Optical tweezers
reveal how proteins alter replication

A dissertation presented by

Kathy Chaurasiya

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, MA
April 18, 2013
Optical tweezers
reveal how proteins alter replication

by

Kathy Chaurasiya

Abstract of Dissertation

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Physics

in the College of Science of

Northeastern University

April 18, 2013
Abstract

Single molecule force spectroscopy is a powerful method that explores the DNA interaction properties of proteins involved in a wide range of fundamental biological processes such as DNA replication, transcription, and repair. We use optical tweezers to capture and stretch a single DNA molecule in the presence of proteins that bind DNA and alter its mechanical properties. We quantitatively characterize the DNA binding mechanisms of proteins in order to provide a detailed understanding of their function. In this work, we focus on proteins involved in replication of *Escherichia coli* (*E. coli*), endogenous eukaryotic retrotransposons Ty3 and LINE-1, and human immunodeficiency virus (HIV).

DNA polymerases replicate the entire genome of the cell, and bind both double-stranded DNA (dsDNA) and single-stranded DNA (ssDNA) during DNA replication. The replicative DNA polymerase in the widely-studied model system *E. coli* is the DNA polymerase III subunit α (DNA pol III α). We use optical tweezers to determine that UmuD, a protein that regulates bacterial mutagenesis through its interactions with DNA polymerases, specifically disrupts α binding to ssDNA. This suggests that UmuD removes α from its ssDNA template to allow DNA repair proteins access to the damaged DNA, and to facilitate exchange of the replicative polymerase for an error-prone translesion synthesis (TLS) polymerase that inserts nucleotides opposite the lesions, so that bacterial DNA replication may proceed. This work demonstrates a biophysical mechanism by which *E. coli* cells tolerate DNA damage.

Retroviruses and retrotransposons reproduce by copying their RNA genome into the nuclear DNA of their eukaryotic hosts. Retroelements encode proteins called nucleic acid chaperones, which rearrange nucleic acid secondary structure and are therefore required for successful replication. The chaperone activity of these proteins requires strong binding affinity for both single- and double-stranded nucleic acids. We use single molecule DNA stretching to show that the nucleocapsid protein (NC) of the yeast retrotransposon Ty3, which is likely to be an ancestor of HIV NC, has optimal nucleic acid chaperone activity with only a single zinc finger. We also show that the chaperone activity of the ORF1 protein is responsible for successful replication of the mouse LINE-1 retrotransposon. LINE-1 is also 17% of the human genome, where it generates insertion mutations and alters gene expression. Retrotransposons such as LINE-1 and Ty3 are likely to be ancestors of retroviruses such as HIV.

Human APOBEC3G (A3G) inhibits HIV-1 replication via cytidine deamination of the viral ssDNA genome, as well as via a distinct deamination-independent mechanism. Efficient deamination requires rapid on-off binding kinetics, but a slow dissociation rate is required for the proposed deaminase-independent mechanism. We resolve this apparent contradiction with a new quantitative single molecule method, which shows that A3G initially binds ssDNA with fast on-off rates and subsequently converts to a slow binding mode. This suggests that oligomerization transforms A3G from a fast enzyme to a slow binding protein, which is the biophysical mechanism that allows A3G to inhibit HIV replication. A complete understanding of the mechanism of A3G-mediated antiviral activity is required to design drugs that disrupt the viral response to A3G, enhance A3G packaging inside the viral core, and other potential strategies for long-term treatment of HIV infection.

We use single molecule biophysics to explore the function of proteins involved in bacterial DNA replication, endogenous retrotransposition of retroelements in eukaryotic hosts such yeast and mice, and HIV replication in human cells. Our quantitative results provide insight into protein function in a range of complex biological systems and have wide-ranging implications for human health.
# Table of Contents

Abstract .......................................................................................................................... 3

Table of Contents ........................................................................................................... 5

1 Introduction .................................................................................................................... 7
   1.1 DNA carries the instructions for life ........................................................................ 7
   1.2 DNA replication ...................................................................................................... 8
   1.3 Transcription and translation ................................................................................. 9
   1.4 Reverse transcription ............................................................................................. 10

2 Methods ......................................................................................................................... 12
   2.1 Single molecule force spectroscopy ...................................................................... 12
   2.2 The optical trap: a spring made of light ................................................................. 13
   2.3 Stretching single DNA molecules ......................................................................... 16
   2.4 Models of polymer elasticity ................................................................................. 16
      2.4.1 Force-induced structural transitions .............................................................. 17
      2.4.2 Glyoxal binds ssDNA bases exposed in the force-induced melting transition 18
      2.4.3 Visualizing force-induced melting with intercalators and SSBs ....................... 19
   2.5 Biophysical characterization of DNA binding proteins ........................................ 20
      2.5.1 The replicative DNA polymerase of *E. coli* ................................................. 20
      2.5.2 Retroviral nucleocapsid proteins .................................................................... 22
      2.5.3 The ORF1 protein from the retrotransposon LINE-1 ....................................... 26

3 Polymerase manager protein UmuD regulates *E. coli* DNA polymerase III α binding to ssDNA ........................................................................................................ 29
   3.1 Introduction ............................................................................................................. 30
   3.2 Results ..................................................................................................................... 31
      3.2.1 UmuD inhibits α binding to ssDNA ................................................................. 31
      3.2.2 Specific UmuD variants disrupt the UmuD-α interaction ............................... 32
      3.2.3 Specific variants disrupt UmuD inhibition of α binding to ssDNA ................. 34
   3.3 Discussion ............................................................................................................... 35
   3.4 Materials and Methods ......................................................................................... 37
      3.4.1 Proteins and plasmids .................................................................................... 37
      3.4.2 Single molecule DNA stretching .................................................................... 37
      3.4.3 Protein-protein docking ................................................................................ 38
      3.4.4 Thermal stability assay .................................................................................. 38
      3.4.5 RecA/ssDNA facilitated cleavage assay ......................................................... 38
      3.4.6 Tryptophan fluorescence assay ..................................................................... 38

4 A single zinc finger optimizes the DNA interactions of the nucleocapsid protein of the yeast retrotransposon Ty3 ........................................................................ 40
   4.1 Introduction ............................................................................................................ 41
   4.2 Results ..................................................................................................................... 42
      4.2.1 Wild-type Ty3 NC binds DNA with rapid kinetics and slightly stabilizes dsDNA 42
      4.2.2 Cationic tail deletions destroy the rapid kinetics and duplex stabilization of wild-type Ty3 NC 44
4.2.3 Zinc finger deletions destroy rapid kinetics and induce strong duplex stabilization

4.3 Discussion

4.4 Materials and Methods

5 Paired mutations abolish and restore the balanced annealing and melting activities of ORF1p required for LINE-1 retrotransposition

5.1 Introduction

5.2 Results

5.3 Discussion

5.4 Materials and Methods

5.4.1 Autonomous retrotransposition assay

5.4.2 DNA, RNA and protein analysis

5.4.3 Recombinant ORF1p purification

5.4.4 Circular dichroism and thermal melting

5.4.5 RNA filter binding

5.4.6 Single molecule DNA stretching

5.4.7 Annealing and displacement assays

6 Oligomerization transforms human APOBEC3G from a fast enzyme to a slow binding protein

6.1 Introduction

6.2 Results

6.2.1 Single molecule measurements of A3G binding to ssDNA

6.2.2 Quantitative binding model

6.3 Discussion

6.4 Materials and Methods

6.4.1 A3G preparation and purification

6.4.2 Single molecule experiments

6.4.3 Two-step binding model

6.4.4 Fast on and off rates

6.4.5 Fast equilibrium binding constant

6.4.6 Slow on and off rates

7 Conclusions

Publications

Abbreviations

References
Introduction

The opposite of a correct statement is an incorrect statement. The opposite of a profound truth is another profound truth.

Niels Bohr

1.1 DNA carries the instructions for life

Deoxyribonucleic acid (DNA) encodes the instructions all living organisms need to develop, function, and reproduce. DNA is a long polymer of nucleotides, which consist of a nitrogenous base, the sugar deoxyribose, and a phosphate group (Figure 1.1) [1]. Phosphates alternate with sugar subunits along the negatively-charged DNA backbone, with two complementary nitrogenous bases paired in between, to form a double helix with two anti-parallel strands. The base guanine (G) pairs with cytosine (C), and adenine (A) pairs with thymine (T). The bases form hydrogen bonds, but the base stacking interactions are primarily responsible for the stability of the double helix [2], and G—C base pairs contribute more to helix stability than A—T base pairs. The distance between the base pairs is 0.34 nanometers (nm) in the B-form right-handed double helix (Figure 1.2), which is the predominant conformation of DNA in cells. This helix is 2 nm in diameter and makes one full turn per ten base pairs. There are two asymmetric grooves, a major groove of 2.2 nm and a minor groove of 1.2 nm, in each 3.4 nm turn. The nucleotide sequence along the backbone contains genetic information, which codes for proteins. However, much of the eukaryotic sequence is noncoding, including more than 98% of the human genome [3], and contains chromosomal structural areas, sequences that regulate protein expression, and endogenous retroelements.

Figure 1.1. DNA is a polymer of nucleotides, with bases A (blue), T (red), C (green), and G (purple). The pyrimidine bases (C, T) have one ring and the purine bases (A, G) have two rings. The pyrimidine—purine bases pairs (A—T and C—G) form hydrogen bonds (dashed lines) along the sugar-phosphate (brown and green) backbone.
1.2 DNA replication

All living organisms pass on genetic information through DNA replication. DNA polymerases synthesize a copy of the entire genome from the original DNA molecule with high accuracy and processivity [1]. As the helicase unwinds the double helix (Figure 1.3, yellow), each complementary strand serves as a template for DNA polymerases, which insert the complementary nucleotide opposite each base. The polymerase catalyzes the formation of a phosphodiester bond between the 3’ OH of the deoxyribose and the phosphate of the incoming nucleotide, tethering the base to the DNA backbone. Most DNA polymerases cannot begin de novo, but begin from an RNA primer, which forms a short double-stranded region of approximately twenty base pairs and donates the 3’ OH [1]. Since all DNA polymerases add nucleotides to the 3’ end of the molecule, DNA synthesis always proceeds from the 5’ to 3’ end. The leading strand is therefore synthesized continuously from 5’ to 3’, while the lagging strand is synthesized discontinuously in Okazaki fragments (Figure 1.3, orange), which are short segments that begin with an RNA primer at the 3’ end (brown). Single-stranded DNA binding proteins (SSBs) coat ssDNA exposed at the replication fork. Another DNA polymerase removes the RNA primer and inserts the remaining bases (Figure 1.3, purple), and DNA ligase fills in the nick in the DNA backbone [1]. The replicative DNA polymerase in the Gram-negative bacteria *E. coli*, DNA pol III α, is a model system for investigating DNA replication (Chapter 3).
ssDNA exposed at the replication fork, and another DNA polymerase (light purple) removes the RNA primer (brown) and replaces the RNA with DNA (purple). DNA ligase joins the 3’ and 5’ ends of two adjacent Okazaki fragments.

1.3 Transcript and translation

Genetic information in a DNA molecule that codes for proteins is transcribed into ribonucleic acid (RNA), which is a polymer of nucleotides that consist of a nitrogenous base, the sugar ribose, and a phosphate group [1]. The four bases in RNA are G, C, A, and uracil (U) (Figure 1.4). Although RNA is single-stranded (Figure 1.5), the bases may pair to form thermodynamically stable structures. RNA duplexes and RNA-DNA hybrids form A-helices, which DNA may also form. The A-form double helix is right-handed, with a diameter of 2.3 nm, 0.23 nm between base pairs, and 11 bp in each full 2.4 nm turn.

The DNA double helix unwinds to allow RNA polymerases to synthesize complementary messenger RNA (mRNA) from the DNA template strand, pairing G with C, T with A, and A with U [1]. The mRNA strand is antiparallel to the complementary DNA strand (Figure 1.6, navy), with the same sequence as the coding DNA strand (black). RNA polymerase synthesizes mRNA in the 5’ to 3’ direction inside a 17 bp transcription bubble with an 8 bp RNA-DNA duplex (Figure 1.6). In eukaryotes, cellular proteins cleave the 3’ end of the new mRNA strand, add a 3’ polyadenine tail and a 5’ guanine cap, and remove the noncoding introns and splice the exons together [1]. This processing allows the mRNA to leave the nucleus and protects it from enzymatic degradation in the cytoplasm during translation [6, 7].

Ribosomes use the genetic code to read each three-nucleotide codon along the mRNA, translating it into the corresponding protein (Figure 1.7) [1]. Each mRNA codon has a specific transfer RNA (tRNA) molecule with the anticodon that delivers the appropriate amino acid to the
ribosome, where a peptide bond links it to the nascent polypeptide. Protein synthesis begins with the start codon AUG, which codes for the amino acid methionine, and ends with stop codons UAA, UAG, or UGA [1]. The degenerate genetic code has $4^3 = 64$ codons, which code for all 20 amino acids that cells use to make proteins.

1.4 Reverse transcription

Retroviruses, retrotransposons, and other retroelements have a linear RNA genome that encodes reverse transcriptase (RT), the DNA polymerase that synthesizes dsDNA from an RNA template (Figure 1.8) [8]. In HIV-1 and other long-terminal repeat (LTR) retroviruses and retrotransposons, reverse transcription begins from a tRNA primer that anneals to the primer binding site (PBS) on the viral genome, which provides the free 3′ OH required for initiation of DNA synthesis [8]. RT synthesizes the minus strand of complementary DNA (cDNA) through the LTR region at the 3′ end of the RNA template, and continuation of reverse transcription requires transfer of this minus strand cDNA to the LTR on the opposite end of the RNA genome. After DNA synthesis, the plus strand anneals to the LTR on the opposite end of the minus DNA strand in a second strand transfer [8]. The dsDNA viral genome is then imported into the nucleus, where the viral enzyme integrase inserts it into the host genome.
into the cellular genomic DNA. The cell transcribes viral RNA and expresses viral proteins, which package two strands of viral RNA into the newly assembled viruses that exit the cell to continue the viral life cycle [8]. Human cellular proteins such as A3G may also be packaged alongside genomic RNA virion [9, 10], and interfere with reverse transcription and viral replication [11-13] (Chapter 6).

LTR retrotransposons such as Ty3 also use genomic RNA repeat regions for minus strand transfer during reverse transcription (Chapter 4). In contrast, some non-LTR retrotransposons such as long interspersed nuclear element type 1 (LINE-1, or L1) reverse transcribe directly at the dsDNA insertion site. L1 encodes an endonuclease that cleaves cellular DNA, which anneals to L1 genomic RNA to provide a primer for RT [14]. In target site-primed reverse transcription (TPRT) [14], the initial cDNA strand is elongated from the insertion site. The L1 endonuclease cleaves the opposite cellular DNA strand at a second target site, which serves as primer for second strand DNA synthesis [15] (Chapter 5). Retrotransposons may be the ancestors of retroviruses, and they have adverse effects on their eukaryotic host cells that include insertional mutagenesis and alterations in gene expression and recombination [16].
Scientific principles and laws do not lie on the surface of nature. They are hidden, and must be wrested from nature by an active and elaborate technique of inquiry.  

John Dewey

2.1 Single molecule force spectroscopy

Single molecule methods have provided a clearer understanding of a wide range of fundamental biological processes, including DNA replication, transcription, and repair. Single molecule force spectroscopy began with the capture and manipulation of single DNA molecules. Techniques such as optical tweezers, magnetic tweezers, and atomic force microscopy (AFM) apply forces to single molecules, probing conformational changes and structural dynamics in a variety of conditions. Such measurements explore the interactions of DNA with molecules ranging from small ligands to complex proteins. Quantifying the thermodynamics and kinetics of these interactions leads to substantial insights into DNA binding mechanisms in important biological systems.

Portions of this work were originally published in Physics of Life Reviews 7: 299-341 (2010), and have been adapted for this dissertation.
Optical tweezers, magnetic tweezers, and AFM are the predominant force spectroscopy techniques used to trap and manipulate single DNA molecules. Single-beam optical tweezers instruments focus a high power laser through a high numerical aperture microscope objective to form an optical trap. Dual-beam optical tweezers instruments use microscope objectives to bring two counter-propagating laser beams to an overlapping focus to form an optical trap. The trap captures one streptavidin-coated polystyrene bead, while a second bead is attached to a micropipette tip fixed to a flow cell or is held in another optical trap. A single biotin-labeled DNA molecule is tethered to the beads through a biotin-streptavidin linkage or some other attachment method that can withstand the forces to be applied. Translation of the flow cell or optical trap pulls the bead affixed to the pipette tip or held in the trap, resulting in extension of the captured DNA molecule. This displaces the bead in the optical trap, which provides a measurement of the force on the DNA molecule with piconewton (pN) accuracy [17].

Magnetic tweezers use a glass slide and magnetic bead, both coated with streptavidin or another attachment ligand, in order to capture a biotin-labeled DNA molecule. Translation of the glass slide through a magnetic field gradient results in a force on the DNA molecule, measured as three-dimensional motion of the magnetic bead in video acquisition. Advantages of this technique include single molecule manipulation in three dimensions and detection of forces as low as 0.05 pN [18].

Although AFM is predominantly used in imaging applications, the technique may be used for single molecule force spectroscopy. A single DNA molecule is immobilized between the surface and the AFM tip, and force is measured as a function of extension and relaxation. A typical DNA attachment technique functionalizes opposite ends of the molecule with thiol and biotin. The thiolated end binds covalently to a gold surface, while the streptavidin-coated AFM tip captures the biotin-labeled end of a single DNA molecule [19]. Resolution of the DNA stretching curves is on the order of 5-10 pN [20, 21].

### 2.2 The optical trap: a spring made of light

When light encounters matter, it exerts a force that captures very small particles. Optical tweezers generated by a tightly focused laser beam can trap particles as large as 100 μm and as small as individual atoms [22], and hold them against forces from 0.1 to 300 pN.

The physics of the optical trap depends upon particle size. In the Rayleigh length scale, where particle radius $r$ is significantly smaller than the wavelength of laser light ($r << \lambda$), the particle is effectively a point dipole induced by the electromagnetic field. The Lorentz force on the particle depends upon the gradient of the electromagnetic field $E$ [23, 24]:

$$ F_\text{L} = \pi n_1^2 r^3 \left( \frac{m^2 - 1}{m^2 + 2} \right) \nabla E^2 $$

where $m = n_2/n_1$ is the ratio of the indices of refraction of the particle and the medium. For high-index materials ($m > 1$), the gradient force attracts the particle toward focus of the laser beam. However, Rayleigh scattering counteracts the gradient force along the direction of light propagation. The particle re-radiates some of the energy imparted by scattered photons, but the energy it accumulates results in the scattering force [24]:

$$ F_\text{s} = \left( \frac{2\pi}{\lambda} \right)^4 \frac{8\pi n_1 r^6}{3c} \frac{m^2 - 1}{m^2 + 2} I $$

where $I$ is the intensity of the electromagnetic field. This change in momentum is always forward along the photon path. In general, the scattering force tends to dominate the axial gradient force. However,
sharpening the laser beam focus increases the intensity gradient of the electric field so the backward gradient force effectively counteracts the forward scattering force, forming a stable three-dimensional optical trap slightly behind the laser beam focus.

When particle radius is larger than laser wavelength \( r >> \lambda \), geometric optics describes the optical trap (Figure 2.1). The light rays (Figure 2.1A, green and blue) are symmetric when the bead is at the focus of the laser beam (maroon dashed line). They refract toward the normal (black line) as they enter the bead and away from the normal when they re-enter the medium. The change in momentum on the photons \( dp_1 \) and \( dp_2 \) cancel out in the lateral direction, leading to a net scattering force on the particle \(-F_s\) along the direction of beam propagation.

![Figure 2.1. Radiation pressure on a particle in an optical trap. Light rays shown as solid lines (green and blue), initial momentum shown below for vector addition (dashed lines), and difference in momentum shown as purple and red arrows, respectively. (A) The scattering force on the particle \(-F_s\) is always along the direction of light propagation. (B) The gradient force restores the particle to the focus of the laser beam (maroon star). The lateral component \(-F_{g_l}\) restores the slightly displaced particle (dashed red line) toward the center of the gradient (dark maroon line). (C) The axial component \(-F_{g_a}\) also restores the particle to the center of the beam, and points away from the direction of light propagation when the particle is below the focus. Therefore a tightly focused laser beam generates a gradient force strong enough to cancel out the scattering force in the forward direction, so the particle sits in an optical trap slightly behind the focus.

