it is found HMGB protein binds to B-form DNA about 20 times stronger than to overstretched DNA (14). The highly reproducible DNA stretching curves with saturating protein concentration suggest that HMGB proteins remain bound through the stretch and release cycles, while the timescale for the transition from B-DNA to overstretched DNA is longer than 100 s. Also, in the presence of HMGB proteins, when increasing the DNA pulling rate, the overstretching forces increases while this is not the case with DNA only. Therefore, HMGB proteins inhibit DNA unwinding and rewinding at high pulling rates (14).

The major result of the study of HMGB proteins is that the measurement of rapid DNA dissociation rates agree with recent bulk measurements (52) but contrast with the slow dissociation rates observed upon removal of other proteins from the solution surrounding the single DNA molecule (53). This contradiction can be resolved by recognizing that there are two distinct protein-DNA dissociation rates: macroscopic and microscopic (14). The slow macroscopic dissociation will allow proteins to escape into the solution. The fast microscopic
dissociation will break those short-range protein-DNA contacts, but the protein remains bound to DNA through long-range electrostatic interactions. Such slow macroscopic dissociation can be accelerated by high salt (54). Furthermore, our study shows short-range HMGB-DNA contacts play the major role in HMGB induced DNA bending and HMGB proteins bind two DNA strands of any secondary structure and winding with fast microscopic kinetics (14).

The studies discussed above illustrate the utility of specific DNA stretching methods in quantifying DNA binding interactions. Studies on SSB proteins demonstrate our ability to quantify the thermodynamics and kinetics of protein binding to ssDNA, and even yield the dynamics of protein sliding on dsDNA to find ssDNA binding sites. In contrast, studies on ActD probe binding to both dsDNA and ssDNA, but we were able to determine that the ligand prefers destabilized dsDNA for binding. After being trapped in a distorted DNA helix, the ligand dissociates very slowly, explaining its potent anti-cancer activity (45). Finally, HMGB proteins bind preferentially to dsDNA, altering both its elasticity and stability, allowing us to quantify its dsDNA binding affinity. In addition, we showed that HMGB proteins require collisions with other proteins to rapidly dissociate from dsDNA, showing that protein binding involves separate microscopic and macroscopic dissociation rates. Here we will probe the thermodynamics and kinetics of HIV-1 NC, which exhibits a combination of the characteristics described above. NC binds strongly to both dsDNA and ssDNA, making characterization of this binding more difficult. We will also show that NC also exhibits microscopic and macroscopic dissociation rates, similar to those observed for HMGB proteins.
Chapter 2

Single molecule studies on the kinetics of HIV-1 NCp7

2.1 HIV-1 retrovirus

2.1.1 HIV-1 retroviral life cycle and HIV-1 NC proteins

HIV-1 is a typical retrovirus that has been well studied. It composed of two identical copies of RNA molecules 7-10 kilo-bases in length in the virion. The structure of HIV-1 virion is shown in figure 2.1 and there is a maturation process before the virion infects the host cell (55). The main components of virions are envelope, RNA and Proteins. Envelop is composed of lipids obtained from the host plasma membrane during budding process as well as glycoprotein encoded by the env gene. The proteins are primarily Gag proteins, protease, pol proteins and env proteins. During the maturation process, Gag is cleaved into three domains, MA, CA and NC and NC is further cleaved from NCp15 to NCp9, and finally NCp7(1).
HIV-1 enters the host cell by fusion. After that, the virus uses its own reverse transcriptase to produce DNA from its RNA genome. The new DNA is incorporated into the host cell genome by integrase, the form of which is referred as a provirus. The host cell treats the viral DNA as part of its own genome, translating and transcribing together to produce the protein required to assemble new virus. The whole retroviral lifecycle is shown in Figure 2.2 (57).
2.1.2 NC chaperone activity and retroviral reverse transcription

Reverse transcription is the process that converts ssRNA into dsDNA. It is known there are at least 3 important steps that require nucleic acid chaperone activity: tRNA primer annealing (58), minus-strand transfer, and plus-strand transfer (59-63). Previous studies have shown that the strand transfer steps require rearrangement of nucleic acid secondary structure to a lower free energy, which is very slow in the absence of protein (64-69). NC can accelerate the interactions in the strand transfer process. The capability to facilitate the nucleic acid (NA) rearrangement is known as chaperone activity (70). The chaperone activity consists of three components: NA aggregation, NA destabilization and rapid kinetics. NA aggregation is usually characterized by the increasing force at the start of the force-extension curve and it facilitates the NA annealing (71-74). NA destabilization is characterized by the increasing transition slope and overall lower transition force in the force-extension curves. The capability of NA destabilization can overcome the energy barrier to melt the TAR RNA and TAR DNA complex to allow NA to reanneal. Furthermore, rapid kinetics of NA dissociation is also important for allowing the proteins quickly
adjust to single and double strands binding states (67), which is characterized by the hysteresis, the area between DNA stretch and release curve (18,26,75).

2.1.3 HIV-1 NCp7

The HIV-1 NCp7 protein is the major nucleic acid (NA) binding domain of the Gag polyprotein. It is essential in the virus life cycle for viral genome selection and packaging, viral assembly and reverse transcription (76-89). HIV-1 NCp7 is a 55-amino acid polypeptide with two zinc fingers (ZF) and 15 basic residues, with net charge of 9 shown in figure 2.3. Each of the zinc fingers contains one aromatic residue interacting with nucleic bases, which could account for destabilizing duplex. The basic residues are likely primarily responsible for NA aggregation (70,90).

Figure 2.3 Sequence and secondary structure of HIV-1 NCp7(69)

HIV-1 NCp7 is a nucleic acid chaperone protein that facilitates the rearrangement of NAs into their lowest energy configuration. NCp7 is known to have all the three components of chaperone activity and studies have shown that all the three steps in the reverse transcription need the participation of NCp7 proteins (37,65). Besides the high efficiency of full-length cDNA production, the RNA aggregation activity of NCp7 is also needed for this activity. All of these
features indicate a significant role for the chaperone activity of HIV-1 NCp7 in reverse transcription. In contrast, NC proteins from other retroviruses do not always exhibit all of the components of an ideal chaperone. For example, HTLV NC is a poor aggregator of NA (91,92), while FIV NC has excellent aggregation but weak NA destabilization (93). Similarly, the NC protein from the LTR retrotransposon Ty3 also has weaker NA destabilization. In this work, we will use single molecule DNA stretching methods to characterize the kinetics of NCp7-DNA interaction, which are critical for HIV-1 NC’s nucleic acid chaperone activity. These results will help us to further understand how the HIV-1 NC protein shapes processes involved in HIV-1 replication.

2.2 Single molecule DNA force-extension curves in the presence of HIV-1 NCp7

We use single molecule methods with dual beam optical tweezers for all the experiments. The λ DNA was labeled with biotin on both 5’ ends. Because of the strong non-covalent bonds between biotin and streptavidin coated on polystyrene beads, we can stretch the DNA molecule caught between two beads up to 300 pN. The stretching experiments here from now on are all performed in 10 mM Hepes, 50 mM Na+, pH 7.5 buffer. After attachment of one single DNA molecule, other DNA molecules were rinsed by buffer flow. The solution with specific protein concentrations were then flowed around the DNA molecule to investigate the DNA-protein interactions on DNA stretching curves. HIV-1 NCp7 proteins were stored at −80 °C in 20 mM HEPES, pH=7.5, 5 mM BME, 0.1 mM tris (2-carboxyethyl) phosphine (TCEP) and 2 mol of zinc per mole of NC. The concentration of the protein was determined using an extinction coefficient of 6060 M−1 cm−1 at 280 nm.