When the particle is laterally displaced from the center of beam (Figure 2.1B, dashed red line), these rays (green and blue) have different intensities due to the Gaussian gradient of the focused beam. This leads to a greater change in momentum from the light ray closer to the center (green) relative to the one further away (blue), resulting in a net restoring force towards the center of the beam. Therefore the lateral gradient force \( F_{g_l} \) restores the particle to the focus (maroon star, Figure 2.1C). A very tightly focused laser beam generates an axial gradient force that counteracts the scattering force, trapping the particle slightly behind the focus.

Moving a trapped particle away from the laser beam focus by a distance \( x \) results in a force \( F \):

\[
F = -kx
\]

measured by deflection of the laser beam. Therefore the trap is a spring made of light, with stiffness \( k \).
The dual-beam optical tweezers instrument used in these experiments brings collimated beams from continuous wave diode lasers (200 mW, $\lambda = 830$ nm) to an overlapping focus ($d = 1 \mu$m) inside a liquid flow cell (Figure 2.2, red), where the scattering forces cancel out. The tight focus by the microscope objectives (water-immersion, NA of 1.0, working distance of 2 mm) forms an optical trap that captures 5 $\mu$m polystyrene beads. Position-sensitive photodiode detectors measure bead displacement. The beads, micropipette tip, and optical trap are imaged on two CCD cameras with a white light source (Figure 2.2, yellow). Professor Mark Williams originally built this instrument in 1999, and Dr. Fei Wang updated it in 2005. I rebuilt the optical tweezers in collaboration with Dr. Kiran Pant in 2010, and optimized the design, electronics, and data acquisition program under the guidance of Dr. Micah McCauley.

Figure 2.2. Dual-beam optical tweezers instrument. Paths of laser beam 2 (red) and light source 1 (yellow) traced through the optics and into cameras (symmetric paths of beam 1 and light source 2 not shown for clarity). A polarized beam splitting cube deflects most of the laser intensity into the detectors, directing only a small amount into the infrared-sensitive CCD camera (dashed red) for beam visualization that aids in instrument alignment. The flow cell (not shown, see Figure 2.3) sits between the microscope objectives on a piezoelectric stage.
In the experiments presented in this work, this instrument was used to tether a single dsDNA molecule (bacteriophage λ, 48,500 bp) labeled with biotin to two streptavidin-coated polystyrene beads, one held in an optical trap and another on the micropipette tip (Figure 2.3). The fixed bead is moved away from the optically-trapped bead at 100 nm/s, and the force on the DNA molecule is measured at each extension. The bead movement reverses direction after 14 μm, releasing the tension. Generating this DNA force-extension curve takes approximately 2 minutes.

2.3 Stretching single DNA molecules

Single molecule DNA stretching experiments determine force as a function of extension (Figure 2.4). At low extensions, the measured tension increases gradually as the duplex uncoils in what is known as the entropic regime. As extension approaches the dsDNA contour length, the backbone resists further extension and the force increases dramatically in an elastic response. At 65 pN, dsDNA undergoes an overstretching transition, increasing to 1.7 times its contour length at nearly constant force. There is a second transition at the end of this overstretching plateau, near the contour length of ssDNA. If the extension is reduced at this point, the relaxation curve will match the stretching curve. Some hysteresis, where the relaxation curve does not match the stretching curve, may occur, depending upon solution conditions. DNA stretching and relaxation cycles exhibit similar force-extension curves on the timescale of the experiment in typical solution conditions, indicating that the process is reversible.

2.4 Models of polymer elasticity

Polymer models of dsDNA and ssDNA effectively characterize DNA force-extension curves. The Worm-Like Chain (WLC) model assumes a smooth distribution of bending angles for a polymer of length $l$ along contour $s$ (Figure 2.5). The extensible WLC model describes dsDNA in terms of observed length $b_{ds}$ of an elastic polymer under the influence of tension $F$ [26-30]. Though no exact solutions to this model are known, an approximate solution is appropriate for high forces:

$$\langle \mathbf{t}(s_1) \mathbf{t}(s_f) \rangle = e^{-\frac{b_{ds}^2}{P}}$$
where $P_{ds}$ is the persistence length, $B_{ds}$ is the end-to-end contour length, and a stretch modulus $S_{ds}$ is added to account for backbone extensibility. Here $k_B$ is Boltzmann’s constant and $T$ is temperature.

The Freely Jointed Chain (FJC) model describes the polymer elasticity of ssDNA as a collection of independent monomers of length $l$ with varying bond angles $\theta$ (Figure 2.6) [31]:

$$b_{ss}(F) = B_{ss} \left[ \coth \left( \frac{2P_{ss}F}{k_B T} \right) - \frac{1}{2} \frac{k_B T}{P_{ss} F} + \frac{F}{S_{ss}} \right]$$  \hspace{1cm} 2.5$$

The WLC (Figure 2.4, green) and FJC (blue) polymer models are shown with the same generally accepted parameter values ($B_{ds} = 0.34$ nm/bp, $P_{ds} = 48$ nm, and $S_{ds} = 1200$ pN in Equation 2.4, $B_{ss} = 0.55$ nm/bp, $P_{ss} = 0.75$ nm, and $S_{ss} = 720$ pN in Equation 2.5) throughout this dissertation [17, 25].

2.4.1 Force-induced structural transitions

A thermodynamic model quantitatively describes the overstretching transition at 65 pN in terms of force-induced melting. The force exerted on the dsDNA molecule does work to increase the length of the DNA, converting dsDNA to ssDNA and disrupting both base pairing and base stacking interactions. In this model, the second transition at the end of the melting plateau is a non-equilibrium process involving the remaining base pairs of dsDNA which must break for strand separation. Force-induced melting is analogous to thermal melting, and the model predicts that solution conditions which influence thermal melting, such as salt, pH, and temperature, also affect the force-induced melting transition. DNA stretching experiments quantitatively confirmed these predictions [30, 32, 33], and recent modeling studies also support a force-induced melting model [34-36]. Furthermore, experiments demonstrated that solution conditions [37] and DNA binding ligands [38-45] known to inhibit DNA reannealing induce strong hysteresis in the force-extension curves, providing additional evidence for melting of the DNA strands.

In an alternate model of the overstretching plateau, B-form duplex DNA lengthens in response to the applied force, undergoing a structural transition to a new form of DNA referred to as “S-DNA” [20, 46-48]. This form of DNA is predicted to preserve base pairing but not base stacking, a distinction based upon the observation that strand separation occurs at high forces [49-51]. An early modeling study predicted a transition to this form of DNA at a significantly larger force than experiments observed. Recent studies use the proposed existence of S-DNA as a means to generate new parameters to fit stretching curves and other experimental results. However, it is not clear that additional fitting parameters are needed to explain DNA stretching experiments. In addition, these models do not make predictions that can be tested with other experiments, making it difficult to test proposed S-DNA models [52-54]. Magnetic tweezers experiments with both strands of a dsDNA molecule tethered to beads did not observe the transition at 65 pN, but instead measured a transition at 110 pN over a similar extension attributed to a combination of S-DNA and P-DNA, a form of melted DNA which is overwound [55]. Although this particular transition of torsionally constrained DNA is consistent with force-induced melting, it was suggested that some features of DNA stretching curves were incompatible with force-
induced melting [56, 57] and that the structure of DNA in the overstretching transition remains unclear [58, 59].

It is essential to establish the nature of the overstretching transition in order to use single molecule force spectroscopy techniques to characterize DNA binding. Recent experiments use glyoxal, intercalating dyes, and SSBs to establish that this conformational transition involves base pair disruption, and therefore DNA overstretching is force-induced melting of dsDNA into ssDNA.

2.4.2 Glyoxal binds ssDNA bases exposed in the force-induced melting transition

Glyoxal (C$_2$H$_2$O$_2$) is a small molecule which binds irreversibly to exposed guanine bases of DNA with slow kinetics [60]. The modified guanine bases have three rings instead of two, introducing steric constraints that hinder base pair reannealing [61]. λ-DNA molecules were held at fixed extensions for 30 min in the presence of glyoxal, which is the timescale required for DNA binding [60]. The DNA stretching curve exhibits a decrease in melting force and strong hysteresis (Figure 2.7A), indicating that guanine bases exposed to solution are subject to glyoxal modification and subsequently prevent DNA reannealing [60]. Therefore extension into the overstretching plateau exposes ssDNA bases to solution, reflecting force-induced melting of the dsDNA molecule.

As the DNA molecule is held at larger fixed extensions, the corresponding relaxation curves exhibit additional hysteresis (Figure 2.7B). These results demonstrate that glyoxal binding increases as the DNA molecules are held further into the overstretching plateau, despite constant solution conditions. This indicates that greater extensions into the stretching transition result in exposure of additional bases, and the relaxation curves in the presence of glyoxal are a combination of dsDNA and ssDNA. The experimental data fits well to a linear combination of the FJC and WLC polymer models (fits shown in Figure 2.7B), where the measured contour length $b$ is a function of the ssDNA fraction $f_{ss}$ [60]:

$$b(f_{ss}) = b_{ds}(1-f_{ss}) + b_{ss}f_{ss}$$

and $b_{ds}$ and $b_{ss}$ are force-dependent DNA extensions from the WLC (Equation 2.4) and the FJC model (Equation 2.5), respectively. The fractional extension along the transition plateau agrees well with the fraction of glyoxal-stabilized ssDNA obtained from fits to Equation 2.6, which provides structural evidence that DNA overstretching is indeed the force-induced melting of dsDNA into ssDNA.

Figure 2.7. Glyoxal binds ssDNA bases exposed in the force-induced melting transition. (A) Extension (solid line) and relaxation
(dashed line) data of a λ-DNA molecule alone is shown in black. After the addition of 500 mM glyoxal, the molecule is extended (solid green line) and held fixed (dashed green arrow) for 30 min. The significant hysteresis upon relaxation (dashed green line) reflects that the two DNA strands do not reanneal, indicating glyoxal binding to exposed nucleotides. The second stretch (solid blue line) follows the previous relaxation curve, which suggests that modification is permanent. (B) Relaxation data (open circles) for a series of fixed extensions (dashed arrows), in which the DNA molecule is stretched in the presence of 500 nM glyoxal. Fits to a linear combination of the WLC and FJC models (Equation 2.6) are shown as solid lines. Figures adapted with permission from [17].

2.4.3 Visualizing force-induced melting with intercalators and SSBs

The significant presence of ssDNA exposed to glyoxal modification in the overstretching transition is unlikely to arise from nicks in the DNA backbone [62], and further experiments with small molecules and SSBs confirm the force-induced melting model. Recent single molecule studies have directly visualized the nature of the structural transition in a combination of optical tweezers and fluorescence imaging techniques [63]. A DNA molecule stretched to a fixed extension in the absence of ligand is briefly transferred into the presence of YOYO, a fluorescent dye which intercalates into the paired bases of dsDNA [64]. Subsequent imaging reveals only regions of dsDNA, to which the intercalator can bind [63]. The fraction of dsDNA present at each fixed extension corresponds directly to fractional extension along transition plateau, illustrating a structural conversion from dsDNA into a form of DNA to which YOYO is unable to bind [63].

Experiments with fluorescent dye-labeled SSBs demonstrate that the form to which dsDNA is converted upon overstretching is ssDNA. Human mitochondrial SSB (mtSSB) binds and wraps relaxed ssDNA [65], but does not affect the overstretching transition or bind ssDNA that is under tension greater than 40 pN. When DNA is extended into the stretching transition and briefly placed in the presence of mtSSB, the images show fluorescent spots at both ends of the DNA molecule, indicating the presence of protein-wrapped ssDNA [63]. These spots increase in brightness and move toward the center of the molecule as a function of extension, illustrating the relative increase of ssDNA with progressive movement into the stretching transition. This method also visualizes nicks in the DNA backbone, since the mtSSB wraps the relaxed ssDNA in the middle of the molecule. Molecules without nicks do not exhibit these binding events, and mtSSB fluorescence is confined to the ends of the DNA. Two-color fluorescent measurements with both YOYO and mtSSB confirm that mtSSB-wrapped ssDNA forms at an interface with YOYO-labeled dsDNA [63].

In contrast with mtSSB, the SSB Replication Protein A (RPA) binds ssDNA under tension of at least 70 pN. RPA is able to bind both ssDNA under tension and relaxed ssDNA without wrapping it [66]. Two-color fluorescence measurements with eGFP-labeled RPA and bis-intercalator POPO-3 show three fluorescent regions [63]. The dsDNA segment has two bright spots of relaxed ssDNA on either side, followed by two ssDNA strands extending out to their respective attachment sites on each bead. Application of flow perpendicular to the axis of the molecule stretches out the relaxed ssDNA, clearly illustrating both strands of ssDNA created upon dsDNA overstretching [63].

Similar experiments with a DNA molecule attached to beads on both strands reveal the torsionally-constrained transition at 110 pN, with sites of POPO-3-labeled dsDNA and RPA-labeled ssDNA throughout the molecule [63]. The negative correlation of dsDNA and ssDNA areas on the same DNA molecule implies spatial separation of melted regions, with no evidence to support an interpretation of separate S-DNA and P-DNA phases [67]. The data also indicate that short regions of dsDNA remain when DNA is stretched to forces beyond the overstretching transition (in the second transition at the end of the overstretching plateau, near contour length of ssDNA). Thus, complete separation of the strands require application of unexpectedly high forces, but most of the DNA has been melted by force during
overstretching [63]. Although this pulling-rate dependent transition [20, 68] is not well-described, it exists even in the presence of ssDNA binding ligands, which is unexpected in the S-DNA model [67].

The results of these single molecule fluorescence imaging experiments are consistent with formation of ssDNA during both structural transitions, an observation which is incompatible with the prediction of unexposed individual bases of the S-DNA model. Thus the overstretching transition is a force-induced melting transition, in which the applied force does work to melt dsDNA into ssDNA. Therefore DNA stretching experiments involve melting of the two strands. This result can be used as a basis for investigation of the biophysical mechanisms of DNA-protein interactions with single molecule force spectroscopy techniques.

2.5 Biophysical characterization of DNA binding proteins

We use optical tweezers to quantitatively characterize the DNA binding mechanisms of proteins involved in replication in biological systems ranging from *E. coli* to human cells. The replicative DNA polymerase of *E. coli* is a model system for studying DNA replication, and we use single molecule biophysics to characterize its dsDNA and ssDNA binding properties. Genetic rearrangements during DNA replication involve homologous recombination, in which nucleic acid strand exchange occurs between similar molecules. This is the most frequent type of recombination for retroviruses and retrotransposons [69], which carry two genomic RNA molecules [70-72], in eukaryotic host cells. Retroviral recombination occurs predominantly between these two RNA strands during strand transfer events [69], which require nucleic acid rearrangements. Thus DNA replication requires proteins such as nucleic acid chaperones and DNA polymerases, which bind both single- and double-stranded DNA in order to facilitate these processes. A comparable affinity for both forms of DNA results in complex binding properties, and single molecule force spectroscopy methods explore the DNA binding mechanisms of these proteins [17, 25].

2.5.1 The replicative DNA polymerase of *E. coli*

Chromosomal DNA molecules are millions of base pairs in length, and the vast amount of information they carry must be replicated with high accuracy in order to sustain life. The enzymes which are largely responsible for this are known as DNA polymerases, and they are intricate complexes with multiple subunits [73-76]. The polymerase subunits of these holoenzymes synthesize the new complementary DNA strands during DNA replication [73-76]. The α subunit is the 1160-residue DNA polymerase of *E. coli* DNA pol III, a complex 10-subunit asymmetric dimer which coordinates simultaneous leading and lagging strand synthesis at the replication fork [73-76]. Although a crystal structure of residues 1-917 demonstrates that α subunit folds into the right hand shape characteristic of DNA polymerases [77], the absence of significant sequence similarities between pol III α and well-known DNA polymerases leaves DNA binding domains unclear [78]. The partial crystal structure reveals a dual helix-hairpin-helix motif (HhH)_2, at residues 833-889 which is the predicted dsDNA binding domain [79, 80]. Although the *E. coli* pol III α crystal structure does not include the C-terminal residues 917-1160, sequence homology modeling suggests an OB fold at residues 978-1078, which is the ssDNA binding site [77, 81-85]. The full-length crystal structure of *Thermus aquaticus* pol III reveals similarities to *E. coli* Pol III [81], and a recent structure of the *T. aquaticus* α subunit shows the OB fold bound to DNA [86]. Single molecule DNA stretching experiments characterized the dsDNA and ssDNA binding activity of *E. coli* pol III α, quantifying binding to both forms of DNA by distinct regions of the protein [87].

DNA stretching curves in the presence of full-length α (Figure 2.8A) exhibit significant hysteresis, indicating that α remains bound to the single DNA strands, such that the DNA molecule cannot reanneal on the timescale of the relaxation cycle. Although the initial stretch with α follows the stretching curve of DNA without protein up to the melting plateau, subsequent stretches reflect that some protein
remains bound, stabilizing a fraction of ssDNA over the length of the molecule. Thus all of the bound protein does not dissociate upon complete relaxation, and the DNA molecule reaches a saturated combination of roughly half ssDNA and half dsDNA [87]. Pol III α demonstrates a strong preference for ssDNA without decreasing the DNA melting force. This indicates that the protein does not destabilize existing dsDNA, but binds ssDNA available due to force-stabilized fluctuations. Furthermore, DNA stretching experiments reveal that pol III α stabilizes the DNA duplex, increasing the DNA melting force in a salt-independent manner [87]. This is due to protein binding to dsDNA, an interaction which must be disrupted prior to force-induced melting of the duplex.

DNA stretching curves in the presence of the 917-residue N-terminal fragment (α1-917) also exhibit a higher melting force, but no hysteresis. This indicates that α1-917 binds dsDNA, stabilizing the duplex, but does not bind ssDNA. Thus the 917 N-terminal residues do not contain the ssDNA binding site, but do contain the dsDNA binding site [87]. Single molecule experiments with residues 1-835 (α1-835) detected a small change in melting force, indicating that the fragment has decreased affinity for dsDNA. Therefore, the dsDNA binding site lies predominantly within residues 836-917, a region of pol III α which includes the (HhH)2 domain at residues 833-889.

DNA stretching curves with the C-terminal fragment α917-1160 exhibit the type of hysteresis evident with full-length α, but lack the higher melting force. Similar features are evident with a fragment of residues 978-1160 (α917-1160), which indicates that both proteins bind ssDNA but not dsDNA [87]. In contrast, stretching experiments with α1076-1160, a fragment which does not include the predicted OB fold at residues 978-1078, do not show any evidence of DNA binding activity [87]. These results are consistent with an ssDNA binding site at the OB fold.
Equilibrium binding constants for the full-length α protein and the three fragments α1-917, α1-835, and α917-1160 were obtained for both dsDNA and ssDNA [87]. Measurements of the observed melting force $F_m$ as a function of protein concentration determine fractional binding to dsDNA:

$$\theta = \frac{F_m(c) - F_m^0}{F_m^{sat} - F_m^0}$$

where $F_m^0$ is the melting force of DNA alone and $F_m^{sat}$ is the force observed upon full protein saturation [88, 89]. The site exclusion binding isotherm of McGhee and von Hippel relates the fractional occupancy $\theta$ to the protein concentration $c$, for a protein with an equilibrium association binding constant $K_a$ and binding site size $n$ [90, 91]:

$$\theta = K_a c \frac{(1-\theta)^n}{\left(1 - \frac{\theta}{n}\right)^{n-1}}$$

Fits to the McGhee-von Hippel binding isotherm yield the dsDNA equilibrium binding constant $K_{ds}$ for the four proteins. Furthermore, protein binding to ssDNA stabilizes a fraction of ssDNA, and fits of this relaxation data to a linear combination of the contour lengths of dsDNA $b_{ds}$ and ssDNA $b_{ss}$ yield the force-dependent contour length $b$, which is proportional to the fraction of protein-stabilized ssDNA:

$$f_{ss} = \frac{b - b_{ds}}{b_{ss} - b_{ds}}$$

Fits of $f_{ss}$ to the McGhee-von Hippel binding isotherm (Equation 2.8, where $\theta$ is $f_{ss}$) determine $K_{ss}$ for each protein [87]. The equilibrium binding constants determined (Figure 2.8B-C) indicate that full-length α has strong binding affinity for both single- and double-stranded DNA. Both the α1-917 (blue) and α1-835 (red) fragments show dsDNA binding, but no affinity for ssDNA. In contrast, α917-1160 (purple) binds ssDNA but not dsDNA. Figures adapted with permission from [17].

2.5.2 Retroviral nucleocapsid proteins
Retroviruses and retrotransposons store genetic information in a linear RNA strand, and their DNA polymerase RT converts the RNA template into dsDNA which will be integrated into cellular genomic DNA. Reverse transcription includes multiple stages which involve rearrangement of nucleic acid secondary structure. These processes require nucleic acid chaperone proteins, which aggregate nucleic acids, destabilize base pairing, and promote complementary strand annealing. The chaperone activity of these proteins requires strong binding affinity for both single- and double-stranded nucleic acids, a property that is not yet well-understood.

The nucleic acid chaperones in retroviruses are NC proteins, which tend to be small cationic proteins with minimal structure other than zinc-stabilized CCHC finger motifs [92-95]. Retroviral reverse transcription requires NC for primer annealing, strand transfer, and other instances of secondary structure rearrangement [69]. NC facilitates the annealing of a tRNA primer to the PBS on a positive strand of viral genomic RNA [96-98], a process which involves destabilization of RNA base pairing. RT then polymerizes the minus strand of complementary DNA as its RNaseH domain degrades the RNA template, proceeding through the repeat region at the end of the genomic RNA strand. Reverse transcription of the remaining viral RNA requires minus strand transfer, in which the newly synthesized cDNA must anneal to the LTR on the opposite end of the RNA genome. However, secondary structures within the cDNA and the LTR RNA are thermodynamically stable, and thus the energy barrier to strand annealing makes duplex formation unlikely. NC destabilizes these secondary structures and facilitates annealing of the complementary DNA-RNA duplex [99-101]. A second strand transfer, where the plus DNA strand anneals to the LTR on the opposite end of the minus DNA strand, is also inefficient without NC [99-101]. The nucleic acid chaperone activity of NC may also be essential for genomic RNA dimerization during retroviral maturation [102-104].
In addition to its nucleic acid chaperone properties, NC also participates in genomic RNA packaging for retroviral assembly, during which it is a domain of the Gag polyprotein [70-72], and viral dsDNA integration [106, 107]. Multiple stages of the HIV-1 life cycle require NC, and antiviral drugs targeted its zinc fingers [89, 108, 109] until non-specific toxicity concerns became clear [110]. The protein has little known structure other than these zinc fingers, and thus further drug targeting requires a detailed understanding of its function. Single molecule DNA stretching experiments have characterized the nucleic acid chaperone activity of several retroviral NC proteins [88, 111-113].

HIV-1 NC is a 55-residue protein with two zinc fingers and a basic N-terminal tail [114-116]. Single molecule DNA stretching experiments reveal that the nucleic acid chaperone aggregates DNA, destabilizes dsDNA, and facilitates duplex reannealing [88, 105, 111, 117]. The presence of 10 nM HIV-1 NC reduces the free energy of DNA melting from approximately $2k_BT$ (Figure 2.9A) to approximately $1k_BT$ (Figure 2.9B) [105, 117], destabilizing the DNA duplex. The increase in the width of the melting transition ($\Delta F_m$) also suggests weak duplex destabilization, and positively correlates with chaperone activity [37, 105, 111]. The DNA stretching and relaxation curves in the presence of HIV-1 NC (Figure 2.9C) do not exhibit the strong hysteresis that is characteristic of SSBs, which bind ssDNA and dissociate slowly, such that the DNA molecule cannot reanneal on the timescale of the DNA release cycle [45]. In contrast, HIV-1 NC (or NCP7) rapidly dissociates from ssDNA, allowing the two strands to reanneal and form the DNA duplex [88, 111]. However, the NC precursor products Gag, NCp15, and NCp9 which have less chaperone activity than NCp7 induce more hysteresis, implying slower DNA binding kinetics [111]. Rapid kinetics of nucleic acid binding is an essential component of HIV-1 NC chaperone activity, and it requires a delicate balance between ssDNA binding and dissociation [88].

**Figure 2.9.** HIV-1 NC lowers the free energy of DNA melting $\Delta G$. (A) Force-extension curve of dsDNA (solid line), and the FJC model (Equation 2.5) for ssDNA stretching (dashed). (B) Stretching curves of dsDNA (solid line) and ssDNA (dashed line, [105]) in the presence of 7 nM HIV-1 NC. (C) Stretch (solid) and release (dashed) curves of dsDNA in the absence of protein (black) and in the presence of 10 nM HIV-1 NC (blue).
The nucleic acid interactions of HIV-1 NC involve aromatic residues in the zinc fingers which stack with ssDNA bases [118, 119]. Interchanging the order of the zinc fingers forms mutant 2-1, which has significantly lower DNA binding affinity and lacks the rapid kinetics of wild-type HIV-1 NC [88]. In mutant 1-1, a copy of the first zinc finger replaces the second zinc finger, also leading to lower binding affinity and slower interaction kinetics than HIV-1 NC [88]. However, the DNA binding and dissociation kinetics of this mutant are the least compromised among those studied, and it retains some ability to facilitate minus strand transfer [120]. Thus rapid kinetics relies on specific zinc finger architecture, and both zinc fingers are crucial for the chaperone function of HIV-1 NC.

Single molecule stretching experiments have also characterized DNA binding of NC proteins from other retroviruses, elucidating the role of rapid kinetics in nucleic acid chaperone activity. NC proteins from Moloney murine leukemia virus (MuLV), Rous sarcoma virus (RSV), and human T-cell lymphotropic virus type 1 (HTLV-1) differ predominantly in ssDNA binding affinities [112]. DNA stretching curves in the presence of HIV-1 NC (Figure 2.9C) and RSV NC both exhibit minimal hysteresis, which indicates rapid DNA binding kinetics for both proteins [112]. The clearest disparity between them is the effect on the DNA melting force. HIV-1 NC lowers the melting force, reflecting duplex destabilization, while RSV NC increases the melting force, suggesting duplex stabilization. This duplex stabilization activity of RSV NC is unique among the retroviral NC proteins examined. MuLV NC and HTLV-1 NC (Figure 2.10A) both lower the DNA melting force, the magnitude of which implies moderate duplex destabilization activity [112]. However, DNA stretching curves in the presence of these two NC proteins exhibit significant hysteresis, reflecting their relatively slow dissociation from ssDNA. Furthermore, HTLV-1 is unable to aggregate nucleic acids, while the remaining retroviral NC proteins have similar strong aggregation abilities [112].

Figure 2.10. The CTD of HTLV-1 NC regulates DNA dissociation kinetics. (A) DNA extension (solid) and release (dashed) curves in the absence of protein (black) and in the presence of 700 nM wild-type HTLV-1 NC (green) and (B) 200 nM ΔC29 HTLV-1 NC. (C) Force as a function of time at fixed extension during DNA release (open circles) in the presence of 700 nM HTLV NC wild-type (green) and 200 nM ΔC29 (blue, data magnified in inset), fit to single exponentials (black). Figures adapted from [113].
These results reveal that nucleic acid chaperone activity correlates with rapid DNA interaction kinetics across a variety of retroviruses. HIV-1 NC and RSV NC are the most effective chaperones, MuLV NC has moderate chaperone ability, and HTLV-1 is an extremely poor chaperone. This comparison suggests that nucleic acid chaperone activity may be related to thermodynamic stability of secondary structures formed in the repeat regions of the retroviral genomic RNA [112]. Only the complex secondary structure of the HTLV-1 LTR region is inconsistent with this observation [112], which suggests that the extremely slow dissociation rate of HTLV-1 NC may serve a specific biological function.

Figure 2.10 shows representative DNA stretching curves in the presence of (A) HTLV-1 NC and (B) ΔC29 HTLV-1 NC, a mutant lacking 26 residues of the acidic C-terminal domain (CTD). The significant hysteresis observed with the full-length protein is characteristic of SSBs [38-42, 44, 45], while the lack of hysteresis with the C-terminal truncate resembles the behavior of nucleic acid chaperones such as HIV-1 NC (Figure 2.9C) and RSV NC. Time-dependent force measurements, in which the force is monitored as a function of time at fixed extension, reveal that wild-type HTLV-1 NC lowers the equilibrium melting force $F_m$ more than its CTD deletion mutant (Figure 2.10C). When the DNA molecule is held at the midpoint during release, the force measured increases until it converges to $F_m$, which is consistent with protein dissociation from ssDNA. Equilibrium constants for DNA reannealing in the presence of the protein obtained from these force relaxation measurements indicate that wild-type HTLV-1 NC dissociates from ssDNA an order of magnitude more slowly than its C-terminal truncate [113].

These results are consistent with a model in which the CTD regulates the chaperone activity of HTLV-1 NC. The strongly salt-dependent nature of the interaction between the anionic CTD and the cationic N-terminal domain (NTD) [113] indicates that the CTD occupies the DNA binding site in low salt, reducing the DNA association rate of wild-type HTLV-1 NC [113]. Lack of a CTD in the mutant leaves the NTD available for binding ssDNA and contributes to the on rate, an electrostatic mechanism of DNA binding regulation by the CTD that resembles the mechanism used by bacteriophage SSBs [45]. Additionally, the off rate of ΔC29 HTLV-1 NC is significantly faster than that of the wild-type protein [113]. This is consistent with a model in which HTLV-1 NC requires the CTD to bind ssDNA cooperatively [113]. The acidic CTD of each bound monomer interacts with the basic NTD of the adjacently bound monomer, and this cooperative interaction reduces the rate of dissociation from ssDNA [113]. However, the ΔC29 HTLV-1 NC mutant has faster on and off rates than wild-type HIV-1 NC, which results in rapid kinetics and thus confers efficient nucleic acid chaperone activity.

Cooperative binding to ssDNA is unexpected for a retroviral nucleocapsid protein such as that of HTLV-1, particularly since the resulting low off rate compromises its nucleic acid chaperone function. A biological function of more importance may require slow dissociation from ssDNA, superseding the need for efficient chaperone activity. Although that biological function remains unknown, one possibility is exclusion of human A3G proteins, for which HTLV-1 NC does not possess a degradation mechanism [121]. The rapid kinetics of HIV-1 NC allow A3G to bind ssRNA [122] and be co-packaged into HIV-1 virions [11, 13, 123], which inhibits viral replication in the absence of the HIV-1 protein Vif. In contrast, HTLV-1 virions have few A3G proteins [124, 125], a reduction which requires the CTD of HTLV-1 NC [121]. It is possible that HTLV-1 NC binds ssRNA cooperatively to exclude binding of APOBEC3G [113]. Thus the slow ssDNA dissociation kinetics that reduce HTLV-1 NC chaperone activity may be responsible for reduced packaging of A3G proteins [113]. We investigate the biophysical mechanisms by which A3G inhibits viral replication in Chapter 6.

2.5.3 The ORF1 protein from the retrotransposon LINE-1

Many retroviruses and retrotransposons, such as Ty3, also use LTR regions of genomic RNA for minus strand transfer during reverse transcription. We investigate the chaperone activity of Ty3 NC, which only
has a single zinc finger, in Chapter 4. Mobile genetic elements which also reproduce via an RNA intermediate but lack this repeat region are called non-LTR retrotransposons. Long interspersed nuclear element type 1 (LINE-1, or L1) is a non-LTR retrotransposon abundant in the human genome. L1 encodes an endonuclease, which cleaves cellular genomic DNA at the site of insertion. This DNA strand anneals to L1 genomic RNA and becomes the primer for RT, a process called target site-primed reverse transcription (TPRT) [14]. The L1 nucleic acid chaperone, open reading frame 1 protein (ORF1p), is essential for TPRT-based retrotransposition, and it may destabilize the dsDNA duplex at the target site and facilitate annealing of the DNA-RNA hybrid [15]. The first cDNA strand transcribed from the RNA template is already joined to cellular DNA cleaved at the target site. The L1 endonuclease also cleaves the opposite strand of cellular dsDNA at the second target site, which anneals to the end of the first cDNA strand and serves as a primer for polymerization of the second cDNA strand [15]. This second strand exchange may also involve the nucleic acid chaperone activity of ORF1p, which may destabilize genomic dsDNA and stabilize the DNA-cDNA duplex [15].

The ORF1 protein from mouse L1 is 357 residues [126], which is significantly larger than NC proteins in LTR retroviruses and retrotransposons. L1 ORF1p also lacks the CCHC domain characteristic of NC proteins and ORF1 proteins from all other non-LTR retrotransposons [127]. The basic CTD contains the nucleic acid binding site, while the acidic NTD contains a 120-residue coiled-coil domain [128], which is required for the protein–protein interactions [126] that allow mouse L1 ORF1p to form a stable trimer [128]. Two arginine (R) residues in the basic CTD, at positions 297-298, are integral to nucleic acid binding and chaperone activity [129]. Single molecule DNA stretching experiments reveal that wild-type ORF1p (RR) increases the transition width, or the change in force from the beginning to end of the DNA melting plateau, moderately relative to that for HIV-1 NC [129]. The transition width $\Delta F_m$ reflects the cooperativity of the melting transition, and an increase in $\Delta F_m$ correlates positively with nucleic acid chaperone activity [37, 105, 111, 117]. The wild-type RR protein also decreases the extension at which the ssDNA transition occurs, reflecting ssDNA aggregation [129]. That is, the protein induces effects that make ssDNA attracted to itself, decreasing the effective length of ssDNA at a particular force. Therefore moderate ssDNA aggregation may contribute to nucleic acid chaperone activity.

Substituting alanine (A) for both arginine residues results in a mutant (AA) with minimal RNA binding affinity and no chaperone activity [129]. DNA stretching curves in the presence of the AA mutant reflect a significant decrease in DNA binding, minimal ssDNA aggregation, and almost no impact on the transition width $\Delta F_m$ [129]. However, replacing one or both of the arginine residues with lysine (K) conserves charge, resulting in mutants (RK, KR, and KK) with nucleic acid binding affinity similar to that of wild-type RR. The RK mutant shows a decrease in nucleic acid chaperone activity, and facilitates L1 retrotransposition less efficiently than wild-type ORF1p. In contrast, the KR and KK mutants also show diminished nucleic acid chaperone activity and are unable to facilitate retrotransposition [129]. DNA stretching studies show that all three mutants increase $\Delta F_m$ and aggregate ssDNA, reflecting some chaperone ability [129]. However, the mutants also significantly aggregate dsDNA, and KR and KK in particular form extremely stable protein-dsDNA aggregates at relatively low concentrations [129]. These results illustrate that mutations in ORF1p that result in severe nucleic acid aggregation inhibit chaperone activity.

The amino acid at position 159 of ORF1p is responsible for a 15-fold difference in retrotransposition efficiency in two variants of mouse L1, $T_{FC}$ and $T_{spa}$ [130]. $T_{FC}$ ORF1p contains aspartic acid (D159), while the ORF1p of $T_{spa}$ contains histidine (H159). Single molecule methods and bulk assays establish that D159 and H159 ORF1p have similar nucleic acid binding affinities [130]. However, D159 is a more efficient nucleic acid chaperone than H159 [130]. DNA stretching curves in the presence of D159 have a larger increase in transition width $\Delta F_m$ than those with H159 [130], which reflects greater chaperone
activity [37, 105, 111, 117]. Although both variants of ORF1p induce similar hysteresis, indicating preferential binding to ssDNA, D159 aggregates ssDNA significantly less than H159. Smaller ssDNA aggregation, and thus faster dissociation from ssDNA, is consistent with bulk experiments which indicate D159 has more rapid nucleic acid binding kinetics than H159 [130]. This implies that ORF1 proteins possess optimal ssDNA aggregation properties to effectively direct nucleic acid rearrangements for L1 retrotransposition. Thus mutations that decrease ΔF_m or induce strong DNA aggregation interfere with the precise balance of nucleic acid aggregation, duplex destabilization, and strand reannealing, which is characteristic of efficient nucleic acid chaperone activity. We further examine this correlation between ORF1p chaperone activity and L1 retrotransposition in Chapter 5.
Polymerase manager protein UmuD regulates *E. coli* DNA polymerase III α binding to ssDNA

**Abstract**

Replication by *E. coli* DNA pol III is disrupted upon encountering DNA damage. Consequently, specialized Y-family DNA polymerases that are able to bypass the damaged DNA are employed. Although the mechanism for exchanging DNA pol III for a damage-bypass polymerase is not understood, the protein UmuD is extensively involved in modulating cellular responses to DNA damage and may play a role in DNA polymerase exchange. UmuD interacts with the α subunit of DNA pol III at two distinct binding sites, one of which is located adjacent to the single-stranded DNA binding site of α. Here, we have used single molecule DNA stretching experiments to demonstrate that UmuD inhibits binding of α to ssDNA, and that UmuD residues D91 and G92 are involved in this function. This suggests that preventing α from binding its ssDNA template is one mechanism by which UmuD inhibits DNA replication and facilitates polymerase exchange.

This work was submitted for consideration for publication, and has been adapted for this dissertation. Ensemble and molecular modeling results included in this chapter are the work of Michelle Silva, Philip Nevin, Lukas Voortman, and Penny Beuning at the Northeastern University Department of Chemistry and Chemical Biology.
3.1 Introduction

The DNA pol III holoenzyme is a ten-subunit protein complex that efficiently and accurately replicates the entire genome of *E. coli* [74, 75]. It is composed of three subassemblies: the polymerase core, the β processivity clamp, and the clamp loader complex. Polymerase and proofreading activities are conducted by the core subassembly which consists of the polymerase subunit α, the proofreading subunit ε, and the θ subunit, which has a role in stabilizing the core [131, 132]. The β processivity clamp encircles the DNA and provides a platform for the polymerase core to bind, providing α with access to the primer-template and facilitating processive replication. The clamp loader complex, which consists of the γ, δ, δ′, τ, χ, and ψ subunits, loads the β clamp onto the DNA [133] with τ tethering the polymerase core to the replisome [134], and coordinating simultaneous replication of the leading and lagging strands of the replication fork [134, 135].

Although DNA pol III efficiently replicates undamaged DNA, replication is disrupted upon encountering damaged bases [136-139]. Formation of a RecA filament on accumulated ssDNA triggers the SOS response [140], resulting in the upregulation of genes encoding numerous proteins involved in DNA damage repair and tolerance [141]. These proteins include the potentially mutagenic Y-family polymerases DNA pol IV (DinB) and DNA pol V (UmuD′,C) [142-144]. Replication of damaged DNA can proceed once DNA pol III α is replaced with one of these Y-family polymerases, which can replicate damaged DNA in a process known as translesion synthesis [145-148]. DNA pol V is composed of two subunits, the cleaved form of UmuD and the UmuC polymerase. DNA polymerase manager protein UmuD regulates the cellular response to DNA damage in part, along with UmuC, by decreasing the rate of replication, thereby allowing time for non-mutagenic DNA repair processes to occur [149-151]. UmuD undergoes a RecA/ssDNA-facilitated auto-cleavage of 24 amino acids of its N-terminal “arms” to form UmuD′. UmuD forms a tight dimer (UmuD2), which is the predominant form for the first 20–40 minutes of the SOS response after which the cleaved form UmuD′ is the predominant species [150, 152]. Although UmuD and UmuD′ are expected to be dimeric under all the conditions studied here [153] for simplicity we will refer to these dimeric forms as UmuD rather than UmuD2 and UmuD′ rather than UmuD′2, respectively. UmuD also regulates mutagenesis in the cell through its interaction with the Y-family DNA polymerase DinB, by inhibiting DinB-dependent -1 frameshift mutagenesis [142, 154, 155].

UmuD interacts with several components of DNA pol III, including the polymerase subunit α, the β clamp, and the proofreading subunit ε [156]. Recent ensemble biochemical experiments have shown that there are two UmuD binding sites on the α subunit, one in the N-terminal domain and one in the C-terminal domain [157] (Figure 3.1). The C-terminal binding site (residues 956-975), which has higher affinity for full-length UmuD relative to the cleaved form UmuD′ [157], is adjacent to the β clamp binding site (residues 920-924) [158], which tethers the polymerase to its DNA template. UmuD, but not UmuD′, releases α from the β clamp, which may inhibit DNA replication and facilitate polymerase exchange [157]. The C-terminal UmuD binding site of α is also adjacent to the OB fold (residues 975-1160), through which α binds ssDNA [87], suggesting that UmuD may be competing with ssDNA for binding to α. We hypothesized that one way UmuD contributes to a DNA damage checkpoint is by disrupting the interaction between α and ssDNA, thereby inhibiting replication. To test this hypothesis, we have used single molecule DNA stretching to quantify α binding to ssDNA in the presence of wild-type UmuD and several UmuD variants designed from a computational docking analysis of the complex. We find that wild-type UmuD competitively inhibits α binding to ssDNA through UmuD-α interactions, while a single amino acid substitution, D91K, in UmuD disrupts this inhibition.
3.2 Results

3.2.1 UmuD inhibits α binding to ssDNA

In DNA stretching experiments, a single double-stranded λ-DNA molecule was captured between two polystyrene beads, one held in an optical trap and the other fixed on a micropipette tip. As the distance between the beads increases, measurements of the force on the DNA molecule yield a force-extension curve (Figure 3.2A, solid black line). At a constant force of 62.6 ± 0.5 pN, the dsDNA molecule undergoes a force-induced melting transition to ssDNA. This overstretching transition has been established as force-induced melting in the presence of DNA binding proteins such as α, which is the case for the experiments presented in this work [159]. When the tension on the DNA molecule is released (Figure 3.2A, dashed black line), the ssDNA generated by force anneals immediately into dsDNA, and the curve exhibits minimal hysteresis, or mismatch between DNA extension and release curves.