Figure 2.4 shows the stretching and release cycle for bacteriophage λ DNA in the absence of the protein and in the presence of 20 nM HIV-1 NCp7. In the absence of protein, at extensions much less than the B-form contour length of double-stranded DNA (dsDNA), very little force is required to stretch the DNA. As the contour length is approached, the force increases dramatically, reflecting the elasticity of the double helix. The plateau from 0.34 nm/bp to 0.6 nm/bp at a constant force of around 60 pN represents a transition from dsDNA to ssDNA, a force-induced melting transition (16,31,94). In the absence of protein, the transition is
cooperative, and as the DNA is relaxed back to lower extensions, the force-extension curve is almost completely reversible, showing little hysteresis. In the presence of wild type HIV-1 NC, the transition is altered significantly. HIV-1 NCp7 shows some aggregation, destabilizing the DNA in melting transition, reduction in the ssDNA contour length and a significant slope. First, the work done by the force to melt dsDNA provides a way to estimate the average base pair stability in the presence of HIV-1 NCp7. While in the absence of HIV-1 NCp7, each base pair’s stability is about 1.5 kcal/mol and in the presence of HIV-1NC, it is reduced 30% to about 1 kcal/mol, therefore proving HIV-1 NC to be a moderate duplex destabilizer. Another important feature is the significant increase of the transition slope, reflecting the intercalation of HIV-1 NC and reduction of the transition cooperativity (66,95). Also, it shows a small amount of hysteresis which indicates that HIV-1 NC exhibits fast kinetics, which is an important component of chaperone activity that describes a protein’s ability to rapidly bind to and dissociate DNA, and is correlated with nucleic acid reannealing activity. The rapid protein-DNA interaction kinetics, marked by small hysteresis in DNA stretching and release experiments, appears to be the physical property of NC proteins that most strongly correlates with in vivo infectivity of the corresponding mutant viruses (67).
Figure 2.4 DNA stretching (solid) and release (dash) curves in absence (black) and in the presence of 20nM NCp7 at non-equilibrium. The green line indicates the transition slope and the blue area indicates the amount of hysteresis.

Figure 2.5(B) shows the DNA stretching curves (black) with 10 nM NCp7 after multiple subsequent stretching cycles, and the small hysteresis vanishes indicating saturated HIV-1 NCp7 binding to both strands, which we refer to a HIV-1 NC/2x ssDNA complex. This curve differs significantly from the HIV-1 NC/1x ssDNA (red) while the latter is obtained by dissociating one of the dsDNA strands under high tension in high pH at first. This difference suggests the saturating HIV-1 NCp7 is capable of keeping the both DNA strands together at force up to ~80 pN, which in the absence of protein would induce strand separation. At the same time, HIV-1 NCp7 also strongly elongates and deforms the DNA duplex, indicated by the increase in extension beyond the B-form contour length. The combined capabilities of HIV-1 NCp7 to destabilize the B-form DNA duplex while also keeping the two strand from separation together make it an optimum NA chaperone by allowing it to decrease NA re-folding transition barriers, while also facilitating annealing of the new structure (75).
In this chapter, we will use several variations of single molecule DNA stretching to probe NC’s DNA interaction kinetics in detail. By stretching and releasing DNA at different pulling rates, we probe timescales for NC’s association to and dissociation from DNA as it is stretched. We find that, while most of the NC-DNA binding occurs on timescales much faster than probed by DNA...
stretching, a small component of bound NC dissociates only on timescales of hundreds of seconds. We also directly probe dissociation of NC from single DNA molecules in solution in the absence of competitor DNA or NC proteins in the solution surrounding the DNA. In this case we observe extremely slow NC dissociation, reflecting a primarily electrostatic and long-lived macroscopic DNA binding mode that it independent of the microscopic modes characterized in previous experiments, similar to that previously observed for several high mobility group B (HMGB) proteins as well as other non-sequence specific dsDNA binding proteins (14,96-99).

2.3 Quantifying the general characteristics of DNA stretching with HIV-1 NCp7

Here we measure the DNA stretching curves at different concentrations of HIV-1 NCp7. Two major features that characterize the DNA melting and annealing is the hysteresis and slope of the stretching curves. By quantifying the amount of hysteresis, the kinetics of interaction between DNA and NC could be characterized. Taking account of the difference in the extent of stretching observed in different DNA molecules, we calculate the relative hysteresis, which is the amount of the actual hysteresis relative to the amount of hysteresis possible in a given stretch and release cycle shown in Figure 2.6.

The observed DNA extension in experiments at a given force can be expressed as a linear combination of the extension of dsDNA given by WLC model and ssDNA given by FJC model(33,100). This combination curve is used to calculate the maximum hysteresis for a given DNA elongation cycle, which is the area between the hybrid WLC and FJC curve and the DNA extension curve, A2. The actual hysteresis is the area between DNA extension and release curve, A1. Thereby the hysteresis area ratio is A1/A2 and the results are presented below. In addition, this ratio has been corrected due to the very small amount of hysteresis existing in the DNA only stretch and release curve (75).
Figure 2.6 Model of Hysteresis Area Ratio Calculation: The red solid line and the red dotted line are the actual force extension and release curves of DNA. The blue line is the standard WLC model for dsDNA elasticity, and the green line is the standard FJC model for ssDNA elasticity, at a normal contour length 0.55nm/bp. The purple line is a linear combination of WLC and FJC models representing a fraction of DNA melted in the initial stretch. The hysteresis ratio is determined by calculating the ratio of the area between the extension and release curve (red solid and dotted lines, respectively) and the area between the extension curve and the linear combination of FJC and WLC models (solid red line and solid purple line).

When the NCp7-DNA complex is stretched, the DNA strands become intercalated and never separate, leading to a small hysteresis (75). The hysteresis area ratio represents the non-equilibrium component of DNA-bound. Measuring the hysteresis area ratio allows us to estimate the overall protein-DNA interaction kinetics. The larger it is, the slower the interaction kinetics.
Figure 2.7 HIV-1 NCp7 concentration dependence of hysteresis area ratio

Figure 2.7 shows the hysteresis of NCp7 dependence on the concentration from 2nM to 30nM from non-equilibrium DNA stretching curves with protein. At first, the hysteresis will decrease with the concentration while above 5 nM, hysteresis reaches saturation at about 0.22.