![Figure 3.1. Diagram of DNA pol III α, with domain labels within the boxes and known interaction sites above the boxes (sequence numbering shown below). The two UmuD binding sites, one in the N-terminal domain and one in the C-terminal domain of α, are shown in yellow [157]. The CTD binding site (residues 956-975) is adjacent to the β clamp binding site (residues 920-924, shown in blue) [158] and recent biochemical experiments show that UmuD displaces α from the β clamp [157]. This UmuD binding site is also adjacent to the OB fold (red), where ssDNA binds α [87], an observation that led us to hypothesize that UmuD also inhibits α binding to ssDNA.](image)

The molecule is a well-characterized combination of dsDNA and ssDNA along the force-induced melting transition [60], so waiting at constant extension for a fixed time exposes ssDNA to proteins in solution. The protein-bound ssDNA exhibits a change in observed length upon DNA release, which is a direct
measurement of protein binding to ssDNA. This single molecule technique has been established as a quantitative method of characterizing \( \alpha \)-ssDNA binding \[87\]. As expected, constant extension experiments in the presence of 250 nM \( \alpha \) exhibit large hysteresis, indicating significant ssDNA binding (Figure 3.2B). Fits to Equation 2.6 yield the fraction of ssDNA bound by \( \alpha \), which agrees with previous results obtained by this method \[87\].

![Figure 3.3](image)

**Figure 3.3.** The fraction of ssDNA bound by \( \alpha \) decreases with increasing UmuD concentration. (A) Force extension (solid black line) and release (open circles) curves for DNA in the absence (black) and presence of \( \alpha \) and UmuD (open circles). Open circles are data points, and solid color lines are fits to Equation 2.6, which yield the fraction of ssDNA bound to \( \alpha \) at each UmuD concentration. Fractions shown in these representative curves are 0.58 ± 0.02 in the absence of UmuD (green), 0.44 ± 0.02 at 200 nM UmuD (blue), 0.27 ± 0.02 at 500 nM UmuD (red), 0.09 ± 0.01 at 1 \( \mu \)M UmuD (brown), and 0.06 ± 0.01 at 3 \( \mu \)M UmuD (purple). (B) Fraction of ssDNA bound by \( \alpha \) as a function of effective UmuD concentration in solution (Equation 3.2). Error bars represent standard error (N ≥ 3) for all points except 2.5 \( \mu \)M, which represents propagated error. The solid line is a \( \chi^2 \) fit to a simple DNA binding isotherm (Equation 3.1) that yields apparent equilibrium dissociation constant \( K_{\text{app}} = 340 \pm 103 \) nM between UmuD and \( \alpha \) in the presence of ssDNA, and a saturated \( \alpha \)-ssDNA binding fraction \( f_{\text{sat}} = 0.51 \pm 0.05 \) consistent with previous single molecule results \[87\].

Introducing the DNA damage response protein UmuD disrupts the binding of \( \alpha \) to ssDNA (Figure 3.3A). Constant extension experiments generate ssDNA for 30 minutes in the presence of both \( \alpha \) and UmuD, and the fraction of ssDNA bound by \( \alpha \) decreases with UmuD concentration. Control experiments demonstrate that UmuD does not bind DNA, since force extension curves in the presence of UmuD are the same within uncertainty as those of DNA only (Figure 3.2B, purple). Therefore the interaction between UmuD and \( \alpha \) inhibits \( \alpha \) binding to ssDNA. A simple DNA binding isotherm (Equation 3.1) fit to the fraction of \( \alpha \) as a function of effective UmuD concentration in solution yields the equilibrium dissociation constant \( K_d = 340 \pm 103 \) nM between UmuD and \( \alpha \) in the presence of ssDNA (Figure 3.3B).

### 3.2.2 Specific UmuD variants disrupt the UmuD-\( \alpha \) interaction

To predict potential \( \alpha \) binding sites on UmuD, we used three global protein-protein docking methods. All three methods predicted an ensemble of complexes with UmuD binding near the N-terminal domain and C-terminal domain of DNA pol III \( \alpha \), consistent with the two previously characterized UmuD binding sites [157]. At the C-terminal domain, a number of docking models suggested the formation of a salt bridge (Figure 3.4A) between the arginine residue of DNA pol III \( \alpha \) at position 1068 (Figure 3.4B) and the aspartic acid residue of UmuD at position 91 (Figure 3.4C). As a result, UmuD residues D91, along with its adjacent neutral residue G92, were each mutated to lysine in order to disrupt this salt bridge (Figure 3.4C).
To verify that the mutations did not destabilize UmuD, the melting temperatures of each UmuD variant were determined (Figure 3.5A). It should be noted that UmuD has two melting transitions, which have been assigned to the dissociation of the arms from the globular domain (see drawing in Figure 3.5A) and the melting of the globular domain, respectively [162]. Both melting transitions of the UmuD variants are comparable to those of wild-type UmuD, indicating that the variants are properly folded. To test the effect of each mutation on the cleavage activity of UmuD, a RecA/ssDNA-facilitated UmuD cleavage assay was also performed. All variants were able to cleave efficiently to form the cleavage product UmuD' (Figure 3.5B), so these mutations do not alter UmuD enzymatic activity. Because UmuD can undergo spontaneous cleavage, control reactions where RecA/ssDNA was not added to the UmuD variants were also carried out. The amount of UmuD' present in both the control reactions and the reactions with RecA/ssDNA were quantified separately to distinguish the amount of UmuD' produced only in the RecA/ssDNA facilitated reaction (Figure 3.5B). Thus, the UmuD variants constructed show similar stability and similar RecA-ssDNA-facilitated cleavage efficiency as wild-type UmuD.

Figure 3.4. Docking model predicts residues involved in the interaction between DNA pol III α and UmuD. (A) The docking model predicts a salt bridge between the C-terminal domain of DNA pol III α (pink) and UmuD (yellow) at α residue R1068 (green) and UmuD residue D91 (blue). UmuD residue G92 (red) may also be involved in this interaction. (B) The homology model of α [160] showing arginine residue R1068 (green). The C-terminal domain (residues 917-1160) containing the binding sites for the β clamp, UmuD, and ssDNA is shown in pink. (C) The homology model of full-length UmuD [161] showing the residues D91 (blue) and G92 (red) predicted to bind to DNA pol III α by protein-protein docking studies. The arms of UmuD are shown here in a “trans” conformation where the arm of one monomer (both arms shown in purple) is bound to the globular domain of the other monomer (both globular domains shown in yellow). The active site residues S60 and K97 (cyan) cleave the N-terminal arms (residues 1-24) to form UmuD'.

To verify that the mutations did not destabilize UmuD, the melting temperatures of each UmuD variant were determined (Figure 3.5A). It should be noted that UmuD has two melting transitions, which have been assigned to the dissociation of the arms from the globular domain (see drawing in Figure 3.5A) and the melting of the globular domain, respectively [162]. Both melting transitions of the UmuD variants are comparable to those of wild-type UmuD, indicating that the variants are properly folded. To test the effect of each mutation on the cleavage activity of UmuD, a RecA/ssDNA-facilitated UmuD cleavage assay was also performed. All variants were able to cleave efficiently to form the cleavage product UmuD' (Figure 3.5B), so these mutations do not alter UmuD enzymatic activity. Because UmuD can undergo spontaneous cleavage, control reactions where RecA/ssDNA was not added to the UmuD variants were also carried out. The amount of UmuD' present in both the control reactions and the reactions with RecA/ssDNA were quantified separately to distinguish the amount of UmuD' produced only in the RecA/ssDNA facilitated reaction (Figure 3.5B). Thus, the UmuD variants constructed show similar stability and similar RecA-ssDNA-facilitated cleavage efficiency as wild-type UmuD.
Figure 3.5. UmuD variants are structurally stable and enzymatically active. (A) Melting temperatures of all UmuD variants. The N-terminal “arms” dissociate from the globular domain (drawing above) in the first transition (dashed), and the globular domain melts in the second transition (solid). (B) Relative amount of UmuD′ produced by each variant in the RecA/ssDNA-facilitated self-cleavage reaction (45 min at 37 °C, drawing above), in the absence (dashed) and presence (solid) of RecA/ssDNA. Thermal stability and self-cleavage activity of all variants is similar to wild-type UmuD, suggesting that the mutations had minimal impact on protein stability and function. Error bars represent standard error (N ≥ 3).

As shown in Figure 3.6, the UmuD variant D91K disrupted the DNA pol III α-UmuD interaction by a 15-fold increase in $K_d$ (18 ± 1.7 µM, compared to 1.1 ± 0.6 µM for wild-type UmuD) while no change was seen with the adjacent UmuD variant G92K ($K_d = 1.3 ± 0.5$ µM). When both positions were changed to lysine, D91K-G92K, a similar effect of decreased binding to α was observed ($K_d = 33 ± 9.6$ µM), confirming this position as a binding site for DNA pol III α. Although it is initially surprising that an adjacent residue mutation does not also disrupt UmuD-α interactions, unlike D91, we predict that a side chain at 92 would be angled away from the surface of UmuD (Figure 3.4A), so it is less likely to participate directly in the interaction.

3.2.3 Specific variants disrupt UmuD inhibition of α binding to ssDNA

We used single molecule DNA stretching experiments to test whether the three UmuD variants retain the ability to disrupt binding between α and ssDNA. The biochemical results show that the D91K mutation compromises the ability of UmuD to bind α, and single molecule experiments demonstrate that the fraction of ssDNA bound by α in the presence of 1 µM UmuD D91K is only slightly smaller than that of α alone (Figure 3.7A). UmuD G92K has an affinity for α similar to that of wild-type UmuD, and DNA stretching shows that this variant retains most of its ability to inhibit α binding to
ssDNA (Figure 3.7B). However, the double mutation D91K G92K yields a UmuD variant whose binding to α is dramatically weakened, and is completely unable to disrupt α-ssDNA binding (Figure 3.7C). UmuD residue D91 is therefore required for UmuD to bind α and disrupt its binding to ssDNA. Residue G92 also participates in the interactions that are responsible for UmuD inhibition of α-ssDNA binding. Collectively, these results show that the interaction between UmuD and α is responsible for the ability of UmuD to inhibit α binding to ssDNA.

Figure 3.7. UmuD variants have compromised ability to disrupt α binding to ssDNA. (A-C) Force extension (solid black line) and release (open circles) curves for DNA in the absence (black) and presence of protein (open circles). Open circles are data points, and solid color lines are fits to Equation 2.6, which yield the fraction of ssDNA bound by α. (A) The fraction of α-bound ssDNA in the presence of 1 μM UmuD D91K (blue, 0.54 ± 0.02) is similar to that of α alone (green). The D91K variant lost the ability to inhibit α-ssDNA binding relative to wild-type UmuD (brown), which indicates that the D91 residue is required for the interaction between α and UmuD that disrupts α binding to ssDNA. (B) The G92K variant (red, 0.23 ± 0.02) partially retains its ability to disrupt the interaction between α and ssDNA. (C) The D91K-G92K mutation (purple) fully abolishes the ability of UmuD to disrupt α-ssDNA binding, and fraction of ssDNA bound to α is the same without the UmuD variant (0.58 ± 0.02). (D) Fraction of ssDNA bound to α in the presence of 1 μM UmuD variants. These single molecule results indicate that UmuD residue D91 is essential for the interaction between UmuD and α that disrupts α binding to ssDNA, but G92 is also involved in the interaction.

3.3 Discussion

In this work, we used a single molecule method to demonstrate that UmuD inhibits the binding of α to ssDNA through its interaction with α. The apparent equilibrium dissociation constant $K_d^{\text{app}}$ between UmuD and α is $340 \pm 103$ nM in the presence of ssDNA, which agrees within uncertainty with previous measurements of the equilibrium dissociation constant between α and UmuD ($K_d = 1.1 \pm 0.6$ μM) determined by using a tryptophan fluorescence binding assay in the absence of DNA. The binding affinity measured in bulk solution reflects UmuD binding to both the N- and C-terminal domain sites on α, while
single molecule results are sensitive to UmuD binding to the C-terminal domain of α only, since that site is adjacent to the ssDNA binding site on α. The measured apparent binding affinity between α and UmuD is within the range of cellular UmuD concentrations, suggesting that this interaction plays an important role in regulating α activity in vivo [163].

In addition to directly demonstrating that UmuD inhibits α binding to ssDNA, we used protein-protein docking models to identify potential UmuD-α interaction sites. The resulting models predicted that binding between UmuD and α involves UmuD residues D91 and G92 and α residue R1068. Biochemical experiments confirmed that corresponding UmuD variants exhibit a compromised ability to bind α, despite the fact that the variants were demonstrated to be thermally stable and active for cleavage. Single molecule experiments showed that the D91K variant completely fails to disrupt the α-ssDNA interaction, while the G92K variant only partially inhibits the ability of α to bind ssDNA. These results demonstrate that a direct α-UmuD interaction at these residues is responsible for UmuD inhibition of α-ssDNA binding. Another UmuD variant at position 91, D91A, has been shown to disrupt the UmuD-DinB interaction [154], suggesting the existence of a general binding site for polymerases on UmuD. Furthermore, this region of UmuD may provide a binding site for the protease ClpXP, which plays a role in modulating mutagenesis by degrading UmuD and UmuD′ as well as DinB [164, 165]. Furthermore, the UmuD G92D mutation is poorly cleavable, but when this mutation was constructed in the context of UmuD′, cells harboring this variant were mutable to a similar extent as cells harboring wild-type UmuD [166, 167], suggesting that the G92D mutation does not alter the ability of UmuD′ to facilitate UmuC-dependent mutagenesis. On the other hand, the UmuD G92C variant was as proficient for cleavage as wild-type UmuD [168]. Therefore, this region of UmuD seems to be an important site for a number of protein interactions.

The C-terminal domain of α facilitates numerous interactions necessary for efficient replication, as it interacts with the β clamp [158], ssDNA [87], and the τ subunit of the clamp loader [169, 170]. The interaction between α and the β clamp is essential for high processivity. However, since replication on the lagging strand is carried out in a series of Okazaki fragments that are synthesized in a discontinuous manner, the polymerase on the lagging strand must be recycled for each Okazaki fragment. Although the exact mechanism of the processivity switch is unknown, the OB fold of α has been implicated as a sensor for ssDNA such that when synthesis of an Okazaki fragment is completed, the affinity of α for the β clamp is decreased and α is released from the clamp and from DNA [171-173]. Thus, the interaction between α and ssDNA is proposed to act as a processivity switch. In this work, we demonstrated that SOS-induced levels of UmuD inhibit binding of α to ssDNA. UmuD also inhibits binding of α to the β clamp, and these two mechanisms likely work together to facilitate polymerase exchange.

UmuD, together with UmuC, specifically decrease the rate of DNA replication and therefore were proposed to participate in a primitive DNA damage checkpoint [149-151]. We previously showed that UmuD, but not the cleaved form UmuD′, disrupts the binding of α to the β clamp, which provides a possible molecular explanation for the role of UmuD in a primitive checkpoint [157]. The question then arises of the role of UmuD disruption of ssDNA binding by α. It has been shown that DNA damage on the lagging strand does not disrupt DNA replication, whereas DNA damage on the leading strand may severely disrupt replication or may cause leading strand replication to occur discontinuously as the polymerase can re-initiate synthesis downstream of the damage [136-138]. Polymerases that encounter DNA damage can also become stalled in a futile cycle of insertion and excision of nucleotides [174] as they fail to copy the damaged DNA. Our observations suggest that UmuD would then rescue the stalled polymerase by preventing α from binding to ssDNA or to the β clamp, thereby allowing DNA repair proteins or translesion DNA polymerases access to the damaged DNA.
Taken together, our findings suggest that UmuD specifically disrupts the interaction between α and ssDNA, inhibiting access to the ssDNA template by the replicative polymerase as part of the primitive DNA damage checkpoint to allow DNA repair. UmuD may also prevent binding of α to ssDNA to facilitate polymerase exchange, allowing an error-prone TLS polymerase to copy damaged DNA so that DNA replication may proceed. Our results demonstrate an additional mechanism by which UmuD may regulate the cellular response to DNA damage.

3.4 Materials and Methods

3.4.1 Proteins and plasmids

Wild-type UmuD was expressed from the pSG5 plasmid in BL21 (DE3) (Novagen) as previously described [161, 175]. UmuD D91 and G92 were changed to lysine by site-directed mutagenesis of pSG5 using the QuikChange Kit (Agilent) and confirmed by sequencing the resulting plasmids (Macrogen USA). Wild-type UmuD and all variants were purified as previously described [175]. Wild-type DNA pol III α was expressed from the pET28a-α plasmid in Tuner competent cells (Novagen) and purified using both a Nickel His-trap column (GE Healthcare) and a heparin column (GE Healthcare) as previously described [157]. Fractions collected after the heparin column were diluted 6-fold with buffer HA (50 mM HEPES, pH 7.5; 1 M NaCl; 2 mM beta-mercaptoethanol; 20% glycerol) and loaded onto a hydroxyapatite column (BioRad Bioscale Mini CHT Type 1, 5-mL, 40 μm cartridge) to concentrate the protein; protein concentrator devices were avoided because they significantly reduced the stability and activity of DNA pol III α. After washing with 10 column volumes (cv) of buffer HA, buffer HB (100 mM sodium phosphate, pH 6.5; 1 M NaCl; 2 mM beta-mercaptoethanol; 20% glycerol) was used to elute the protein from the column isocratically in 2 cv. Fractions containing DNA pol III α were dialyzed against protein storage buffer (50 mM HEPES, pH 7.5; 150 mM NaCl; and 50% glycerol) and stored at -20 °C.

3.4.2 Single molecule DNA stretching

In DNA stretching experiments with optical tweezers, a single λ DNA molecule is captured between two polystyrene beads inside a flow cell. One bead is fixed on a micropipette tip, and the other is held in a dual-beam optical trap [159]. As the fixed bead is extended at 100 nm/s, the tethered DNA molecule exerts a force on the trapped bead, which is measured by deflection of the trapping laser beams. The force on the DNA molecule is measured as a function of DNA extension. As the DNA is stretched, the dsDNA helix undergoes a force-induced melting transition into ssDNA (Figure 3.2A). The bases anneal upon DNA release, exhibiting minimal hysteresis, or mismatch between the extension and release curves. After the DNA is stretched and released in buffer only (10 mM HEPES, 100 mM Na⁺, pH 7.5), the solution in the flow cell is exchanged to include protein. Subsequent force-extension curves are obtained in the presence of α, UmuD, or both proteins.

Force-extension curves in the presence of 250 nM α alone exhibit significant hysteresis, since protein bound to the exposed ssDNA prohibits the two strands from annealing upon DNA release. All single molecule experiments were performed by waiting at fixed extension for 30 min prior to DNA release, which has been established as a quantitative method of characterizing α binding to ssDNA [87]. The fraction of ssDNA bound $f_{ss}$ may be described in terms of observed length $b$ (Equation 2.6) [87] where the WLC model (Equation 2.4) describes the polymer elasticity of dsDNA and the FJC model (Equation 2.5) describes ssDNA. As previously described, the fits were confined to forces below 40 pN [87] to eliminate effects from changes in the force-extension curve of ssDNA due to protein binding.

These experiments were repeated in the presence of both α and UmuD, and the fraction of α-bound ssDNA $f_{ss}$ as a function of UmuD concentration $c$ was fit to a simple competitive DNA binding isotherm:
\[ f_{ss} = \left( 1 - \frac{c_s}{K_d^{app}} \right) f_{sat} \]

where \( K_d^{app} \) is the apparent equilibrium dissociation constant between UmuD and \( \alpha \) in the presence of ssDNA, and \( f_{sat} \) is saturated \( \alpha \)-ssDNA binding. A minor correction to added UmuD concentration \( c \) accounts for UmuD bound to \( \alpha \) in solution, so effective UmuD solution concentration \( c_s \) is [176]:

\[ c_s = \frac{c}{1 + K_a c_\alpha} \]

where \( K_a = 9.1 \times 10^{-5} \text{ M}^{-1} \) is the equilibrium association constant between UmuD and \( \alpha \) in bulk solution [157] and \( \alpha \) concentration \( c_\alpha \) is 250 nM.