Another important parameter that can be obtained from stretching curves is the transition slope, which is a primary quantitative characteristic of protein-DNA binding. In the presence of NCp7, the overstretching transition is significantly sloped when DNA elongates beyond its B-from contour length from 20-30 pN(67,90). Above 30 pN, DNA elongation results from stacking between dsDNA bases and a weak intercalation process. This duplex elongation induced primarily by base stacking and intercalation can only occur upon application of the force. In this case, the binding of NCp7 with DNA occurs as the protein inserts its two planar aromatic residues (Phe16 and Trp37) between the DNA base pairs (64,75,101,102). However, the ability of aromatic residues to stack with the bases is too weak to intercalate stable DNA or RNA duplex without loops or mismatches in the absence of force, so the elongation is only observed at high force.
Figure 2.8 shows the slope of the transition with different concentration for HIV-1 NCp7. The slope of the transition is measured as the tangent line to force-extension stretching curves at the transition midpoint about 0.455 nm/bp. The protein-free slope is approximately 21 pN/bm/bp. It can be seen the slope is increasing with the concentration, close to the protein-saturated maximum slope, which reflects the protein’s ability to intercalate dsDNA upon saturated binding.

![Figure 2.8](image)

Figure 2.8 HIV-1 NCp7 concentration dependence of the transition slope.

2.4 Quantifying the interaction kinetics between HIV-1 NCp7 and DNA by varying the DNA pulling rate

As pointed out previously, in the presence of NCp7, the two DNA strands do not separate during the overstretching. Ideally, if the protein-DNA binding is very fast and completely reaches equilibrium in the timescale of 10s seconds, the stretch and release curve should almost overlap, as is the case for simple intercalators, like ethidium (103). For the NCp7-DNA complex,
although the stretch and release curves are similar in shape, the release curve is always shifted to a longer extension, indicating more NC binding at any given force (Fig. 2.4 and Fig. 2.5(A)). Meanwhile, the difference in the extension between the stretch and release curves for the NCp7-DNA complex is small compared to the total NCp7 effect on DNA stretching (Fig. 2.5(B)). The result strongly implies there are two rates in the binding kinetics of NCp7-DNA, which we will denote “fast” and “slow”. Most NCp7 molecules bind and intercalate dsDNA rapidly, with a timescale of 10-100s seconds, leading to strong dsDNA elongation beyond its contour length. But there is also a small fraction of NC molecules that bind to DNA slowly with the timescale larger than 100-1000s, leading to apparent hysteresis between the stretch and release curves of NCp7-DNA complex.

To quantify the binding rates of NCp7-DNA complex, we repeat these measurements as a function of pulling rate. The force-extension curves under different pulling rate measurements show different hysteresis. In our experiment, the applied pulling rates are from 20 nm/s to 500 nm/s. Figure 2.9 shows typical DNA stretch and release cycles in the presence of 20 nM HIV-1 NCp7, performed at two different pulling rates: 50 nm/s and 500 nm/s.

The amount of NCp7 bound will be less than the equilibrium amount during stretching and more than the equilibrium amount during release, resulting in disagreement between the stretch and release curves, or hysteresis. Because the maximum hysteresis would be obtained for an infinitely slow protein that does not have time to bind during stretching or dissociate during release, we can use the hysteresis area ratio to obtain the non-equilibrium component of NCp7 bound. The hysteresis is quantified in Figure 2.10, showing the non-equilibrium component of NCp7 bound as a function of DNA stretching time, which we define as the time it takes for one single stretch or release cycle at a given pulling rate. This non-equilibrium fraction represents the fraction of NCp7 bound to DNA that exhibits binding kinetics slower than the timescale of stretching (104).
The time dependence of the non-equilibrium component of NCp7 bound, $F(t)$, exhibits two multiple timescales, likely due to the continuum of the sequence-dependent binding sites. $F(t)$ can be fit to a double exponential function:

$$F(t) = f_{\text{fast}} \cdot e^{-t/\tau_{\text{fast}}} + f_{\text{slow}} \cdot e^{-t/\tau_{\text{slow}}} \quad (\text{Eq. 2.1})$$

Where $f_{\text{fast}}$, $f_{\text{slow}}$ and $\tau_{\text{fast}}$, $\tau_{\text{slow}}$ are the fast and slow fractions of the non-equilibrium NC binding component and the corresponding relaxation times, respectively. The resulting parameters are $f_{\text{fast}} = 0.14 \pm 0.05$, $\tau_{\text{fast}} = 25 \pm 10$ s, $f_{\text{slow}} = 0.22 \pm 0.02$, and $\tau_{\text{slow}} = 1200 \pm 500$ s. Thus, the fast fraction is 0.15 with relaxation time of 25 s, and the slow fraction is 0.22 with relaxation time of 1200 s, indicating a large difference in timescales between two different measured processes. The pulling rate measurements show that DNA-NCp7 binding and dissociation occurs on multiple timescale and the slow rates occur on a timescale of over 1000 seconds.
Figure 2.9 DNA stretching (solid) and release (dash) curves in absence (black) and in the presence of 20nM NCp7, 50mM Na\(^+\), at 50 nm/s (blue), 500 nm/s (red). The faster pulling rate curve exhibits the larger hysteresis.

![Graph showing DNA stretching and release curves](image)

Figure 2.10 Non-equilibrium component of DNA-bound 20 nM HIV-1 NCp7 as a function of the DNA stretching time. The non-equilibrium component is obtained from the hysteresis area ratio of the DNA stretching curves with NCp7 under different pulling rates.

2.5 Multiple stretching experiments on the same DNA molecule with NCp7

The DNA force-extension curves are not well reproduced while the subsequent stretching curves of the same DNA molecule vary systematically. This is mostly due to the multiple timescales existing in the binding kinetics of NCp7-DNA that is demonstrated by pulling rate dependence studies. In order to fully understand the mechanism of interaction between NCp7 and DNA, we will show how these subsequent stretches behave on different timescales, before and after rinsing with buffer and competitor DNA (plasmid DNA - pbr322). These behaviors will also be characterized by transition slope and hysteresis from the DNA stretch and release curves with NCp7. The total time for this process during the experiment is up to more than 50 minutes. There
are primarily three stages for this process: First, we stretch the DNA with 20 nM NCp7 4 times; Second, we flow the 10 mM Hepes, 50 mM Na\(^+\), pH 7.5 buffer into the chamber to rinse out any protein remaining in the solution surrounding the single DNA molecule; Third, we flow the competitor DNA pbr322 to remove any remaining protein that is bound to the DNA; Fourth, we then stretch the DNA for two more cycles. Figure 2.11 (a) shows two typical DNA stretching curves with NCp7 before and after rinsing buffer. The red curve is the very first stretch before buffer rising and the blue curve is the last stretch after buffer rising. The elapsed time is about 35 minutes. Figure 2.11 (b) also shows two DNA stretching curves with NCp7. The yellow curve is the stretch before buffer rinsing, while the green curve is the stretch after buffer rinsing and waiting for an hour. Figure 2.11 (c) shows the typical DNA stretching curve after flowing competitor DNA.
Figure 2.11 (a) Stretching (solid line) and release (dashed line) in the presence of 20nM NCp7 in 10 mM HEPES, 50 mM Na\(^+\), pH 7.5 buffer with various wait times. (b) Stretching (solid line) and release (dashed line) in the presence of 20nM NC in 10 mM HEPES, 50 mM Na\(^+\), pH 7.5 buffer before and after rinsing, waiting for an hour. (c) Stretching (solid line) and release (dashed line) in the presence of 0.35 nM pbr322 competitor DNA.

The trend in this multiple stretching experiments is that both the transition slope and the hysteresis decrease with the time. Figure 2.12 shows the quantitative results for this process.
Stretching the DNA-NCp7 complex many times equilibrates binding and dissociates some of the extra NCp7. In 20 minutes, multiple stretches dissociate 20-30% of the NCp7. Rinsing with 50 mM Na\(^+\) buffer solution for 10 minutes dissociates an additional 10% of NCp7. Stretching for 15 minutes in the same buffer dissociates additional 10% of NCp7, but a few very slow and strongly bound NCp7 remain and produce measurable effects on the slope and hysteresis. These very slow strongly bound proteins will not come off from the DNA even for an hour. Washing with variable concentrations of the competitor DNA for about 5 minutes removes the NCp7 almost completely from the single stretched DNA molecule, after which the stretching curve behaves very similar to the DNA only curve. Figure 2.13 shows the comparison between changes in slope for DNA stretching curves of different concentrations of NCp7 and after flowing different concentrations of pbr322 competitor DNA. A few strongly bound NCp7 could only be dissociated with the flow of pbr322 competitor DNA.

On the other hand, we also try to rinse with 10 mM Mg\(^{2+}\) buffer instead of 50 mM Na\(^+\) buffer. The result is shown in figure 2.14. After rinsing the single DNA molecule with buffer containing Mg\(^{2+}\) for a very short time (less than 15 s), almost all the NCp7 molecules are washed off of the stretched DNA molecule. These results suggest the presence of an additional, macroscopic DNA binding mode that is much slower than the microscopic modes probed in most experiments. This additional mode is likely primarily electrostatic, as observed for other DNA binding proteins (10,14,97,98).
Figure 2.12 (a) (b) shows the change in slope and hysteresis of the DNA during the multiple stretches process. **0 min**: DNA only; **10 min**: 20nM NCp7 first stretch; **24 min**: 20nM NCp7 last stretch; **32 min**: first stretch after rinsing; **45 min**: last stretch after rinsing; **50 min**: stretch after flowing 0.87nM pbr322.

Figure 2.13 (a) shows the slope of DNA stretching curves with different concentration of NCp7. (b) shows the slope of DNA stretching curves after washing by competitor DNA pbr322 of different concentration. The few strongly bound NCp7 could be dissociated with the flow of competitor DNA pbr322.
Figure 2.14 DNA stretching curves of 20nM NCp7 in 50 mM Na\(^+\) buffer (red) and in 10nM Mg\(^2+\) buffer (green). Once with the Mg\(^{2+}\) buffer flowing for a very short time less than 15 s, almost all the NCp7 molecule are rinsed.

2.6 Discussion

In this chapter, we measure the binding properties and DNA interaction of HIV-1 NCp7 using single molecule biophysical methods. To be an optimal NA chaperone, a protein must be able to promote NA aggregation, destabilize dsDNA and exhibit rapid kinetics simultaneously, and NCp7 exhibits all the three components (13,67-70,83,105). However, while most of the DNA-bound NCp7 exhibits rapid kinetics (Fig. 2.10), a small component of bound NC dissociates very slowly from DNA. This component likely represent the component bound to high affinity sequence-specific sites, and this slow component may be important in determining the rate at which reverse transcription can proceed, such that NC may not represent a roadblock for reverse transcription (106,107).

From the stretching curves with the protein, it is shown that NCp7 can destabilize the B-form DNA duplex by deforming it, while keeping both strands together. Therefore, NCp7 can act as
an optimal NA chaperone by decreasing the NA refolding transition barriers as well as facilitating annealing of the new structure. The small hysteresis of the DNA stretch and release cycle, reaching saturation after 5 to 10 nM shown in figure 2.7, along with the lower force required to stretch DNA beyond its B-form contour length (>0.34 nm/bp) illustrates that HIV-1 NC destabilizes the DNA duplex, and these results were previously correlated with the optimum chaperone activity of this protein (70,92). It has been shown that both zinc fingers work together to facilitate rapid nucleic acid annealing, where both of the aromatics play a critical role in the NA chaperone activity (75). The ability of NCp7 to stack with nucleic acids requires the aromatic residues in the zinc finger. In contrast, the ability of NCp7 to facilitate strand re-annealing is primarily due to its cationic residues, and is related to protein’s ability to aggregate nucleic acids.

The change in slope from the DNA-NCp7 stretching curves reflects the ability of the protein to elongate dsDNA without allowing the strands to separate. This elongation at high forces represents DNA intercalation by NC, which only occurs at high forces. Therefore, the NCp7-DNA binding is strongly promoted by the stretching force, similar as the more conventional intercalators studied previously (95,108). These interactions also resemble the hemi-intercalation observed between the aromatic rings of the dsDNA intercalator ActD and ssDNA bases (44). While such intercalation is too weak to occur frequently at zero force when NAs are fully double-stranded, the intercalation measured in these single molecule experiments likely becomes important in cases of locally unstable elements of NA secondary structure, such as duplexes containing mismatches, loops or bulges.

To further quantify the kinetics of NCp7-DNA complex, we stretch the DNA molecule with protein under different pulling rates. Most of the DNA-bound NCp7 molecules are in equilibrium on the timescales of 10-100 s, as previously observed (67). These DNA-bound NCp7 molecules behave like mobile multivalent cations, optimizing their positions to maximize electrostatic self-attraction (109,110) and aggregate density (111-114). In contrast, there is also a small fraction of NC molecules that binds to DNA slowly with a timescale larger than 100-1000 s, exhibiting apparent hysteresis between stretch and release curves of NCp7-DNA complex.
The multiple stretching curves on one single DNA molecule with NCp7 further strengthen the point that multiple binding and association rates exist in the NCp7-DNA complex. The strongly bound NCp7 molecules corresponding to the slow fraction of the long timescale of 1000 s will stay bound even after rinsing with Na\(^+\) buffer. However, rinsing with competitor DNA or Mg\(^{2+}\) removes the bound NC. This result shows that many of these slowly bound proteins remain bound to DNA primarily electrostatically. Therefore, the bound NC can be exchanged for another highly charged ligand, Mg\(^{2+}\). Alternatively, when another DNA molecule collides with the NC-bound and stretched single DNA molecule, the bound NC can transfer to the other DNA molecule. Therefore, the complex kinetics of NC binding are determined by the interplay between strong electrostatic binding as well as strong non-electrostatic interactions at specific sites on DNA.
Chapter 3

DNA binding properties and kinetics of wild type and mutant HIV-1 Gag

3.1 Background on wild type and mutant HIV-1 Gag

The structure of HIV-1 Gag is shown in Figure 2.1(b). NC is first synthesized as a domain of HIV-1 Gag. During viral maturation, the viral protease cleaves Gag into MA, CA, NCp7, p6 and two small peptides, spacer peptide 1 and 2 (SP1 and SP2). HIV-1 protease first cleaves after SP1, generating a form of NC called NCp15, which consists of mature NCp7 linked to SP2 and p6. Further processing of NCp15 results in NCp9 (NCp7+SP2), and finally cleavage between NCp7 and SP2 results in release of mature NCp7 (115-117). While NCp7 is known as a good nucleic acid chaperone, HIV-1 Gag primarily functions as a nucleic acid binding and packaging protein (65,70,118,119). HIV-1 Gag is capable of packaging viral RNA from a large pool of cellular nucleic acid molecules and it relies on the interaction of RNA with the two zinc fingers of NC domain and also with its cationic residues. For all the Gag studies, we will use the Gag-Δp6 protein that has been used in most in vitro studies, and refer to them as Gag for simplicity. Figure 3.1 shows examples of Gag wild type and mutants studied here.