### 3.4.3 Protein-protein docking

Protein-protein docking models were used to predict residues involved in the binding interaction between \( \alpha \) and UmuD. The structures used for these docking models were a homology model of UmuD [161] and a homology model of full-length DNA pol III \( \alpha \) [160]. Protein complexes where predicted by docking DNA pol III \( \alpha \) with UmuD using ClusPro 2.0 [177-180], GRAMM-X [181], and PatchDock [182]. The top 10 results from each method were analyzed and compared. Local docking was performed using the RosettaDock server [183].

### 3.4.4 Thermal stability assay

A thermal stability assay was used to determine the melting temperature of UmuD variants relative to wild-type, as previously described [184]. To determine whether mutations at positions D91 and G92 disrupt the stability of UmuD, melting temperatures of these variants were compared to those of wild-type. Samples containing 20 \( \mu \text{M} \) of each variant in 50 mM HEPES, pH 7.5, 100 mM NaCl and 25x Sypro Orange (Invitrogen) were exposed to temperatures from 25 °C to 80 °C while monitoring the fluorescence emission intensity at 575 nm. Melting temperatures \( T_m \) were determined by taking the first derivative of the melting curves, as previously described [157].

### 3.4.5 RecA/ssDNA facilitated cleavage assay

Reactions were assembled as previously described [161] and incubated at 37 °C for 45 minutes. After incubation, the cleaved product UmuD' was separated from full-length UmuD using 18% SDS-PAGE. Bands were then analyzed using the image analysis software ImageQuant TL (Amersham Biosciences). Control reactions in the absence of RecA, ssDNA and ATPγS were carried out to determine the amount of UmuD' present due to spontaneous cleavage.

### 3.4.6 Tryptophan fluorescence assay

The equilibrium dissociation constant \( K_d \) between DNA pol III \( \alpha \) and the UmuD variants were determined with a Varian Cary Eclipse Fluorescence Spectrophotometer, as previously described [157]. DNA pol III \( \alpha \) (5 \( \mu \text{M} \) in 50mM HEPES, pH 7.5 and 100 mM NaCl) was titrated with varying volumes of 200 - 400 \( \mu \text{M} \) UmuD variants. Tryptophan fluorescence quenching was used to quantitate binding constants, as previously described [157].
Acknowledgements

We thank Micah McCauley for discussions. We also thank Prof. Zhaohui Zhou for providing access to the fluorimeter. This work was supported in part by NIH [GM072462] (MCW), NSF [MCB-1243883] (MCW), the American Cancer Society [Research Scholar Grant RSG-12-161-01-DMC] (PJB), NSF [Career Award MCB-0845033] (PJB), and the NU Office of the Provost (PJB, LV). SL was supported by the NSF [MCB-1243883, ROA supplement]. KRC was supported by the NSF IGERT Program [DGE-0504331]. PJB is a Cottrell Scholar of the Research Corporation for Science Advancement.
A single zinc finger optimizes the DNA interactions of the nucleocapsid protein of the yeast retrotransposon Ty3

Abstract

Reverse transcription in retroviruses and retrotransposons requires nucleic acid chaperones, which drive the rearrangement of nucleic acid conformation. The nucleic acid chaperone properties of the HIV-1 NC protein have been extensively studied, and nucleic acid aggregation, duplex destabilization, and rapid binding kinetics have been identified as major components of its activity. However, the properties of other nucleic acid chaperone proteins, such as retrotransposon Ty3 NC, a likely ancestor of HIV-1 NC, are not well understood. In addition, it is unclear whether a single zinc finger is sufficient to optimize the properties characteristic of HIV-1 NC. We used single molecule DNA stretching as a method for detailed characterization of Ty3 NC chaperone activity. We found that wild-type Ty3 NC aggregates single- and double-stranded DNA, weakly stabilizes dsDNA, and exhibits rapid binding kinetics. Single molecule studies in the presence of Ty3 NC mutants show that the N-terminal basic residues and the unique zinc finger at the C-terminus are required for optimum chaperone activity in this system. While the single zinc finger is capable of optimizing Ty3 NC’s DNA interaction kinetics, two zinc fingers may be necessary in order to facilitate the DNA destabilization exhibited by HIV-1 NC.

This work was originally published in Nucleic Acids Research 40: 751-760 (2012), and has been adapted for this dissertation.
4.1 Introduction

Retrotransposons are mobile genetic elements, closely related to retroviruses, which use their own proteins in conjunction with cellular machinery to replicate independently of the genome. The yeast retrotransposon Ty3, and retroviruses such as HIV-1, encode RT, the DNA polymerase that converts the ssRNA genome of positive polarity into dsDNA. During this reverse transcription process, two obligatory DNA strand transfers occur to generate the long terminal repeats that control genomic DNA integration and its expression (reviewed in [114]). In the early stages of replication, a tRNA primer anneals to the PBS close to the 5′ end of the ssRNA, where RT synthesizes the cDNA strand, referred to as minus strand strong stop DNA (-sssDNA). While RT synthesizes the minus strand -sssDNA, its RNaseH domain degrades the 5′ end of the RNA template. Further DNA polymerization requires -sssDNA transfer, where the newly-synthesized single-stranded cDNA anneals to the 3′ untranslated region (UTR) of the RNA template [114-116]. However, thermodynamically stable structures within the cDNA and the 3′ UTR RNA make duplex formation improbable. The NC protein acts as a nucleic acid chaperone that destabilizes these secondary structures and facilitates DNA–RNA hybrid formation [96, 185-188]. NC also participates in the second DNA strand transfer, where the plus DNA strand anneals to the structured LTR of the minus DNA strand [99-101, 189]. These transfer events are required for RT to transcribe the remaining genomic RNA into dsDNA flanked by the LTR, which the homologous enzyme integrase inserts into the cellular genome.

Replication of retroelements requires the nucleic acid chaperone activity of NC proteins, which tend to be small cationic proteins with minimal structure other than one or two CCHC-motif zinc fingers [92-95, 114-116, 190]. HIV-1 NC is 55 amino acids in length, with two zinc fingers and a basic N-terminal tail [114-116]. It directs annealing of the tRNA\textsuperscript{Lys} primer to the PBS of the genomic RNA [96-98]. The viral RNA has 5′ and 3′ repeat regions that include stable trans-activation response element (TAR) hairpins, and HIV-1 NC facilitates rearrangement of these nucleic acids during the minus strand transfer step of reverse transcription [99-101]. NC is also an essential nucleic acid chaperone involved in genomic RNA dimerization and virus assembly [102-104, 191, 192]. The chaperone activity of HIV-1 NC involves nucleic acid aggregation, duplex destabilization, and rapid binding kinetics [88, 105, 111, 117, 193, 194]. Nucleic acid aggregation, an effect due to protein-induced interactions that make the DNA molecule attracted to itself, is associated with the cationic domain of HIV-1 NC. Duplex destabilization involves aromatic residues in the zinc fingers, which stack with ssDNA bases [118, 119]. Thus optimal chaperone activity of HIV-1 NC is highly sensitive to the zinc finger architecture, and relies on a delicate balance between destabilizing secondary structures and promoting complementary strand annealing.

A recent study examining the nucleic acid chaperone activity of NC proteins from several retroviruses established that the nucleic acid interaction characteristics that are important for nucleic acid chaperone activity varied significantly for different viruses [112, 113]. From these studies it is clear that HIV-1 NC has optimal chaperone activity compared to other NC proteins. Although nucleic acid chaperones from HIV-1 and other retroviruses have been extensively studied, less is known about chaperone activity in LTR retrotransposons [159, 185, 195]. Because Ty3 NC is likely to be an ancestor of HIV-1 NC, here we examine its nucleic acid chaperone properties and compare them to those of HIV-1 NC. In addition to allowing comparison of Ty3 NC to the paradigm of nucleic acid chaperone proteins, understanding these differences may also be useful for the development of drugs that target the rapidly evolving retrovirus HIV-1 NC [89, 108, 196]. Ty3 NC is a 57-residue protein with one zinc finger and a basic N-terminal tail [197]. Ty3 NC is monomeric in solution, and it is required for efficient annealing of the tRNA\textsuperscript{Met} primer on the bipartite Ty3 PBS and RNA dimerization [198]. The untranslated regions of the Ty3 genome also form secondary structures, and minus strand transfer requires Ty3 NC to overcome the energetic barrier in forming the DNA-RNA duplex necessary for further reverse transcription.
There are distinct structural differences between Ty3 NC, which has only one zinc finger with two aromatic residues, and HIV-1 NC, which has two zinc fingers with one aromatic residue each. The cationic tail of Ty3 NC, which has minimal structure, also contains the aromatic residue tyrosine. Ty3 NC, with 20 basic residues (pI ≈ 11.5), has a slightly higher overall charge than HIV-1 NC (pI ≈ 10), with 15 basic residues. Although minimal information is available about the three-dimensional structure of Ty3 NC, it is extremely likely to have high charge density, consistent with retroviral NC proteins of known structure such as HIV-1 NC and MuLV NC [199]. Furthermore, the length and structural complexity of the terminal untranslated regions seem to be related to the chaperone activity of a number of retroviral NC proteins, and the secondary structures in the Ty3 UTRs are not as thermodynamically stable as the HIV-1 TAR hairpins [193].

DNA stretching and other biophysical methods have been used to show that the properties of HIV-1 NC are specifically tuned to optimize its interactions with both dsDNA and ssDNA. Specifically, both zinc fingers are required to maintain their native order and structure, and seemingly minor mutations result in loss of rapid kinetics of the DNA-protein interaction [88]. Other retroviral NC proteins exhibited less rapid kinetics in DNA stretching experiments, particularly MuLV and HTLV-1 NC [112, 113]. Notably, MuLV NC is the only NC protein studied that contained a single zinc finger, and it exhibited relatively slow kinetics. Thus, previous results suggest that two zinc fingers are generally required to optimize the kinetics of NC-DNA interactions [88, 111-113]. To determine the capability of a single zinc finger to facilitate NC-DNA interactions and to understand the connection between Ty3 NC's structure and function as a nucleic acid chaperone, we probed the thermodynamics and kinetics of wild-type and mutant NC-DNA interactions with single DNA molecules. We found that, despite the presence of only a single zinc finger, wild-type Ty3 NC aggregates nucleic acids, promotes complementary strand annealing, and exhibits rapid kinetics resembling that of HIV-1 NC. Four mutants of Ty3 NC demonstrate that these chaperone properties require the zinc finger as well as the presence of its cationic N-terminal tail. Thus, Ty3 NC's single zinc finger is necessary and sufficient to facilitate rapid DNA-protein interactions, but a second zinc finger may be required for more effective nucleic acid destabilization by NC proteins, such as that observed for HIV-1 NC and TAR hairpins.

4.2 Results

4.2.1 Wild-type Ty3 NC binds DNA with rapid kinetics and slightly stabilizes dsDNA

Optical tweezers were used to stretch and release a single molecule of bacteriophage λ DNA in the absence of protein and in the presence of a nearly saturated concentration of wild-type Ty3 NC (Figure 4.1A). While bacteriophage λ DNA is not the sequence acted on by Ty3 NC, it represents a random sequence, allowing us to measure the ability of the protein to remodel general nucleic acid structures. In the absence of protein, B-form dsDNA uncoils with a gradual increase in force at low extensions. As the molecule approaches its contour length of 0.34 nm/bp, the elasticity of the backbone causes a sharp increase in force. At approximately 60 pN, the DNA molecule lengthens more than 1.5 times with minimal increase in force as it undergoes a cooperative force-induced melting transition from dsDNA to ssDNA [37, 60, 63]. The force increases rapidly again at the end of this cooperative phase transition, at 0.6 nm/bp. As the DNA is released back to low extensions, the force-extension curve is almost completely reversible, exhibiting only minor hysteresis, which is the difference between the extension and release curves. The area between these curves characterizes the amount of hysteresis.
In the presence of 5 nM wild-type Ty3 NC, the force increases at low extensions, below the λ DNA contour length (Figure 4.1A). The transition at the end of the melting plateau also shifts to lower extensions, decreasing by approximately 0.025 nm.bp. These effects are due to nucleic acid aggregation, in which protein-induced interactions make the DNA molecule attracted to itself. A five-fold increase in protein concentration leads to small increases in these two effects, reflecting additional nucleic acid aggregation (Figure 4.1B).

The force-induced melting plateau is slightly sloped in the presence of wild-type Ty3 NC. This increase in transition width reflects a loss of DNA melting cooperativity, which indicates that the DNA molecule is more likely to undergo conformational rearrangements, an effect which has been observed for multiple nucleic acid chaperone proteins [111, 112, 130, 200]. A simple DNA binding isotherm (Equations 4.1 and 4.2) fit to the change in transition width ∆F as a function of protein concentration c yields an equilibrium dissociation constant $K_d = 3.5 \pm 0.5$ nM and saturated transition width $\Delta F_{sat} = 6.2 \pm 0.4$ pN. Protein concentrations significantly above saturation (80 – 150 nM) were also included in the $\chi^2$ fit (data not shown).

Figure 4.1. Typical force-extension (solid) and release (dashed) curves of λ-DNA in the presence of wild-type Ty3 NC. (A) DNA only (black) and 5 nM wild-type Ty3 NC (green). (B) 5 nM (green) and 25 nM (blue) wild-type Ty3 NC. (C) Change in the transition width $\Delta F$ of DNA force-induced melting as a function of wild-type Ty3 NC concentration. $\Delta F = \delta F - \delta F_0$, where $\delta F$ is the melting transition width in the presence of protein and $\delta F_0 = 3.6 \pm 0.3$ pN, the melting transition width of DNA only. Standard error determined from at least three measurements was used to compute error bars for $\Delta F$. A $\chi^2$ fit (green line) to a simple DNA binding isotherm (Equations 4.1 and 4.2) yields $K_d = 3.5 (\pm 0.5)$ nM and $\Delta F_{sat} = 6.2 (\pm 0.4)$ pN. Protein concentrations significantly above saturation (80 – 150 nM) were also included in the $\chi^2$ fit (data not shown).

Figure 4.2. Structure of (A) wild-type Ty3 NC, (B) Δ1-NCP9, (C) Δ2-NCP9, (D) NCP9 dd, and (E) Δ2-NCP9 dd. Basic residues shown in blue and zinc coordinating residues shown in green.
0.4) pN (Figure 4.1C, green line).

The lack of hysteresis indicates that Ty3 NC exhibits rapid kinetics, allowing it to dissociate quickly from ssDNA. Rapid kinetics, which is characteristic of nucleic acid chaperones, is correlated with a protein’s ability to facilitate reannealing. In contrast, proteins that preferentially bind ssDNA, such as T4 gp32, induce significant hysteresis, and proteins that preferentially bind dsDNA, such as HMG, stabilize the duplex and increase the melting force [17]. The melting force in the presence of wild-type Ty3 NC is slightly higher than that of DNA only (Table 4.1), which indicates net stabilization of the DNA duplex. In order to elucidate how a protein that stabilizes duplex DNA can act as a nucleic acid chaperone, four mutants of Ty3 NC were used to further investigate the DNA binding role of the zinc finger and N-terminal tail (Figure 4.2). The N-terminal tail was deleted to varying degrees in mutants Δ1-NCp9 and Δ2-NCp9, leaving the zinc finger intact. In mutant Δ2-NCp9 dd, however, the zinc finger has also been deleted. Mutant NCp9 dd has a largely intact N-terminal tail, but no zinc finger.

4.2.2 Cationic tail deletions destroy the rapid kinetics and duplex stabilization of wild-type Ty3 NC

The N-terminal tail of mutant Δ1-NCp9 is missing the first 16 residues, compromising the protein’s DNA binding affinity by approximately five-fold (Table 4.1). $K_d$ was estimated as 12 (± 3) nM using a simple DNA binding isotherm to approximate the decrease in melting force as a function of protein concentration (Equations 2.7, 4.1).

In striking contrast to wild-type Ty3 NC, force-extension curves in the presence of 20 nM Δ1-NCp9 exhibit significant hysteresis (Figure 4.3A). The hysteresis ratio for Δ1-NCp9 is 0.68 (±0.1), relative to 0.26 (±0.03) for wild-type (Table 4.1). This indicates that the mutant does not dissociate from ssDNA on the timescale of the experiment, leading to the loss of rapid kinetics. Subsequent stretches reflect incomplete protein dissociation (Figure 4.3B). Δ1-NCp9 induces less aggregation than wild-type Ty3 NC, and has a minimal effect on the force-induced melting transition width. Furthermore, the melting force is 59.4 (± 0.7) pN in the presence of the mutant, which is significantly lower than the 68.0 (± 1.0) pN melting force in the presence of wild-type Ty3 NC (Table 4.1). This result reflects weak duplex destabilization by Δ1-NCp9.

![Figure 4.3](image-url). Typical force-extension (solid) and release (dashed) curves of (A) DNA only (black) and DNA in the presence of (A, B) 20 nM Ty3 NC mutant Δ1-NCp9. First stretch-release curve shown in green, and (B) second stretch-release curve shown in blue.
To identify the DNA binding activity of the zinc finger, we examined the Δ2-NCp9 mutant. Deletion of the N-terminal tail leaves a mutant that primarily consists of the zinc finger, and has DNA binding affinity more than an order of magnitude smaller than that of wild-type. Although Δ2-NCp9 does not affect DNA melting force or transition width to allow for an estimate of \( K_d \), protein concentrations smaller than 50 nM reflected no DNA binding. In the presence of 50 nM protein, the first force-extension curve lies almost exactly on top of the DNA only curve, showing a negligible effect (Figure 4.4A). The second stretch-release cycle exhibits some hysteresis, indicating slow kinetics (Figure 4.4B), which increases further in the third force-extension curve (Figure 4.4C). The Δ2-NCp9 mutant is the only Ty3 NC protein for which successive stretch-release cycles show additional protein binding. The third stretch-release cycle is required to observe an effect comparable to the other proteins presented in Table 4.1 because Δ2-NCp9 binds DNA with significantly lower affinity than wild-type or the other mutants. Furthermore, these effects remain the same within uncertainty, even up to 100 nM protein concentration. Therefore the hysteresis ratio of 0.49 (±0.1) is reported in Table 4.1 for comparison. There is less hysteresis than that for Δ1-NCp9 (Table 4.1), which is consistent with the likelihood that Δ2-NCp9 has higher charge density.

DNA stretching curves in the presence of mutants dominated by the zinc finger imply that the zinc finger of Ty3 NC contributes to nucleic acid aggregation and duplex destabilization. Both mutants with a compromised N-terminal tail also lack the rapid kinetics of wild-type Ty3 NC, demonstrating that the cationic tail is required to promote nucleic acid annealing.

4.2.3 Zinc finger deletions destroy rapid kinetics and induce strong duplex stabilization

To identify the DNA binding properties of the N-terminal tail, we examined two mutants that lack the zinc finger entirely, NCP9 dd and Δ2-NCp9 dd. The zinc finger deletion in mutant NCP9 dd leaves the cationic tail nearly intact, resulting in DNA binding affinity most similar to that of wild-type Ty3 NC, with...
\( K_d \) estimated as 3 (±2) nM (Table 4.1). The initial force-extension curve in the presence of 3 nM NCp9 dd exhibits a significantly higher melting force of 71.5 (±0.4) pN, which reflects strong dsDNA stabilization (Figure 4.5A, Table 4.1). The second stretch exactly follows the previous release curve, indicating that the mutant binds ssDNA irreversibly on the timescale of the experiment (Figure 4.5B). NCp9 dd also induces strong DNA aggregation, indicated by an increase in the DNA stretching force to 10-20 pN below the DNA contour length, an effect that dominates subsequent stretches.

Figure 4.5. Typical force-extension (solid) and release (dashed) curves of (A) DNA only (black) and DNA in the presence of (A, B) 3 nM Ty3 NC mutant NCp9 dd. First stretch-release curve shown in green, and (B) second stretch-release curve shown in blue.

The Δ2-NCp9 dd mutant is Δ2-NCp9 without the zinc finger. Deletion of both the N-terminal tail and the zinc finger results in a small protein likely to have a high charge density, with DNA binding affinity an order of magnitude smaller than that of wild-type Ty3 NC. A DNA binding isotherm (Equations 2.7, 4.1) fit to change in melting force \( \Delta F_m \) as a function of protein concentration \( c \) yields \( K_d = 20 (± 1) \) nM with saturated change in melting force \( \Delta F_{m, sat} = 16 (± 0.5) \) pN. The force-extension curve in the presence of 13 nM Δ2-NCp9 dd exhibits duplex stabilization, DNA aggregation (10 pN force is required to stretch DNA below the dsDNA contour length), and large hysteresis (Figure 4.6A). In contrast with Δ2-NCp9 (Figure 4.5B), subsequent stretch-release cycles reflect incomplete protein dissociation (Figure 4.6B).