Figure 3.1 HIV-1 Gag wild type and mutants
CANC is the mutation that has the MA domain removed, resulting in Gag containing only the CA and NC domains. For DZF1 and DZF1+2, either the first zinc finger or both of the zinc fingers have been removed in the NC domain of Gag, respectively. Previous research shows that CANC is much more effective at tRNA annealing (120,121). The result suggests that the MA domain interferes with annealing due to additional interactions with DNA, causing Gag’s dissociation rate to be slow. By single molecule stretching methods, the role of the individual domains in Gag’s DNA binding will be further elucidated.

3.2 Interactions of HIV-1 Gag wild type with DNA

In order to characterize the extent to which HIV-1 Gag act as a nucleic acid chaperone, we first use single molecule stretching to determine how Gag alters the stability of dsDNA and also the kinetics of DNA binding and annealing. Fig 3.2 (a) shows a DNA stretching and release cycle in the presence of 2 nM wild type HIV-1 Gag using optical tweezers. In the presence of HIV-1 Gag, the transition is altered significantly. First, HIV-1 Gag shows very large aggregation, significant increase in the plateau and reduction in the ssDNA contour length. The large aggregation could possibly make it a chaperone when duplex stabilization is not needed. Second, HIV-1 Gag stabilizes the DNA and makes the force plateau much higher. Third, the large amount of hysteresis indicates that wild type HIV-1 Gag exhibits slow kinetics. Rapid kinetics is an important component of chaperone activity that describes a protein’s ability to rapidly bind to and dissociate from single- and double-stranded nucleic acids, and is correlated with nucleic acid annealing activity. Gag’s slow kinetics indicates it is unlikely to be a good chaperone when rapid protein binding and annealing is needed.
Figure 3.2 DNA stretching (solid line) and release (dash line) curves in the presence (red) and absence (black) of (a) 2 nM HIV-1 Gag (b) 20 nM HIV-1 NCp7.

To compare HIV-1 and NCp7 interactions with DNA, we find that NC proteins interact with most of the DNA molecules simultaneously, extend dsDNA easily at relatively low forces, and destabilize the DNA, with fast kinetics by showing small hysteresis. On the other hand, Gag binds 10 times stronger than NCp7 since the concentration of NCp7 is 10 times the concentration of Gag, while showing similar effect on the stretching curves (70). Gag proteins form 3D clusters called virus-like particles (VLPs) (122). The VLPs prevent strands separation and make the strand unwinding and this torsional constraint will lead to a higher melting force, stabilizing the DNA (123). Also, the HIV-1 Gag-DNA complex shows much larger hysteresis in the stretching curve, indicating its slow kinetics, consistent with observations that Gag can form a roadblock for some reverse transcription processes (124). Figure 3.3 shows the different structures of binding simplified between HIV-1 NCp7 and Gag.
3.3 Interactions of HIV-1 Gag mutants with DNA

3.3.1 CANC

CANC is the mutation with removal of the MA domain. Figure 3.4 shows the stretching and release cycle of 15 nM CANC. We increase the concentration in experiments from 2 nM which is the concentration used for Gag with no obvious effect, until at 15 nM we see an increasing slope of the plateau. Without the MA domain, the CANC stretching curves are much more similar to the DNA only curve than to that observed in the presence of Gag. The binding of CANC-DNA is weak in two aspects: a small slope at high concentration and a small amount of hysteresis. This shows the MA domain is also capable of NA binding and MA could be the major factor for slow kinetics of Gag-DNA interactions. This result is consistent with \textit{in vivo} studies, which showed that the MA domain of Gag inhibits $tRNA_{Lys}$ annealing (125). Therefore both MA and NC bind to nucleic acids and the MA binding causes the Gag annealing rate to be slow, inhibiting its chaperone activity.
3.3.2 HIV-1 DZF1 and DZF1+2

HIV-1 Gag mutants DZF1 and DZF1+2 are two mutants in which either the first zinc finger or both zinc fingers have been removed in the NC domain of Gag. Zinc fingers are found to play crucial roles throughout the HIV-1 virus replication cycle and required for nucleic acids chaperone activity (90). Research in vivo also found deleting either one or the two NC zinc fingers leads to an unexpected premature viral DNA synthesis and the production of non-infectious particles with a high level of viral DNA (126). By studying these two mutants using a single molecule method with optical tweezers, we show how the zinc fingers affect the interaction between HIV-1 Gag and DNA.

Figure 3.5 shows the stretching and release cycle of 6 nM DFZ1 and DZF1+2. Compared with HIV-1 Gag, they all stabilize the DNA duplex and show similar slope. However, the aggregation effect is much weaker at low force. When deleting the first zinc finger, the hysteresis stays about the same. When deleting the both zinc fingers, the hysteresis becomes much larger, showing
obvious slower kinetics. Therefore, at least one zinc finger is required for Gag to exhibit wild-type kinetics. Both zinc fingers appear to be required for wild-type level binding affinity.

Figure 3.5 DNA stretching (solid line) and release (dash line) curves in the presence (red) and absence (black) of (a) 6 nM DFZ1; (b) 6 nM DFZ1+2

3.4 Discussion

In this chapter, we summarize some general characteristics of interactions between DNA and HIV-1 Gag wild type and mutants. HIV-1 Gag is a much larger protein in the structure that NCp7, the binding is 10 times stronger, showing larger aggregation, stabilizing the DNA duplex and slowing the kinetics. To quantify the DNA-protein kinetics, we can use the same method given in chapter 2 to calculate the hysteresis area ratio. The result is shown in Table 1:
<table>
<thead>
<tr>
<th></th>
<th>2nM</th>
<th>6nM</th>
<th>15nM</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gag Wild Type</strong></td>
<td>0.24 ± 0.03</td>
<td>0.23 ± 0.03</td>
<td>NA</td>
</tr>
<tr>
<td><strong>DZF1</strong></td>
<td>0.18 ± 0.02</td>
<td>0.21 ± 0.03</td>
<td>0.19 ± 0.02</td>
</tr>
<tr>
<td><strong>DZF1+2</strong></td>
<td>0.14±0.02</td>
<td>0.30±0.04</td>
<td>0.38±0.06</td>
</tr>
<tr>
<td><strong>CANC</strong></td>
<td>0.10±0.02</td>
<td>0.10±0.02</td>
<td>0.13±0.02</td>
</tr>
</tbody>
</table>

Table 1 Hysteresis area ratio of HIV-1 Gag wild type and mutants in different concentration

It can be concluded that higher binding affinity might make HIV-1 Gag an effective chaperone when duplex destabilization and rapid protein-DNA interactions kinetics are not needed. From the hysteresis area ratio, the MA domain is also shown to be another major factor in NA binding. In addition, deleting the two zinc fingers further slows the kinetics of protein-DNA interactions.
Chapter 4

Distinct nucleic acid interaction properties of HIV-1 nucleocapsid protein precursor NCp15 explain defects in viral infectivity

Portions of this work were originally published in Nucleic Acids Res, 42, 7145-7159 (2014), and have been adapted for this dissertation.