Figure 4.6. Typical force-extension (solid) and release (dashed) curves of (A) DNA only (black) and DNA in the presence of (A, B) 13 nM Ty3 NC mutant Δ2-NCp9 dd. First stretch-release curve shown in green, and (B) second stretch-release curve shown in blue.

DNA force-extension curves in the presence of mutants without the zinc finger indicate that the cationic tail of Ty3 NC contributes strongly to nucleic acid aggregation, and is largely responsible for dsDNA
stabilization. Both mutants also induce strong hysteresis, and this inhibition of DNA annealing abolishes rapid kinetics, which requires both the zinc finger and the N-terminal tail of Ty3 NC.

<table>
<thead>
<tr>
<th>Ty3 NC Protein</th>
<th>c (nM)</th>
<th>$F_m$ (pN) (A)</th>
<th>Hysteresis (Ratio) (B)</th>
<th>$K_d$ (nM) (C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>no protein</td>
<td></td>
<td>61.0 (±0.5)</td>
<td>0.10 (±0.03)</td>
<td></td>
</tr>
<tr>
<td>wild-type</td>
<td>5</td>
<td>68.0 (±1.0)</td>
<td>0.26 (±0.03)</td>
<td>3.5 (±0.5)</td>
</tr>
<tr>
<td>Δ1-NCp9</td>
<td>20</td>
<td>59.4 (±0.7)</td>
<td>0.68 (±0.1)</td>
<td>12 (±3)</td>
</tr>
<tr>
<td>Δ2-NCp9</td>
<td>50</td>
<td>60.6 (±0.1)</td>
<td>0.49 (±0.1)</td>
<td></td>
</tr>
<tr>
<td>NCP9 dd</td>
<td>3</td>
<td>71.5 (±0.4)</td>
<td>0.55 (±0.1)</td>
<td>3 (±2)</td>
</tr>
<tr>
<td>Δ2-NCp9 dd</td>
<td>13</td>
<td>67.1 (±0.5)</td>
<td>0.67 (±0.1)</td>
<td>20 (±1)</td>
</tr>
</tbody>
</table>

Table 4.1. Melting force and hysteresis area in the presence of wild-type Ty3 NC and its mutants, measured at protein concentrations $c$ (force-extension curves shown in Figures 4.1, 4.3-4.6) near or above $K_d$. Equilibrium dissociation constants $K_d$ reported for 50 nM Na⁺. All values were calculated with at least three measurements, reflecting the uncertainty reported (standard error). (A) The melting force is an average along the length of the force-induced melting plateau in the presence of each protein. (B) Hysteresis is an area ratio that reflects amount of protein still bound upon DNA release (see Figure 4.7B). The ratio increases with the amount of ssDNA observed upon DNA release, up to a maximum value of 1. (C) $K_d$ was quantified from the change in transition width in the case of wild-type Ty3 NC (Figure 4.1C). Change in melting force was used to determine $K_d$ for Ty3 NC mutants (Figure 4.7C), with the exception of Δ2-NCp9, which did not affect melting force upon DNA binding. (D) Ty3 NC mutant Δ2-NCp9 did not appreciably bind DNA on the first stretch-release cycle (Figure 4.4A) over two orders of magnitude in protein concentration (1 nM – 100 nM). At protein concentrations 50 nM and above, binding that altered hysteresis was achieved only upon the third stretch-release cycle (Figure 4.4B-C) and this effect did not change significantly at higher concentrations. Therefore data from the third force-extension curve is included in Table 4.1 to provide a comparison of melting force and hysteresis ratio for this mutant.

4.3 Discussion

Recent studies of nucleic acid chaperone proteins from retroviruses and retrotransposons have shown that aggregation, duplex destabilization and rapid kinetics are often characteristic of their chaperone activity. Single molecule stretching experiments have also shown that an increase in the slope of the DNA melting plateau is positively correlated with nucleic acid chaperone activity [112, 130, 200]. Force-extension curves in the presence of wild-type Ty3 NC exhibit this increase in transition width, and demonstrate that aggregation of both single- and double-stranded nucleic acids is a key component of its chaperone activity. The increase in the force at extensions below the DNA contour length has also been observed with retroviral NCs such as HIV-1 NC and RSV NC [112]. In addition, the decrease in ssDNA contour length has been reported for ORF1p, the nucleic acid chaperone from the LINE-1 retrotransposon [130, 200]. DNA stretching curves in the presence of Ty3 NC also exhibit very little hysteresis, indicating that the protein dissociates quickly from ssDNA, allowing the strands to anneal, and rapid kinetics is a key component of its nucleic acid chaperone activity. These results contrast significantly with those obtained in the presence of MuLV NC, the only other single zinc finger NC protein studied by DNA stretching [112]. DNA stretching curves in the presence of MuLV NC exhibited significant hysteresis and much less aggregation. Although the previous results suggested that two zinc fingers may be required for rapid DNA interaction kinetics, this study shows that rapid kinetics can be observed even with a single zinc finger NC protein.

We examined several mutants to investigate the relationship between the structure of Ty3 NC and its nucleic acid chaperone function. Stretching curves in the presence of Δ1-NCp9, which has a partially deleted N-terminal tail, indicate that the zinc finger contributes to weak duplex destabilization, an effect that balances the strong duplex stabilization of the cationic tail. Although the distribution of charge depends upon the structure of Ty3 NC mutants bound to nucleic acids, which is unavailable, both
mutants without the zinc finger must have high charge density. Both of these mutants, NCp9 dd and Δ2-NCp9 dd, significantly increase the melting force, to 71.5 (±0.4) pN and 67.1 (± 0.5) pN, respectively (Table 4.1). The N-terminal tail mutants demonstrate that the cationic residues on the N-terminus stabilize dsDNA, while the zinc finger destabilizes dsDNA. When the cationic region is deleted, DNA destabilization by the zinc finger, which contains two aromatic residues that could potentially stack with ssDNA, dominates the overall thermodynamics of the DNA-protein interaction. When the zinc finger is deleted, the protein strongly stabilizes dsDNA.

While duplex stabilization such as that observed for wild-type Ty3 NC is likely somewhat detrimental to nucleic acid chaperone function, this property is a consequence of the high charge density of the protein. A high charge density is required for efficient nucleic acid aggregation, which is a primary requirement for nucleic acid chaperone activity. Thus, wild-type Ty3 NC balances aggregation by its cationic residues and duplex destabilization by its zinc finger, resulting in overall mild duplex stabilization in conjunction with strong aggregation. Although all four mutants aggregated both dsDNA and ssDNA, the effect was strongest with zinc finger deletion. Mutant Δ2-NCp9, which is primarily the zinc finger, showed the least DNA aggregation as well as the weakest DNA binding. Even at 100 nM protein concentration, nearly four-fold higher than saturated binding of wild-type Ty3 NC, multiple stretch-release cycles were required to see an effect. In contrast, the zinc finger deletion of NCp9 dd is likely to result in higher charge density, preserving DNA binding affinity. These properties are consistent with the ability of NCp9 dd to facilitate tRNA primer annealing, promote genomic RNA dimerization, and initiate cDNA synthesis *in vitro* because these processes do not require significant duplex destabilization to proceed [201].

The rapid kinetics of wild-type Ty3 NC is highly sensitive to protein charge and structure, and deletion of either the zinc finger or the cationic tail forms mutants that induce hysteresis, which demonstrates that rapid kinetics requires both the zinc finger and the N-terminal tail. In fact, only wild-type Ty3 NC has rapid kinetics, quickly switching between binding single- and double-stranded nucleic acids. The weak duplex destabilization effect of the zinc finger and the strong duplex stabilization effect of the cationic tail work in concert to transiently destabilize secondary structures in the 5‘ and 3‘ UTR and facilitate annealing of the DNA:RNA duplex during reverse transcription [115, 202]. Thus, a single zinc finger is sufficient for a highly charged NC protein to exhibit rapid DNA-protein interaction kinetics. Although the high charge density of the protein strongly stabilizes dsDNA, the single zinc finger also destabilizes DNA, and the resulting balance of these two effects makes Ty3 NC a very strong nucleic acid aggregation protein with only weak duplex stabilization. If Ty3 required significant duplex destabilization to facilitate minus strand transfer, as is the case for HIV-1, a second zinc finger would likely be needed. In fact, HIV-1 NC probably evolved a second zinc finger for this purpose, in conjunction with the requirements of other critical HIV-1 replication processes, due to the high thermodynamic stability of its TAR hairpins. This suggests that nucleic acid chaperone activity of NC proteins is specifically tuned to secondary structures in the untranslated regions of their genomic RNA. The nucleic acid chaperone properties are then optimized to maximize DNA aggregation while facilitating only the necessary secondary structure rearrangements required for RT to synthesize a complete functional viral DNA.

### 4.4 Materials and Methods

Single molecule DNA stretching experiments were performed as previously described [17, 159]. Dual beam optical tweezers were used to capture biotin-labeled bacteriophage λ DNA between two streptavidin-coated polystyrene beads (Bangs Labs). Surrounding DNA molecules were rinsed out of solution with buffer (50 mM Na⁺, 10 mM HEPES, pH 7.5), and the captured DNA molecule was then stretched and released at a pulling rate of 100 nm/s to obtain DNA-only force-extension curves (Figure
4.7A, black), which are typically characterized by models of polymer elasticity. The WLC (Equation 2.4) and FJC (Equation 2.5) models describe dsDNA (Figure 4.7A, blue) and ssDNA (red), respectively.

**Figure 4.7.** (A) Typical force-extension (solid) and release (dashed) curves of λ-DNA (black) obtained with optical tweezers. The WLC model (blue) describes dsDNA. Near the dsDNA contour length, the molecule undergoes a force-induced melting transition, from dsDNA to ssDNA. The FJC model describes ssDNA (red). Minimal hysteresis is evident in these solution conditions (50 mM Na+, 10 mM HEPES, pH 7.5). (B) Quantification of the hysteresis area ratio for a typical DNA extension and release curve. Force-extension (solid) and release (dashed) curve of DNA in the presence of 13 nM Ty3 NC Δ2-NCp9 dd are shown in green. The WLC and FJC models are shown in blue and red, respectively. A linear combination of these two models is shown in black, indicating the fraction of ssDNA exposed to solution upon DNA extension (Equation 2.6). Relative hysteresis is the ratio of A1, the area between the stretch (solid green) and release (dashed green) curves, and A2, the area between the stretch (solid green) and melted DNA fraction (black) curves. (C) Equilibrium dissociation constant $K_d$ determined from change in average melting force $\Delta F_m$ as a function of protein concentration c fit to a simple DNA binding isotherm (Equations 2.7, 4.1). Data points for mutant Δ2-NCp9 dd are shown with standard error bars, with a fit (green line) that yields $K_d = 20 \pm 1$ nM and saturated melting force $\Delta F_{m\text{sat}} = 16 \pm 0.5$ pN. $K_d$ was estimated for mutants Δ1-NCp9 and NCp9 dd with this method, but could not be obtained for mutant Δ2-NCp9, which did not affect DNA melting force.

After obtaining force-extension curves of a single DNA molecule, the buffer was replaced with protein dilutions. Ty3 NC and its mutants were introduced into solution with the DNA held under slight tension in an effort to reduce formation of extremely stable protein-DNA aggregates during solution exchange. The first force-extension curve after protein exchange was therefore stretched from zero force to the original starting force after DNA release (example in Figure 4.7B, solid green curve, starting force of 10 pN). Subsequent force-extension curves begin and end at zero force. At least three stretch-release cycles were performed for each measurement, and subsequent force-extension curves were the same within uncertainty in the case of wild-type Ty3 NC. The 50 mM Na+ concentration was used for comparison with earlier studies of other NC proteins. Both the salt and protein concentration in vivo are higher than those used here, but the increased binding in low salt tends to compensate for the lower protein concentration.

Melting forces were determined by averaging along the length of the force-induced melting plateau in the presence of each protein. Hysteresis was quantified by calculating $A_1$, the area between the force-extension (Figure 4.7B, solid green) and release (Figure 4.7B, dashed green) curves of DNA in the
presence of wild-type Ty3 NC and its mutants. A linear combination of the WLC and FJC models (Figure 4.7B, blue and red, respectively) indicates the fraction of ssDNA generated upon DNA extension (Figure 4.7B, black), as described in [60] (Equation 2.6). The total amount of hysteresis possible is $A_2$, the area between the extension curve (Figure 4.7B, solid green) and this linear combination (Figure 4.7B, black). Relative hysteresis is the ratio of $A_1$ and $A_2$, ranging from 0.1 (±0.03) in the absence of protein, to a theoretical maximum of 1, which would indicate that all the ssDNA generated is bound by protein upon DNA release and does not dissociate on the time scale of the release (1 minute). Averages and uncertainties (standard error) for all values reported were calculated using at least 3 measurements.

To quantify the effect of wild-type Ty3 NC on the shape of the force-induced melting plateau, transition width $\delta F$ was determined from the slope at the midpoint of melting transition, as described in [111, 200]. The change in transition width is $\Delta F = \delta F - \delta F_0$, where $\delta F_0$ is the transition width in the absence of protein. Change in transition width as a function of protein concentration $c$ may be described by a simple DNA binding isotherm (neglecting site exclusion, $n = 1$ in Equation 2.8) [130]:

$$\theta = \frac{c}{K_d}(1 - \theta)$$

where $K_d$ is the equilibrium dissociation constant and the fractional DNA binding $\theta$ is given by the change in transition width:

$$\theta = \frac{\delta F(c) - \delta F_0}{\delta F_{sat} - \delta F_0}$$

with $\delta F_{sat}$ defined as the transition width at saturated protein concentration. Standard error for measured data points was calculated from at least three measurements, and fits to Equations 4.1 and 4.2 were performed using $\chi^2$ analysis.

Ty3 NC mutants have negligible effect on the transition width, so $K_d$ was obtained from their effect on the average melting force $F_m$, as reviewed in [17]. Change in melting force is $\Delta F_m = F_m - F_m^0$, where $F_m^0 = 61.0$ (±0.5) pN, the melting force in the absence of protein. $\Delta F_m$ as a function of protein concentration $c$ may be described by Equation 2.7, where the fractional binding $\theta$ is given by the change in melting force. Data points with standard error were fit with Equations 2.7 and 4.1 to quantify $K_d$ for mutant $\Delta 2$-NCp9 dd (Figure 4.7C). This method was used to estimate $K_d$ for mutants $\Delta 1$-NCp9 and NCp9 dd, but mutant $\Delta 2$-NCp9 did not affect melting force over two orders of magnitude change in protein concentration.

Ty3 NC protein of 57 amino acids and its mutants were synthesized on the solid phase using fmoc (9-fluorenylmethoxycarbonyl)- and opfp (pentafluorophenyl ester)-protected amino acids and purified to homogeneity by HPLC [201].

Acknowledgements

We thank Micah McCauley and Ioulia Rouzina for insightful discussions. D. Ficheux (IBCP du CNRS, Lyon, France) is gratefully acknowledged for the synthesis of Ty3 NC and its derivatives. This work was supported by NIH [R01GM75965]; NSF [MCB-0744456]; Institut national de la santé et de la recherche médicale; and Agence nationale de recherches sur le sida et les hépatites virales. KC was supported by the NSF IGERT Program [DGE-0504331].
Paired mutations abolish and restore the balanced annealing and melting activities of ORF1p required for LINE-1 retrotransposition

Abstract
Retrotransposition amplifies LINE-1 to high copy number in mammalian genomes. The L1 protein encoded by ORF1, ORF1p, is required for retrotransposition. This dependence on ORF1p was investigated by mutating three highly conserved residues, R238, R284 and Y318 to alanine, thereby inactivating retrotransposition. R284A and Y318A were rescued by further substituting the alanine with the appropriate conservative amino acid, e.g., lysine or phenylalanine, respectively, whereas R238K remained inactive. Quantification of the steady-state levels of L1 RNA and ORF1p failed to discriminate active from inactive variants, indicating loss of L1 retrotransposition resulted from loss of function rather than reduced expression. The two biochemical properties known for ORF1p are high-affinity RNA binding and nucleic acid chaperone activity. Only R238A/K exhibited significantly reduced RNA affinities. The nucleic acid chaperone activities of the remaining paired mutants were assessed by single molecule DNA stretching and found to mirror retrotransposition activity. To further examine ORF1p chaperone function, their energetic barriers to DNA annealing and melting were derived from kinetic work. When plotted against each other, the ratio of these two activities distinguished functional from non-functional ORF1p variants. These findings enhance our understanding of the requirements for ORF1p in LINE-1 retrotransposition and, more generally, nucleic acid chaperone function.

This work was originally published in Nucleic Acids Research 39: 5611-5621 (2011), and has been adapted for this dissertation. Ensemble results included in this chapter are the work of James Evans, Suresh Peddigari, and Sandra Martin at the University of Colorado School of Medicine.
5.1 Introduction

LINE-1 is a successful and active retrotransponson in mammals, and L1 sequence directly accounts for 17% of the human and 19% of the mouse genomes. L1 also mobilizes non-autonomous elements including short interspersed elements (SINEs) and processed pseudogenes, making it responsible for at least 30% of mammalian DNA [203]. The structure and function of the genome is impacted by L1 at multiple levels, which range from insertional mutagenesis to altering gene expression and promoting recombination [16].

A majority of the 660,000 L1 elements in the mouse genome are truncated and therefore cannot retrotranspose, but approximately 3,000 are estimated to be competent for retrotransposition [204]. Active mouse L1 is characterized by a series of monomer promoter motifs followed by a 5′ UTR, two open reading frames (ORF1 and ORF2), a 3′ UTR, poly-A signal and an A-rich region, all embedded within a short target site duplication. Retrotransposition begins with transcription of one of the active elements, followed by translation of the two L1-encoded proteins, ORF1p and ORF2p. Both L1 proteins are essential for retrotransposition [205], and both are required in cis [206]. The L1 RNA ultimately serves as a template for target-primed reverse transcription [14] resulting in a new L1 insertion in the genome.

The role of ORF1p in L1 retrotransposition is incompletely understood. The protein forms a stable trimer via an N-terminal coiled-coil domain [128], followed by an RNA recognition motif (RRM) [207] and a CTD [208] (Figure 5.1A). Two biochemical properties of ORF1p are known: RNA binding and nucleic acid chaperone activity. ORF1p from mouse L1 is a high-affinity, non-sequence-specific RNA binding protein [209], which is likely critical for packaging the L1 RNA during retrotransposition [210-213]. The RNA binding region of the protein is C-terminal of the coiled-coil, encompassing the RRM and CTD domains, which are both required for high-affinity RNA binding [126, 207]. ORF1p is also a nucleic acid chaperone protein, a property which is required for retrotransposition [129], and may facilitate TPRT [15, 130]. Mutations exist in ORF1p that do not disrupt the high-affinity interaction of the protein with RNA but nevertheless abolish retrotransposition [129, 130, 214], consistent with a critical role for the nucleic acid chaperone activity of ORF1p in L1 replication.

Nucleic acid chaperones paradoxically promote both annealing of complementary single-stranded and melting of double-stranded nucleic acids. A consequence of these two fundamental activities is rearrangement of nucleic acid secondary structure, a property which is essential in a variety of biological processes including retroviral replication [112, 215],

![Figure 5.1](image_url). Mutations to conserved ORF1p residues in the murine ORF1p sequence perturb retrotransposition. (A) Schematic of ORF1p with single amino acid substitutions studied here (bold) or shown previously to affect nucleic acid chaperone activity [129, 130]. (B) Average percent (± SD, N = 3) eGFP-expressing cells 6 days after transfection with antisense intron-containing eGFP reporter in wt or ORF1p mutant L1.
functional splicing of the T4 phage thymidylate synthase gene [216], and proper folding of a yeast tRNA anticodon [217].

Single molecule stretching is a sensitive assay of nucleic acid chaperone activity. When single dsDNA molecules are stretched, the applied force facilitates melting of the DNA. When the DNA molecule is relaxed, DNA reannealing occurs. By measuring DNA stretching and relaxation in the presence of nucleic acid chaperone proteins, the extent to which the proteins facilitate DNA melting and annealing, important components of nucleic acid chaperone activity, can be determined [159, 194].