4.1 Backgrounds on HIV-1 NCp7, NCp9 and NCp15

During HIV-1 maturation, three different forms of nucleocapsid (NC) protein, NCp15 (p9 + p6), NCp9 (p7 + SP2), and NCp7 appear successively, shown in figure 2.1 (b). It is known HIV-1 NCp7’s chaperone function is essential during reverse transcription of the highly structured viral RNA genome (vRNA) (127). A previous study shows the mutation of protease cleavage site proximal to the NC domain have different effects on HIV-1 infectivity (128). When the site between SP2 and p6 are mutated to eliminate viral protease cleavage, viral infectivity dropped over six orders of magnitude. However, when the site between NCp7 and SP1 was mutated, the virus is still viable. Therefore HIV-1 containing NCp15 is not infectious, while a virus with NCp9 behaves similar to the wild type virus.

HIV-1 NCp15 has some resemblance to HTLV-1 NC, which both has a basic N-terminal domain (NTD) containing the two zinc finger motifs and an acidic C-terminal domain (CTD). The acidic CTD of HITLV-1 NC will negatively regulate its chaperone activity (39). Based on a fluorescence study, the CTD folds back and interacts with the zinc finger domain (129). Additional biochemical studies suggest that the basic N-terminal domain of HTLV-1 NC interacts with the CTD either intramolecularly or intermolecularly, in the absence or presence of NAs, respectively. In this study, we will use single molecule method with optical tweezers to investigate the protein effects on DNA stretching curves.

4.2 Quantifying the general characteristics of DNA stretching with HIV-1 NCp7, NCp9 and NCp15
In this work, we measure the DNA stretching curves under different concentration of HIV-1 NCp7, NCp9 and NCp15. As it is pointed out previously, the two major features of the DNA force-extension curves that characterize the DNA protein interaction is the observed hysteresis and slope. The transition slope reflects the protein’s ability to intercalate dsDNA upon saturated binding. By quantifying the amount of hysteresis, the kinetics of interaction between DNA and NC can be characterized.

Figure 4.1 shows the stretching and release cycle for bacteriophage λ DNA in the absence of the protein and in the presence of 20nM HIV-1 NCp7, NCp9 and NCp15 respectively. All three proteins show a large overstretching transition slope, indicating that all three proteins intercalate DNA under force. This is to be expected, as all three proteins retain the aromatic residues on the zinc fingers that are believed to be responsible for intercalation(75). In contrast, NCp9 and NCp15 show a larger amount of hysteresis, which indicates they may not act as optimum chaperone proteins when rapid kinetics is needed, such as for the minus-strand transfer step during reverse transcription.
Figure 4.1 DNA stretching (solid line) and release (dash line) curves in the presence (red) and absence (black) of (a) 20 nM NCp7; (b) 20 nM NCp9; (c) 20 nM NCp15.

Figure 4.2 shows the dependence of the slope of the transition on concentration for HIV-1 NCp7, NCp9 and NCp15. The slope during the overstretching transition is measured as the slope of the tangent line to force-extension curve at the transition midpoint ~0.455 nm/bp. This parameter reflects the protein’s ability to intercalate dsDNA upon saturated binding. For all the 3 proteins, the slope increases with the concentration. For all three proteins, the slope appears to be reaching saturation after 30 nM. Although NCp15 shows the greatest saturated slope, all three proteins show very high saturated slopes, indicating that all three proteins bind strongly to DNA and intercalate under force.
Figure 4.2 Slope of the transition versus concentration curves for HIV-1 NCp7 (green), NCp9 (blue) and NCp15 (red).

Figure 4.3 shows the hysteresis area ratio dependence on concentration for HIV-1 NCp7, NCp9 and NCp15. The way to quantify hysteresis area ratio is described in Figure 2.6. This parameter represents the non-equilibrium component of DNA-bound to the protein. Measuring the hysteresis area ratio allows us to estimate the overall protein-DNA interaction kinetics. The larger it is, the slower the interaction kinetics.
The combination of a large transition slope and small hysteresis observed for all three proteins indicates that the DNA is slightly destabilized by protein binding, but intercalation keeps the strands together, which should facilitate nucleic acid rearrangement. The fact that NCp15 has the greatest hysteresis despite also strongly intercalating DNA under force indicates that this proteins intercalates only very slowly. The slow kinetics of NCp15 intercalation suggest that its nucleic acid chaperone activity will be inhibited for processes that require rapid protein-DNA interactions, as suggested above. Below we will quantify these slow kinetics to better understand the extent to which this protein is expected to show defects in nucleic acid chaperone activity.

4.3 Quantifying the interaction kinetics between HIV-1 NC proteins with DNA by pulling rate measurements

As it is discussed previously, when protein-DNA binding is very fast and completely reaches equilibrium on a timescale of tens of seconds, the stretch and release curves should almost overlap. For NC-DNA complex, though the stretch and release curves are similar in shape, the release curve is always shifted to a longer extension, indicating more NC binding at any given
force (Fig. 4.1 and Fig. 2.4(A)) Meanwhile, the difference in the extension between the stretch and release curves of NC-DNA complex is small compared to the total NC protein effects on DNA stretching (Fig. 2.5(B)). The result strongly implies there are multiple rates in the binding rates of NC-DNA: “fast” and “slow”. Most NC protein molecules bind and intercalate dsDNA rapidly, with a timescale of tens to hundreds of seconds, leading to strong dsDNA elongation beyond its contour length. However, there is also a small fraction of NC molecules that binds to DNA slowly, with a timescale larger than hundreds to thousands of seconds, leading to apparent hysteresis between stretch and release curves of the NC-DNA complex.

To quantify the binding rates of NC-DNA complex, we apply the pulling rate dependence measurements. The force-extension curves under different pulling rate measurements show different hysteresis. In this experiment, the applied pulling rates are from 20 nm/s to 500 nm/s and figure 4.4 shows the typical DNA stretch and release cycles in the presence of 20 nM HIV-1 NCp7, NCp9, and NCp15 respectively, performed at two different pulling rates: 50 nm/s and 500 nm/s.
Figure 4.4 DNA stretching (solid) and release (dash) curves in absence (black) and in the presence of (a) 20 nM NCp7; (b) 20 nM NCp9; (c) 20 nM NCp15; in 50mM Na⁺, at 50 nm/s (blue), 500 nm/s (red). The faster pulling rate curve exhibits the larger hysteresis.

All three proteins lengthen the DNA at high forces due to intercalation (94,95,108), while also destabilizing the DNA duplex (45). In the same way, we can use the hysteresis area ratio to obtain the non-equilibrium component of NC bound which is quantified in figure 4.5, showing
the non-equilibrium component of NCp7, NCp9 and NCp15 bound as a function of DNA stretching time, the time it takes for one single stretch or release cycle at a given pulling rate. This non-equilibrium fraction represents the fraction of NC bound to DNA that exhibits binding kinetics slower than the timescale of stretching. For all measured stretching times, the non-equilibrium component is at least 44\% greater for NCp15 relative to NCp7.

The time dependence of the non-equilibrium component of NC bound, \( F(t) \), exhibits two multiple timescales, likely due to the continuum of the sequence-dependent binding sites. \( F(t) \) can be fit to a double exponential function in Eq 2.1:

\[
F(t) = f_{\text{fast}} e^{-t/t_{\text{fast}}} + f_{\text{slow}} e^{-t/t_{\text{slow}}} \quad \text{(Eq 2.1)}
\]

where \( f_{\text{fast}}, f_{\text{slow}} \) and \( t_{\text{fast}}, t_{\text{slow}} \) are the fast and the slow fractions of the non-equilibrium NC binding component and the corresponding relaxation times, respectively. The resulting parameters are given in Table 2. The non-equilibrium components of the DNA-bound NC molecules have fast fractions varying from 0.14 to 0.19, with relaxation times ranging from 25 to 45 s, and slow fractions varying from 0.22 to 0.30, with relaxation times in the range of 1000 to 5000 s. NCp15 exhibits a significantly larger slow fraction of \(~0.3\), relative to \(~0.2\) for NCp7 and NCp9.
Figure 4.5 Non-equilibrium component of DNA-bound 20 nM HIV-1 NCp7 (green), NCp9 (blue) and NCp15 (red) as a function of the DNA stretching time. The non-equilibrium component is obtained from the hysteresis area ratio of the DNA stretching curves with NC proteins under different pulling rates.

<table>
<thead>
<tr>
<th>Protein, 20nM</th>
<th>$f_{fast}$</th>
<th>$\Delta_{fast}, s$</th>
<th>$f_{slow}$</th>
<th>$\Delta_{slow}, s$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCp7</td>
<td>0.14±0.05</td>
<td>25±10</td>
<td>0.22±0.02</td>
<td>1,200±500</td>
</tr>
<tr>
<td>NCp9</td>
<td>0.16±0.05</td>
<td>45±15</td>
<td>0.23±0.02</td>
<td>5,220±1000</td>
</tr>
<tr>
<td>NCp15</td>
<td>0.19±0.05</td>
<td>40±15</td>
<td>0.30±0.02</td>
<td>3,370±1000</td>
</tr>
</tbody>
</table>

Table 2. Kinetic parameters describing the non-equilibrium component of NC proteins bound to DNA

4.4 Discussion

In this chapter, we compared the protein-DNA interactions of HIV-1 NCp7, NCp9 and NCp15. In the previous study, sedimentation assays and dynamic light scattering experiments showed that NCp7 and NCp9 both promote the formation of large protein-RNA aggregates. In contrast, NCp15-RNA complexes failed to scatter light effectively (104). This result is consistent with the
previous electron microscopy study showing that NCp15 binds ssDNA like “beads on a string”, unlike NCp7 or NCp9, which form condensed aggregates (130,131).

The distinct NA interaction of NCp15 is regulated by the acidic p6 domain. The NMR studies provided support for a fold-back conformation in solution for NCp15, wherein the acidic p6 domain interacts with the basic ZF domain. The interaction seemed to be electrostatic, since at 500 mM NaCl the chemical shift perturbations are weaker. Neutralizing the C-terminal acidic residues in p6 improves NCp15’s chaperone function. Adding p6 peptide in trans also partially inhibits the annealing activity of NCp7 (104).

In the single molecule DNA stretching studies here, HIV-1 NCp7, NCp9 and NCp15 show similar overall features from the titrated stretching curves. The significant change in the transition slope shows all three proteins strongly bind and intercalate the dsDNA, intercalation that is also promoted when increasing the concentration from 2 nM to 30 nM. Furthermore, the hysteresis reaches saturation at 5 to 10 nM for all three NC proteins, while HIV-1 NCp15 has the largest hysteresis area ratio, indicating additional slow interactions with DNA. This result is further quantified by stretching the DNA molecules with protein under different pulling rates.

In the pulling rate dependence measurement, we find NCp15 molecules lead to slower protein dissociation from NA compared to NCp7 and NCp9. According to the hysteresis observed in Table 2, most of the DNA-bound NC molecules are in equilibrium on the timescales of 10-100 s, as previously observed (67). These DNA-bound NC molecules behave like mobile multivalent cations, optimizing their positions to maximize electrostatic self-attraction (109,110) and aggregate density (111-114). NCp15 has the largest overall non-equilibrium fraction of DNA-bound NC molecules. In addition, NCp15 also has a larger fraction of DNA-bound NC molecules which dissociate on slow timescales of ~1000 s, and do not contribute to self-attraction and aggregation. A critical concentration of highly charged mobile cations is known to be required for NA aggregation (109). This condition appears to be satisfied for NCp7 and NCp9, but not for NCp15.
The observed slower NCp15 dissociation kinetics relative to NCp7 and NCp9 is likely due to the formation of additional electrostatic inter-protein contacts between the cationic NTD and anionic CTD domains of NA-bound NCp15 molecules. Acidic CTD domains will lead to cooperative binding and slow NA dissociation. The schematic model is shown in figure 4.6. Similar CTD/NTD interactions have been proposed for HTLV-1 NC (39,129). HTLV-1 NC is a poor NA chaperone, but deletion of its acidic CTD will restore chaperone function (39,132). The presence of the CTD also greatly slows down the NA interaction kinetics of HTLV-1 NC. Therefore, for both HTLV-1 NC and HIV-1 NCp15, the acidic CTDs will negatively regulate their chaperone activity and NA binding kinetics, though the effect in the case of NCp15 is milder. It is known that the interaction between macromolecules correlates with their charge density not their net charge (133,134). The region of highest negative charge density within the CTD of HTLV-1 NC contains 5 anionic residues. In contrast, the highest negative charge density region of p6 has 5 negatively charged amino acids and one positive residue. As a result, NCp15 shows a slower kinetics compared to NCp7 and NCp9, but the differences between the slow fractions of DNA-bound are moderate.

![Figure 4.6 Schematic model of NCp15 binding to DNA: NCp15 unbound and bound. Additional electrostatic inter-protein contacts between the cationic NTD and anionic CTD domains of NA-bound NCp15 molecules result in slow kinetics](image-url)
Chapter 5

DNA binding properties of HIV Viral protein R (Vpr)

5.1 Background on HIV-1 Vpr

HIV-1 Vpr is a 96 amino acid 14 kDa protein. It is an accessory protein and plays an important role in regulating nuclear import of the During the HIV-1 retroviral cycle. Vpr mediates many processes that aid HIV-1 infection, evasion of the immune system, and persistence in the host, thus contributing to the morbidity and mortality of acquired immunodeficiency syndrome (135-138). Figure 5.1 shows the several functions of Vpr during the HIV-1 viral lifecycle (139). Quantitative understanding of Vpr will help target efficiently the virus and reduce the proliferation of HIV-1. Single molecule stretching method can provide quantitative analysis on Vpr-DNA interactions.

Figure 5.1 Vpr performs several functions in the early stages of the HIV-1 life cycle. Schematic of the early stages of HIV-1 infection. ❶ A mature virus containing Vpr proteins arrives to the host-cell. ❷ The virus gets into the host-cell by interacting with cell-surface receptors and co-receptors and introduces the viral RNA and the viral proteins. ❸ Vpr modulates reverse transcription. ❹ Vpr binds to the pre-integration complex (PIC) and facilitates the translocation of the viral DNA into the nuclear pore. ❼ Vpr
is involved in the transactivation of both HIV LTR and host cell genes. (Adapted from a drawing by Divakaran Murugesapillai, Williams lab, Northeastern University).