Here, we investigate the effect of six amino acid substitutions in ORF1p on L1 retrotransposition, expression, protein stability, RNA binding, the cooperativity of DNA melting, and the kinetics of ORF1p-mediated annealing and strand exchange. We present a novel analysis of the latter data, and use it to quantify the balance between these two opposing components of nucleic acid chaperone activity by comparing wild-type (wt) to mutant proteins. Taken together, these data demonstrate that L1 retrotransposition is exquisitely sensitive to changes in the biophysical properties of ORF1p that determine its nucleic acid chaperone activity.

5.2 Results

Previous studies demonstrated that retrotransposition of mouse L1 was reduced or abolished by the substitutions D159H and R297K in ORF1p [129, 130]. These residues lie in the coiled-coil and CTD, respectively, far apart in the primary sequence of ORF1p (Figure 5.1A). To further probe the functional requirements for ORF1p in L1 retrotransposition we substituted alanine for three conserved residues in the RRM and CTD domains of the protein in the context of an otherwise active L1: R238A, R284A or Y318A (Figure 5.1B). All three of these mutations rendered L1 inactive for retrotransposition in a cultured cell assay. Replacing the alanine to make the variants R284K and Y318F nearly or completely restored retrotransposition, respectively, but R238K remained inactive (Figure 5.1B). Retrotransposition detected by eGFP fluorescence using flow cytometry was confirmed by PCR amplification of the spliced eGFP from genomic DNA (data not shown).

![Figure 5.2](image_url). Effect of ORF1p mutations on steady-state levels of L1 RNA and ORF1p in cells transfected with retrotransposition assay vectors. (A) L1 RNA normalized to 18S rRNA measured by qRT-PCR on days 1-5 following transfection of 143B cells, average (± SD, N = 3). (B) Western blots of ORF1p. Top panel, triplicate wt extracts days 1-4 post-transfection as indicated; ORF1p is not detectable on days 5 and 6 (data not shown). Right arrow points to ORF1p, left arrow to signal obtained from 2.5 ng his-tagged recombinant ORF1p (rORF1p); (-) untransfected control cell extract. Middle panel shows hnRNP-Q detected on the same blot. Below are triplicate lanes containing protein extracts on day 2 post-transfection with the indicated ORF1

53
Single residue mutations may decrease protein stability [218, 219]. Mutations in ORF1p would likely interfere with L1 retrotransposition if the protein became unstable or unable to interact with L1 RNA, which might in turn cause instability of L1 RNA. Therefore the steady-state levels of L1 RNA and ORF1p were measured daily from cells transfected with wt and mutant L1 constructs. Based on qRT-PCR, L1 RNA was not reduced below wt levels in any of the inactive mutants. Moreover, differences in abundance of L1 RNA did not distinguish among active and inactive elements (Figure 5.2A). The highest steady-state levels of ORF1p detected in the lysates by western blotting were consistently observed two days post-transfection. Small fold changes of ORF1p abundance among mutants again failed to distinguish between active and inactive ORF1p variants (Figure 5.2B). These data indicate the defects in retrotransposition must be due to something other than instability or loss of L1 RNA or ORF1p.

Mutant proteins were purified (Figure 5.3A) in order to study their biochemical properties. The final purification step is size exclusion chromatography. All of the mutant proteins eluted in the same fraction as wt ORF1p, indicating that they share its homotrimeric asymmetric dumbbell-shaped structure [128].

Mutant proteins were characterized by their differential absorption of circularly-polarized light. All mutant and wt ORF1p produced absorption patterns consistent with a generally helical molecule (data not shown). If local secondary structure were perturbed by some of these mutations, it was insufficient to affect the overall absorption characteristics of the molecule. Additionally, the effect of increasing temperature on elliptical absorption at 222 nm revealed that the proteins denatured between 38 and 46 °C (Figure 5.3B). Although R238K and Y318F were more temperature sensitive than wt ORF1p, this property was not correlated with retrotransposition competence.

A nitrocellulose filter binding assay was used to assess whether any of the ORF1p mutants were compromised for RNA binding. As shown in Figure 5.4, the relative affinities were: Y318F ≥ Y318A ≥ wt ≥ R284K ≥ R284A >> R238K ≥ R238A. Because of experimental variation, however, only the two substitutions at R238 were significantly different than wt, each exhibiting an approximately 10-fold reduction in their apparent affinity for RNA. It is likely that the retrotransposition defects in R238A and R238K are due to their reduced affinity for RNA [214]; however, altered RNA binding does not explain the loss of retrotransposition in R284A or Y318A. The inability of these two mutants to retrotranspose cannot be explained by loss of L1 RNA or ORF1 protein, a failure to trimerize, gross perturbation of
secondary structure, decreased thermal stability or altered $K_d^{\text{app}}$ for RNA. Thus, some other property of ORF1p must be affected by these mutations.

Single molecule stretching experiments were used to assess the nucleic acid chaperone activities of the paired mutants at R284 and Y318. Figure 5.5A shows a stretching and relaxation cycle for bacteriophage λ DNA in the absence of protein and in the presence of wt ORF1p using optical tweezers. In the absence of protein, very little force is required to stretch the DNA to its full contour length. Near the contour length the force increases dramatically, reflecting the elasticity of the double helix. The approximately constant force plateau of about 60 pN represents a cooperative transition from dsDNA to ssDNA, or a force-induced melting transition [25, 37, 60, 220]. In addition, the force-extension curve in the absence of protein was almost completely reversible, showing little hysteresis, or disagreement between stretching and relaxation. The amount of hysteresis is characterized by the area between the stretching and relaxation curves. In the presence of wt ORF1p, the transition was altered significantly, as also shown in Figure 5.5A. The DNA force-induced melting plateau was sloped, indicating that the transition cooperativity was significantly reduced. A reduction in melting cooperativity indicates that DNA melting can be more easily initiated and smaller regions of DNA can be melted at once to facilitate DNA rearrangements. This change in shape of the transition was quantified by measuring the transition width, as described in Figure 5.5A.

A greater transition width indicates a less cooperative transition. Therefore, when the transition width is greater DNA will more easily undergo conformational rearrangements, as expected for a nucleic acid chaperone protein [105, 117]. The amount of hysteresis observed in the presence of wt ORF1p is greater than that observed for the nucleic acid chaperone HIV-1 NC [194], but significantly less than that observed for the slower nucleic acid chaperone protein HTLV-1 NC [113] and SSBs such as T4 gene 32 protein [45]. Thus, wt ORF1p exhibits moderately rapid kinetics, intermediate between other nucleic acid chaperones.

While Y318F was similar to wt ORF1p, Y318A had markedly less effect on DNA stretching (average $\Delta F = 14.7, 17.8, \text{ and } 6.9$, respectively, Figure 5.5B, D), consistent with the relative retrotransposition results for this mutant pair. Likewise, R284A, which was also defective for retrotransposition, had the least effect on DNA stretching while R284K strongly altered the DNA stretching curve (average $\Delta F = 2.1$ and 9.7, respectively, Figure 5.5C, D). Thus, as observed with other mutations in ORF1p [129, 130], inactivation of L1 retrotransposition in Y318A and R284A was correlated with loss of nucleic acid chaperone activity as measured by single molecule stretching. Similarly, mutations that restored retrotransposition also restored chaperone activity measured by DNA stretching.
Figure 5.5. Single molecule stretching. Typical force-extension (solid) and relaxation (dashed) curves of DNA only (black) and in the presence of the indicated ORF1p variants; blue lines are used for retrotransposition active variants and red lines for retrotransposition inactive variants. (A) wt, $\Delta F = 17.8 \pm 2.2$ pN; red lines illustrate the calculation of the transition width, $\delta F$. (B) R284A, $\Delta F = 2.1 \pm 0.5$ pN and R284K, $\Delta F = 9.7 \pm 1.2$ pN. (C) Y318A, $\Delta F = 6.9 \pm 2.0$ and Y318F, $\Delta F = 14.7 \pm 1.2$ pN. (D) Change in the transition width of DNA force-induced melting in the presence of wt and variant ORF1 proteins, where we have subtracted the value of the transition width in the absence of protein ($\Delta F = \Delta F - \delta F_0$), with $\delta F = 3.66 \pm 0.16$ pN [111]. Error bars are standard error for at least three measurements. Significant differences among the five ORF1p variants were found by ANOVA ($p < 0.001$); the active and inactive pairs at each site are different, Y318F $\neq$ Y318A ($p = 0.036$), R284K $\neq$ R284A ($p = 0.017$), but the active R284K is not distinguishable from the inactive Y318A ($p = 0.658$).

Although the results from single molecule DNA stretching for the paired mutations at each site correlated well with retrotransposition, the single molecule data were not able to distinguish within uncertainty between two mutants that clearly differed in retrotransposition, R284K and Y318F (Figure 5.5D). This result may be partially explained by the fact that R284K is less active than wt and Y318F (Figure 5.1B); however, our understanding of the relationship between the nucleic acid chaperone activity of ORF1p and L1 retrotransposition activity was clearly incomplete.
Figure 5.6. Annealing of oligonucleotides with increasing ORF1p. The fraction of oligonucleotide converted to duplex was determined by phosphorimage analysis of dried gels (see methods). (A) 1.0 nM complementary oligonucleotides were incubated with increasing concentrations of ORF1p trimer at 21 °C for 2.5 min. The data fit to a line defined by the equation:

\[
   f(c) = \frac{c/c_i}{1 + (c/c_i)}
\]

where \( f \) is the fraction labeled oligo incorporated into duplex as a function, \( c \) is the ORF1p trimer concentration in nM, and \( c_i = 1.15 \) nM, the ORF1p concentration at the inflection point of the curve, when the reaction is proceeding at half-maximum velocity. (B) Annealing timecourse for 1nM complementary oligonucleotides in the presence of 0 (open diamonds), 0.1 (circles), 1.0 (squares) and 10.0 nM (triangles) ORF1p.

Two aspects of nucleic acid chaperone function [221], the kinetics of DNA annealing and melting, were then examined in detail. To accomplish this, we developed assays to measure the magnitude of the energetic barriers to annealing and melting of complementary DNA oligonucleotides in the presence of ORF1p. First, complementary DNA oligonucleotides were incubated for 2.5 min at 21 °C in the presence of increasing concentrations of ORF1p trimer to determine the mid-point and saturating nt:ORF1p ratio for annealing under these conditions. The annealing rate saturated when the concentration of protein to nucleic acid was approximately 9:1, and the half-maximal point was 1.15 nM (Figure 5.6). Thus, all further experiments were performed with equimolar protein and oligonucleotides.

The effect of wt and mutant ORF1p on DNA annealing was determined at varying temperatures. The reaction can be expressed as:

\[
   S_1 + S_2 \xleftrightarrow{k_{\text{anneal}}} S_1S_2
\]

Here \( S_1 \) and \( S_2 \) are the ssDNA strands that anneal to form duplex \( S_1S_2 \) with total rate \( k_{\text{anneal}} \). The fraction of labeled oligonucleotide incorporated into duplex was plotted against time and these points were fitted to a capped, inverse exponential decay (Figure 5.7A):

\[
   f_t(t) = f^\infty \left(1 - e^{-k_{\text{anneal}}t}\right)
\]

where \( f_t \) is the fraction of labeled oligonucleotide incorporated into duplex as a function of time \( f^\infty \) is the equilibrium fraction of annealed oligonucleotide, \( t \) is the reaction time in seconds and \( k_{\text{anneal}} \) is the reaction rate constant in inverse seconds.
To find the temperature dependence of the reaction rate, the natural logarithm of the resulting rate coefficient was then plotted against the inverse temperature according to transition state theory (Figure 5.7B):

\[
\ln(k) = \ln\left(\frac{k_B T}{h}\right) + \frac{\Delta S_a}{R} - \frac{\Delta H_a}{R}\left(1 - \frac{1}{T}\right)
\]

where \( R \) is the ideal gas constant 8.31 J·mol\(^{-1}\)·K\(^{-1}\), \( T \) is reaction temperature in Kelvin, \( k_B \) is Boltzmann’s constant, \( h \) is Planck’s constant, \( \Delta S_a \) is the entropic contribution to the activation energy for annealing in J mol\(^{-1}\) K\(^{-1}\), and \( \Delta H_a \) is the enthalpic contribution to the activation energy for annealing in J mol\(^{-1}\).

A line was fitted to these points, allowing us to determine the enthalpic and entropic activation energies for this process. Strand exchange assays were used to determine the energetic barrier to melting and subsequent strand exchange in order to probe the other fundamental property of nucleic acid chaperones. A complementary DNA duplex, one strand labeled with \(^{32}\)P, was incubated with a 10-fold excess of competing, unlabeled oligonucleotide and equimolar ORF1p, again varying only temperature. The excess of unlabeled complementary oligonucleotide assured that the displaced \(^{32}\)P-labeled strand remained displaced for quantification. The time dependence of the reaction could only be fit to two exponentials, and was therefore biphasic:

\[
S_1 + S_2S_3 \rightleftharpoons (S_1S_2S_3)^* \rightleftharpoons S_1S_3 + S_2
\]

here \( S_1 \) is the unlabeled ssDNA strand and \( S_2S_3 \) is the duplex that contains the labeled ssDNA strand as well as one unlabeled strand. The intermediate complex \((S_1S_2S_3)^*\) will be formed when an unlabeled strand encounters a partially melted duplex. This represents the first step in the reaction, which occurs with the faster rate \( k_1 \). When the final duplex \( S_1S_3 \) is formed, \( S_2 \) must be dissociated with rate \( k_2 \). Based on this model, the data points were fit to a two-factor exponential decay (Figure 5.8A):

\[
f_r(t) = f_r^\infty + \left[1 - f_r^\infty\right] \left[P(e^{-k_1 t}) + (1 - P)(e^{-k_2 t})\right]
\]

where \( P \) is the fractional contribution to the overall rate from the first rate and \( k_1 \) and \( k_2 \) are the first and second rates, respectively.
Based on previous observations of a biphasic reaction for DNA and RNA hairpin annealing [112, 221], the fast rate was assumed to be the primary DNA melting step required for initial annealing of the complex. This hypothesis is also supported by the observation that the fast rate is similar to that observed for other reported chaperone-mediated DNA-duplex melting reactions [112], whereas the rate constant of the slow phase was 2-3 orders of magnitude slower than that observed for other nucleic acid chaperones. We will show later that the analysis of the fast rates correlate with mutations in ORF1p, further supporting our assumption that the fast rate involves DNA melting needed to initiate strand transfer, a process facilitated by nucleic acid chaperones. In contrast, the slow rates did not depend significantly on the presence of ORF1p. The temperature dependence of the fast rate was plotted and fit to the transition state model as before (Figure 5.8B).

The enthalpic and entropic contributions to the activation energy for product formation were quantified from Equation 5.2 (Table 5.1) for both annealing and exchange and these terms were combined into a total free energy of activation according to:

$$
\delta \Delta G_a = \delta \Delta H_a - T \delta \Delta S_a
$$

where $\delta \Delta H_a$ is the difference in enthalpic contributions between a particular mutant and the no protein condition, $\delta \Delta S_a$ is the difference between entropic contributions between a mutant and no protein, $T$ is the temperature in K (here set to 310, or 37 °C), and $\delta \Delta G_a$ is the total difference in the free energy of activation to product formation in the presence of ORF1p.

All mutant and wt ORF1 proteins catalyzed both annealing and strand displacement, although to differing magnitudes. The activation energy for annealing was reduced by 400-900 J mol$^{-1}$ bp$^{-1}$ compared to reactions without protein, and the energetic barrier to strand displacement was reduced by 9-15 kJ mol$^{-1}$ bp$^{-1}$.
Table 5.1. Slopes and y-intercepts of lines fitted to points plotted according to transition state theory. Slopes of lines fitted to points as graphed from Equation 5.2 express the temperature dependence of the reaction rate; more negative values denote a more temperature-dependent reaction and thus a higher enthalpic barrier to product formation. Intercepts are used according to Equation 5.2 to determine the relative magnitude of entropic contributions to energetic barriers. Values are expressed as ± standard deviation and (standard error). The points on these lines were derived for each mutant from experimental data collected as in Figures 5.7-5.8.

The change in free energy of melting was plotted as a function of that of annealing (Figure 5.9). This plot revealed an optimum magnitude for ORF1p-mediated annealing and strand-displacement catalytic activities in the retrotransposition-competent elements. All of the tested ORF1p variants exhibited an enhanced reduction in the energetic barrier to annealing, or improved catalysis of annealing, compared to wt independent of retrotransposition activity. However, the retrotransposition-active mutants have both very similar reductions in the barrier to melting and annealing. In contrast, R284A shows strongly enhanced annealing and reduced melting. This strongly enhanced annealing is unable to overcome the reduction in the protein’s ability to melt nucleic acids, which is clearly required for chaperone activity. Perhaps most interesting is the result for Y318A, which shows strongly enhanced melting and annealing, yet does not facilitate retrotransposition. This demonstrates that optimum melting and annealing is necessary for retrotransposition. This optimum melting and annealing is restored by substituting to form Y318F, which shows melting and annealing kinetics in these studies that are similar to wt.

5.3 Discussion

ORF1p from mouse L1 is a robust nucleic acid chaperone protein that is required for retrotransposition. Previously, substitutions that compromise L1 retrotransposition were mapped to single amino acids in either the coiled-coil [130] or the CTD [129] domains (D159H and R297K in Figure 5.1A, respectively). Both of these substitutions affected nucleic acid chaperone activity and not RNA binding affinity.

Here, an additional three highly-conserved amino acids, two (R238, R284) in the recently-described RRM domain [207] and the third, Y318, in the CTD [208],
were replaced with alanine, and also found to inactivate L1 retrotransposition. Two of these three new alanine mutations were rescued by substitution of the appropriate amino acid to restore the basic (R284K) or aromatic (Y318F) character of the original arginine or tyrosine, whereas the third (R238K) was not (Figure 5.1). Retrotransposition deficits were independent of changes observed in steady-state levels of L1 RNA or ORF1p, or aberrant ORF1p structure (Figures 5.2 and 5.3B). Both RNA binding and nucleic acid chaperone activities of ORF1p appear to be critical for L1 retrotransposition because R238A and R238K compromised high-affinity RNA binding (Figure 5.4), and R284A and Y318A compromised nucleic acid chaperone activity based on reduced cooperativity of DNA melting (Figure 5.5), and poorly optimized annealing and strand displacement catalysis (Figure 5.9).

Residues R238 and R284 lie in the RRM domain; the structure of the corresponding fragment from human L1 ORF1p has been solved by x-ray crystallography, revealing that R202 (homologous to R238 in mouse L1) forms a salt bridge together with E169. Disruption of the salt bridge is expected to destabilize ORF1p [207]. In the case of mouse L1, this salt bridge does not appear to be important for ORF1p stability because eliminating it with R238A did not alter the thermal melting behavior compared to wt (Figure 5.3B) or reduce its abundance in cells (Figure 5.2B). Moreover, the salt bridge partner of R202, E169 (D204 in mouse), is rotated to the outside of ORF1p in the modeled structure; thus it is unlikely that a salt bridge can form between these two residues in the mouse protein. Interestingly, R238K fails to restore the RNA affinity of ORF1p. Taken together, the available data indicate that R238 is important for interactions between ORF1p and nucleic acids, although the interactions involve more than simple electrostatics.

The second residue of the RRM domain studied here, R284, has not been examined previously. Based upon the crystal structure of the human homolog, R284 is on the surface of the RRM domain in the vicinity of several other arginine residues that form electrostatic interactions with RNA [207]. The interpretation that this residue is important for electrostatic interactions is supported by the observed inactivation of retrotransposition by R284A and restoration of activity by R284K. We were unable to measure a significant reduction in RNA affinity in R284A, although a reproducible trend in this direction was observed. The nucleic acid chaperone activity of R284A is, however, significantly altered by an imbalance of annealing and displacement activities because of a markedly decreased ability to catalyze the displacement reaction. In contrast, R284K restores the wt balance between annealing and strand displacement, suggesting that the electrostatic interaction between R284 and nucleic acids is important for the melting component of the nucleic acid chaperone activity.

The final residue examined, Y318 in the CTD, is conserved in placental mammals, marsupials and frogs. The homologous site, however, is substituted with phenylalanine in swimmer, a LINE element in Japanese medaka [207], consistent with the observed retrotransposition with Y318F here. Y318F also restored the balance between the melting and annealing activities of ORF1p that was lost in Y318A. Because annealing catalysis was not significantly different between the alanine and phenylalanine variants at residue 318, the biochemical problem with Y318A appears to be its excessive basepair melting catalysis. This residue is likely involved in hydrophobic or steric interactions because the terminal hydroxyl group on tyrosine is dispensable.