5.2 Single molecule studies of Vpr-DNA interactions

We measure the DNA stretching curves with different concentrations of the Vpr protein from 10 nM to 100 nM. All of the curves show similar effect that stretching forces increase significantly below the DNA contour length at 0.34 nm/bp. Figure 5.2 shows the example at 50 nM Vpr. From this feature, we expect the Vpr protein will condense the DNA molecule. In these stretching experiments, we minimized the overall condensation and aggregation forces by holding the DNA molecule at 30 pN while flowing protein into the sample chamber.

![Figure 5.2 DNA stretching (solid) and release (dash) curves in absence (black) and in the presence of 50 nM Vpr (red). The curve shows significant stretching forces below the DNA contour length of 0.34 nm/bp indicating Vpr strongly condenses DNA.](image-url)
5.2.1 Formation of loops on DNA in the presence of Vpr

When a single DNA molecule is held at low forces below 2 pN, the subsequent stretching curves do not change. However, when the DNA molecule is held at such low forces below 2 pN for about ten minutes in the presence of the Vpr protein, the force-extension curve exhibits jumps when pulled at high pulling rates of 240 nm/s. These jumps show there are breakings of loops on the DNA molecule. The formation of such loops is likely due to the flexibility of DNA (140) and the DNA molecule is also stabilized by Vpr. Figure 5.3 (a) shows the example of formation of loops at low forces on the DNA molecule with 100 nM Vpr. Figure 5.3 (b) shows the distribution of forces at which loops break. The most probable force of these loops is around 9 pN. This force is slightly lower than that observed for the chromatin architectural protein HMO1 using a similar method (141).

Figure 5.3 (a) DNA stretching (solid) and release (dash) curves in absence (black) and in the presence of 100 nM Vpr (red). The DNA molecule is held at low extension (<0.25nm/bp) when flowing the protein. Loops are formed at different locations below 25 pN (b) Histogram of loop formation and breaking forces
A histogram of loop forces for 42 unbinding events is recorded in figure 5.3 (b). During the experiment, we found the formation of loops needs an incubation time of about 7 minutes holding the DNA molecule at a short extension around 0.25 nm(bp. Also, the breaking of loops is favorable on a comparatively high pulling rate, 240 nm/s and the breaking events happen at a force of $9.5 \pm 0.5$ pN.

5.2.2 DNA compaction with Vpr

The protein induced looping suggest the Vpr protein can compact the DNA molecule at low forces. To further investigate the kinetics of DNA compaction by Vpr, we measured the time dependence of DNA extension in the presence of 100 nM Vpr at a constant force of 10 pN. In figure 5.4 (a), the first stretch (black) is DNA only and the subsequent stretch (orange) is overlapped with the first up to 10 pN. Then, the DNA molecule is held at 10 pN and exposed to 100 nM Vpr resulting an obvious decrease in the DNA extension in a couple of seconds. The release curve (dash orange) no longer overlaps with the DNA only curve. The result shows the DNA molecule has been compacted by Vpr at such low forces. The rate of the compaction is fitted to single exponential in figure 5.4 (b). We found the time constant for the compacting process is $6.6 \pm 0.5$ s. Meanwhile, we also measured the DNA compaction at 15pN and the time constant is $16.4 \pm 1.4$ s. This suggests the rate of compaction is force dependent.
Figure 5.4 (a) Compaction of a single DNA molecule by Vpr by constant force measurements. The orange solid line shows the stretch of the DNA molecule before flowing 100nM Vpr. The orange dash line shows the release curve after the DNA molecule being held at a constant force of 10pN for several minutes. (b) The green solid line shows the compaction of the DNA molecule by Vpr. The extension decreased with time and the red line is the single exponential fit for the compaction process.

5.3 Discussion

In this work, we characterize Vpr-DNA interactions by single molecule force spectroscopy. When the DNA molecule is held at low force below 2 pN, the Vpr protein is found to stabilize DNA loops. Under such condition, we observe an increase in force at extension less than the DNA contour length, followed by a sudden decrease in force as the loops break during extension. We also observe the DNA shortening at a constant force of 10 pN in the presence of Vpr. Vpr initially compacts DNA on the timescale of several seconds, followed by Vpr mediated DNA looping on the timescale of several minutes. These characteristics are found to be similar with the yeast architectural protein HMO1, which has been studied previously using a similar method. The most probable force to break the loop is around 9.5 pN, similar to forces exerted by molecular motors such as RNA polymerase (142-144). Therefore, the presence of Vpr is not likely to represent a block to reverse transcription. However, the compact form induced by Vpr is
likely very helpful to facilitate transport of viral DNA into the nuclear pore after reverse transcription is complete (138,145,146).
Chapter 6

Conclusions

In this work, we use single molecule stretching experiments to investigate the NA chaperone activity of HIV-1 NC proteins, HIV-1 Gag proteins and its mutants. Many steps of the virus replication process require NC’s chaperone activity to facilitate NA rearrangement. Our single molecule DNA stretching experiments with HIV-1 NCp7 suggest it is an optimal chaperone protein, but that its NA interaction kinetics are complex. Rapid DNA annealing kinetics, as one of the three major components of nucleic acid chaperone activity, is further elucidated by measuring pulling rate dependence and multiple stretches on one DNA molecule. Surprisingly, although most of the bound NC molecules exhibit rapid kinetics, a small component exhibits slow kinetics, and this slow kinetics may have a significant effect on reverse transcription.

We have also applied our newly developed method for measuring NA interaction kinetics to NCp9 and NCp15 to determine how the progressive cleavage of the Gag protein may regulate HIV-1 replication. The results show NC proteins have multiple binding and annealing rates. The time dependence of the non-equilibrium component of NC bound exhibits two multiple timescales. The non-equilibrium components of the DNA-bound NC molecules have fast fractions with relaxation times ranging in 25 to 45 s, and slow fractions with relaxation times in the range of 1000 to 5000 s. The fast and slow fraction of DNA bound is further illustrated by a timeline of multiple stretching curves on the same DNA molecule with NCp7.

In the pulling rate dependence measurements, NCp15 exhibits a significantly larger slow fraction than NCp9 and NCp7, mainly due to the formation of additional electrostatic inter-protein contacts between the cationic NTD and anionic CTD domains. This result is consistent with the research in vivo and NMR study.

In order to better understand the mechanism of the function of each specific domain of HIV-1 Gag in the retroviral process, we also compare the characteristics of stretching curves of HIV-1 Gag and its mutants, including CANC, DZF1 and DZF 1+2. Single molecule experiments on these proteins demonstrate that the MA domain is also capable of binding to DNA other than
NC, and the two zinc fingers play a significant role for the rapid kinetics of protein DNA interactions. Different effects in NA aggregation, NA destabilization and rapid kinetics suggest these proteins are not optimal chaperones compared to HIV-1 NCp7. Moreover, we also apply single molecule methods to study the HIV-1 Vpr protein interaction with DNA combines DNA compaction and looping.

These new approaches for studying DNA binding and dissociation kinetics provide a way to elucidate the biophysical mechanisms regulate the DNA interaction properties of HIV-1 NC, Gag and Vpr proteins involved in complex retroviral replication processes. Quantitative characterization of these mechanisms provides fundamental insight into protein function, which may help in the development of anti-HIV drugs.
References