Based upon the results of earlier studies of wt and mutant ORF1 proteins, we proposed that at least some of the interactions involved in high affinity binding of ORF1p to RNA are distinct from those involved in the nucleic acid chaperone activity [129, 130], that is, these reactions involve distinct sites on the protein. The present results provide further support for this hypothesis. If the melting activity in the strand displacement assay is functionally equivalent to the RNA binding affinity, then the paired variants
at residues 284 and 318 should have been indistinguishable in the strand displacement assay as they were in the RNA binding assay, but they were not.

A three-dimensional structure of the intact trimer with nucleic acids will be necessary to fully understand the relationship between the structure of ORF1p and its RNA binding and nucleic acid chaperone activities. Nevertheless, it seems unlikely that all of the residues that affect the chaperone activity of ORF1p are closely juxtaposed in the tertiary structure of the protein, because they are so widely distributed along its primary structure. Instead, we predict that residues responsible for chaperone activity are distributed in the folded structure as well as the primary sequence of this protein.

When taken together and compared with retrotransposition experiments, the in vitro single molecule and solution experiments demonstrate the importance of the finely balanced nucleic acid chaperone capabilities of ORF1p for biological activity. They also show that the nucleic acid chaperone activity of ORF1p requires optimal interactions with ssDNA and dsDNA that result in a reduction in melting cooperativity, as well as an optimal enhancement of strand annealing and strand displacement capabilities. This enhancement of annealing and displacement is likely a result of the combination of the ability of ORF1p to facilitate strand attraction as well as nucleic acid unwinding. Specific single amino acid substitutions in ORF1p inhibit this chaperone activity by interfering with at least one of these capabilities, as demonstrated by the properties of the mutations at R284 and Y318. A mutation that strongly enhanced both melting and annealing (Y318A) inhibited retrotransposition, but a compensatory mutation that reduced the melting and annealing capabilities of ORF1p and restored these properties to wt levels also restored retrotransposition. Similarly, a mutation that enhanced annealing but significantly reduced the ability of ORF1p to facilitate melting also inhibited retrotransposition. A compensatory mutation that restored the kinetics to wt values also restored retrotransposition activity (Figure 5.9).

Other groups have published kinetic data on DNA or RNA annealing [112, 113, 221-223], DNA melting as an intermediate step to annealing [112], steady-state duplex destabilization, and chaperone effects on force-induced DNA melting [17, 88, 111]. Although it has been suggested that the melting and annealing activities of nucleic acid chaperones are in a tightly constrained balance [115] this is, to our knowledge, the first work to quantitatively determine an optimum kinetics of strand displacement and annealing that can be destroyed and restored with specific point mutations.

Finally, it is noteworthy that all of the ORF1p mutations that retained wt RNA affinity were nevertheless nucleic acid chaperone proteins– all ORF1p variants tested accelerated both basepair annealing and melting, and presumably allow for the rearrangement of secondary structure as a consequence, regardless of their retrotransposition competence. The property that discriminates between biologically active and inactive variants of ORF1p, however, is the balanced ratio of the energetic barriers to annealing and melting. This finding implies that L1 retrotransposition depends upon an ORF1p that provides the two optimized energetic reductions in specific kinetic properties, consistent with observations that dispersed mutations in ORF1p could disrupt this finely-tuned balance and inactivate retrotransposition. These quantities must have a basis in the specific nucleic acid rearrangements that are required for L1 retrotransposition. It may be a general feature of nucleic acid chaperone proteins to maintain specifically tuned energetic reductions to annealing and melting that satisfy their biological function, as determined here for L1 ORF1p.

5.4 Materials and Methods

5.4.1 Autonomous retrotransposition assay
Mutations were introduced into L1 ORF1 by site-directed mutagenesis and assayed for effects on retrotransposition in 143B cells using eGFP expression as a marker of retrotransposition as previously described [129, 206].

5.4.2 DNA, RNA and protein analysis

Whole cell lysates (WCL) were prepared from L1 transfected 143B cells every 24 hours post-transfection as described [130]. Protein was quantified (Bradford assay, Bio-Rad) and ORF1p was detected by western blot in 20 μg of WCL as described [130]. hnRNP-Q was detected on the same blots using a mouse monoclonal hnRNP-Q antibody (Abcam). Blot images were captured on a GE Typhoon 9400. L1 RNA was quantified by quantitative real time PCR using primers with a probe specific to the ORF2 region. Total RNA was isolated from WCL using TRIzol-LS (Invitrogen) and quantified (Nanodrop, Thermo Scientific). RNA (2 μg) was treated with RQ1 DNase (Promega) and cDNA was synthesized from 1 μg using a High-Capacity cDNA Reverse Transcription kit (Applied Biosystems). Complementary DNA (2 μL) was used for quantitative PCR using TaqMan probe assay with 5′-CTCAGAATGAAAGGCTGGAAAAC as the forward primer, 5′-AGGATGGCTACTCCTGCTTGTT as the reverse primer, and 5′-FAM-CCAAGCATAATGATGAAG-NFQMGB as the probe. A standard curve was generated using 10-fold dilutions of L1-containing plasmid. 18S rRNA was quantified using the TaqMan Gene Expression assay for 18S rRNA (Applied Biosystems). The quantities of L1 and 18S rRNA in the cDNAs were calculated from their respective standard curves using the absolute standard curve method (Applied Biosystems) and then L1 RNA values were normalized to 18S rRNA values.

5.4.3 Recombinant ORF1p purification

ORF1p coding sequences with single residue mutations were cloned into a modified pFastBac bacmid and used for recombinant expression in SF9 cells as before [15], according to the Bac-to-Bac baculovirus expression system protocol (Invitrogen). We replaced the 6His to ORF1 AUG sequence of pFastBac with the corresponding sequence from pBluBac to facilitate ORF1p purification as described previously [128], except that the ammonium sulfate precipitation was 27% NH₄(SO₄)₂ followed by 73% NH₄(SO₄)₂.

5.4.4 Circular dichroism and thermal melting

CD and melt scans were performed on a Jasco 815 spectropolarimeter with purified ORF1p between 2.1 and 4.2 μM trimer concentration in 50 mM phosphate, pH 7.6; 250 mM NaCl and 0.1 mM EDTA, and data were collected as the observed ellipticity in mdeg. Thermal stability was determined by observing the differential absorption at 222 nm while the temperature was raised at a rate of 2 °C per minute, from 4 °C to 70 °C. Data were normalized to the 4 °C data point.

5.4.5 RNA filter binding

Assays were performed in 250 mM NaCl and analyzed as described previously [209]. Differences in $K_{d}$ of 2-3 fold were not significant because of experimental variation [129].

5.4.6 Single molecule DNA stretching

A dual beam optical tweezers instrument was used to stretch 5′ biotin-end-labeled bacteriophage λ DNA as described elsewhere in detail [25, 37, 105]. Briefly, the DNA was captured on streptavidin-coated polystyrene beads (Bangs Labs), then stretched and relaxed after rinsing out other DNA molecules. All stretching experiments were performed at a pulling rate of 100 nm/s in 10 mM HEPES; 100 mM Na⁺; pH 7.5 buffer. Buffer was exchanged for specific protein solutions to probe protein-DNA interactions. To quantify the effects of protein on the DNA stretching curves, the transition width was determined from the transition slope at its midpoint as described previously [111]. Significant differences among protein
groups were assessed by one way ANOVA followed by Tukey’s pairwise comparisons using KaleidaGraph 4.03 (Synergy Software).
5.4.7 Annealing and displacement assays

For annealing, ORF1p and two complementary, 48-nt DNA oligonucleotides, one end-labeled with $^{32}$P, were incubated at equimolar concentrations in 30 μL reaction buffer (20 mM HEPES pH 7.6; 25 mM NaCl; 1 mM EDTA pH 8.0; 1 mM MgCl$_2$; 1 mM dithiothreitol and 0.1% Triton-X 100 w/v). Reactions with ORF1p were incubated at 1, 15, 22 and 37 °C; those without protein were incubated at 15, 22, 37 and 42 °C. Aliquots (3 μL) were taken at the noted times, mixed with 3 μL stop buffer (40 mM HEPES pH 7.6; 0.4 mg/mL tRNA; 0.2% sodium dodecyl sulfate w/v; 10 mM EDTA pH 8.0; 3% Ficoll-400 w/v; 0.25% bromophenol blue w/v and 0.25% xylene cyanol w/v) and placed on ice. Samples were fractionated through a 15%, 19:1 bis:acrylamide gel at 4 °C in 0.5X TBE, the gel dried and exposed to a storage phosphor screen. Images were collected on a Typhoon 9400 and bands containing single and double-stranded oligonucleotides were quantified with ImageQuaNT 5.2 (GE). KaleidaGraph was used for curve fitting and graphical representation of data.

Full duplex exchange experiments to quantify melting catalysis facilitating strand exchange were performed by mixing 1 nM 38-bp dsDNA, 5′ end-labeled with $^{32}$P on one of the strands, with 10 nM of the identical unlabeled single-stranded oligonucleotide competitor and 10 nM ORF1p trimer in reaction buffer. This strand exchange assay differs from a strand-exchange assay used previously to study ORF1p because the large thermodynamic advantage that drove the previous reactions (replacement of a short [15] or imperfectly matched [129] strand in a preformed duplex with the perfect complement, provided at 50x molar excess) has been removed. Reactions were incubated with ORF1p at 22, 27, 32 and 37 °C, or 37, 40, 42, 45, 47 and 50 °C for reactions without protein, and then treated as described for the annealing reactions.

Acknowledgements

We thank D. Branciforte, D. Bushman, and J. Cummiskey for constructs, and Drs. I. Rouzina, D. Bain, and K. Maluff for valuable discussion. This work was supported by NIH [R01GM40367] (SLM) [R01GM75965] (MCW), and NSF [MCB-0744456] (MCW). KRC was supported by the NSF IGERT Program [DGE-0504331].
Oligomerization transforms human APOBEC3G from a fast enzyme to a slow binding protein

Abstract

Human A3G inhibits HIV-1 replication in the absence of the viral protein Vif via cytidine deamination of viral ssDNA, as well as via a distinct deamination-independent mechanism. Approximately seven A3G proteins are packaged per virion, yet there are as many as 1000 deamination sites on viral ssDNA. Therefore, efficient deamination requires rapid on-off binding kinetics. In contrast, a relatively slow dissociation rate is required for the proposed deaminase-independent roadblock mechanism, whereby A3G binds the viral template strand and blocks reverse transcriptase-catalyzed DNA elongation. To resolve this apparent contradiction, we developed a quantitative single molecule method, which showed that A3G initially binds ssDNA with fast on-off rates and subsequently converts to a slow binding mode. We propose that catalytically active monomers or dimers slowly oligomerize on the viral genome and inhibit reverse transcription. Oligomerization may represent a general mechanism by which proteins assume multiple functional roles.

This work is currently under consideration for publication, and has been adapted for this dissertation.
6.1 Introduction

The HIV-1 genome encodes several accessory proteins essential for viral replication, including virion infectivity factor (Vif) [224, 225]. HIV-1 Vif counteracts members of the human APOBEC3 (A3) family of deoxycytidine deaminases [123, 226-233], which convert deoxycytidine bases in ssDNA to deoxyuridine [13, 234-239]. These proteins function as cellular restriction factors and are part of the innate immune response to viral pathogens [13, 237, 240]. Although all of the A3 family members inhibit replication of retroviruses and retroelements [12, 13], A3G [226] is the most potent inhibitor of HIV-1 replication [11-13], reducing viral infectivity by several orders of magnitude in the absence of Vif [123].

Although A3G is the most studied of all the APOBEC proteins, the molecular mechanism for A3G-mediated HIV-1 restriction is still not fully understood. A3G deamination of minus-strand viral DNA formed during reverse transcription leads to G to A hypermutation in the plus-strand [235, 236, 241], which effectively impairs viral replication. However, A3G catalytic mutants retain antiviral activity [242-245], suggesting that a deaminase-independent mechanism is also involved [11, 13, 246]. Moreover, A3G inhibits hepatitis B virus replication without G to A hypermutation [247], and A3A inhibition of LINE-1 and Alu retrotransposition [248-252] as well as parovirus replication [250, 253] is not dependent on A3A’s editing activity. Further evidence for a non-editing mechanism is based on the reduction of minus-strand viral DNA levels in HIV-1 particles during endogenous reverse transcription [254], inhibition of RT-catalyzed viral DNA elongation in vitro by a catalytic A3G mutant [122], inhibition of strand transfer reactions in vitro and in cell-based assays [122, 255, 256], and A3G-induced inhibition of reverse transcription in viruses from human CD4+ T cells [257]. Therefore, the roadblock model, in which A3G binds the template strand and physically blocks viral DNA synthesis, has been proposed as a molecular mechanism for deaminase-independent inhibition [122].

Since only 7 (± 4) A3G molecules are incorporated into each vif-deficient virion [258], RT inhibition by an A3G roadblock requires a slow A3G off-rate from single-stranded nucleic acids. In contrast, these few A3G molecules must have fast on-off rates to deaminate up to 1000 sites in several minutes [239] using a rapid search mechanism on viral ssDNA [259, 260]. To resolve this apparent paradox, we hypothesize that A3G exhibits fast binding kinetics as a monomer or dimer in order to function as an efficient enzyme, and slow kinetics upon oligomerization in order to block RT from elongating viral DNA. To test this idea, we used optical tweezers to monitor A3G binding kinetics on a single DNA molecule.

6.2 Results

6.2.1 Single molecule measurements of A3G binding to ssDNA

For these studies, a single double-stranded λ-DNA molecule was tethered to two polystyrene beads, one held in an optical trap and another on a micropipette tip. As the fixed bead is gradually moved away from the optically-trapped bead, the force on the DNA is measured at each extension (Figure 6.1, solid black line). Force-induced melting occurs at a constant force of 61.0 ± 0.5 pN, generating ssDNA. (It is well-established that overstretching is force-induced melting in the presence of ssDNA binding proteins, which are the conditions relevant in this work [159].) At the end of this transition, the bead movement reverses direction and DNA tension is gradually released (Figure 6.1, dashed black line). The minimal hysteresis, or mismatch between DNA extension and release, indicates that the ssDNA generated by force reanneals immediately into dsDNA during the return. At any given point along the transition, the molecule is a well-characterized combination of dsDNA and ssDNA (Equation 2.6) [159]. Pausing at fixed extension during the melting transition allows precise control of the fraction of ssDNA substrate available for protein binding.
The force-extension curve probes the length of the captured DNA molecule at a given force. At forces above 7 pN, ssDNA is longer than dsDNA. We exploit this force-dependent difference in length to measure the ssDNA binding properties of A3G. A3G-saturated ssDNA, obtained at high protein concentration, is longer than dsDNA and shorter than ssDNA (Figure 6.1). Therefore, A3G bound to ssDNA increases the molecule in length by $\Delta x_a$ below the melting transition, and decreases the length by $\Delta x_b$ above the melting transition.

In the presence of 50 nM A3G, the extension curve follows the DNA-only curve before the melting transition, reflecting no measurable binding to dsDNA (Figure 6.2A, solid line). A3G only binds after force-induced melting generates ssDNA. Based on the observed hysteresis (Figure 6.2A, dashed line), most of the protein does not dissociate upon DNA release and prevents the two strands from fully reannealing. A3G-bound ssDNA is longer than dsDNA (Figure 6.1), so the change in length at a given force (Figure 6.2A, $\Delta x_t$) describes the total fraction of A3G bound to ssDNA ($f_{total}$, see Figure 6.5A).

A second stretch of the same molecule does not retrace the release curve, revealing that some fraction of the protein has dissociated during the 30 s incubation at zero force between stretch-release cycles (Figure 6.2B). As soon as any A3G dissociates below the force-induced melting transition, the two strands reanneal into dsDNA, which is shorter than A3G-bound ssDNA (Figure 6.2B, $\Delta x_b$). Therefore the second stretch reflects the fraction of A3G that remains bound ($f_{slow}$), which allows the fraction that dissociates quickly ($f_{fast}$) to be quantified as $f_{fast} = f_{total} - f_{slow}$.

A3G was exposed to ssDNA for 50 s in this experiment. However, A3G oligomerization observed in bulk experiments occurs on much longer timescales [259]. To measure slow binding, A3G was incubated for 250 s with ssDNA generated by force-induced melting (Figure 6.2C). The DNA release curve obtained after this incubation exhibits a length increase relative to the initial release curve (Figure 6.2C, $\Delta x_b$) because additional A3G binds during incubation. This effect increases at longer incubation times (Figure 6.2D), approaching the A3G-saturated ssDNA curve (shown in Figure 6.1). Fits to the DNA release curves at increasing incubation time yield $f_{total}(t)$, while the subsequent stretch (data not shown for clarity) quantifies $f_{slow}(t)$, and $f_{fast}(t)$ is the difference between the two. These measurements for 50 nM A3G are presented in Figure 6.3A.

![Figure 6.1. Typical extension (solid black) and return (dashed black) of a single DNA molecule. At 61.0 ± 0.5 pN, the molecule undergoes a force-induced melting transition [159] from dsDNA (green, Equation 2.4) to ssDNA (blue). A3G-saturated ssDNA (200 nM A3G, $t > 500$ s, data points fit to Equation 6.2, solid purple) is longer than dsDNA ($\Delta x_a$ below the melting transition) and shorter than ssDNA ($\Delta x_b$ above the melting transition). A3G-saturated ssDNA is significantly shorter than ssDNA only (blue, Equation 2.5), which suggests that A3G may wrap ssDNA upon binding.](image)
Figure 6.2. Single molecule method to measure fast and slow fractions of total A3G binding. (A) In the absence of protein (black), a single DNA molecule reanneals immediately upon release, exhibiting minimal hysteresis, or mismatch between the stretch (solid) and release (dashed) curves. In the presence of 50 nM A3G, the stretch curve follows the dsDNA-only curve (solid green), indicating no A3G binding to dsDNA. A3G binds some fraction of the exposed ssDNA generated by force-induced melting, and does not dissociate during 50 s of DNA release. The bound protein prohibits the DNA strands from reannealing, resulting in significant hysteresis (dashed green). For a given force (shown for 40 pN), there is a corresponding change in DNA length $\Delta x_t$ between A3G-free dsDNA (left arrow, drawing 1) and partially A3G-bound ssDNA (right arrow, drawing 2). This force-dependent change in length is a measure of A3G binding to ssDNA (see Figure 6.5). (B) The second stretch (solid blue) lies in between the first stretch and release curves, distinguishing the fraction of A3G that remains bound ($f_{\text{slow}}$) from the fraction that dissociated ($f_{\text{fast}}$) during the 30 s before the second stretch. The A3G that dissociates rapidly allows the strands to reanneal immediately into dsDNA (drawing 3), resulting in decrease in length $\Delta x_f$. (C) Pausing at fixed DNA extension after the melting transition and incubating ssDNA with 50 nM A3G results in additional binding (drawing 4), indicated by the corresponding increase in length $\Delta x_i$ measured during DNA release. (D) A3G binding increases with total exposure time to ssDNA (dashed lines).
6.2.2 Quantitative binding model

The slow binding component increases at the expense of the fast component, suggesting a two-step A3G-ssDNA reaction:

\[
A3G + \text{ssDNA} \xrightarrow{k_{c}} (A3G/\text{ssDNA})_{\text{fast}} \xrightarrow{k_{1}} (A3G/\text{ssDNA})_{\text{slow}}
\]

6.1

in which an initial bimolecular process leads to a fast complex that converts to a slow, more stable complex in the second unimolecular step.

Figures 6.3. Quantifying A3G binding reveals association and dissociation rates for fast and slow binding modes. (A) Total binding at 50 nM A3G \(f_{\text{total}}\) (red) separated into a fast fraction \(f_{\text{fast}}\) (blue) and slow fraction \(f_{\text{slow}}\) (green), as a function of ssDNA-A3G incubation time. Fits to the binding model (solid lines, Equations 6.6-6.8) yield observed rates \(k_{\text{fast}}\) and \(k_{\text{slow}}\), and the equilibrium fractions of fast and total binding, \(P_{\text{fast}}\) and \(P_{\text{total}}\) (see Table 6.1.). (B) Slow fraction bound as a function of time for five A3G concentrations. Solid lines are fits to Equation 6.7. Error bars (A, B) are standard error \((N \geq 3)\) for 50-200 nM A3G and propagated error for 10-20 nM A3G. (C) Fast rates \((k_{\text{fast}}, \text{blue data points})\) and their uncertainty obtained from fits to the binding model (Equations 6.6-6.8, shown in panel A for 50 nM A3G). The linear dependence on protein concentration (solid blue line, Equation 6.10) yields forward bimolecular rate constant \(k_{1}\) and backward rate \(k_{-1}\). The forward rate \(k_{C}\) (purple data points) and the backward rate \(k_{1}\) (red data points) were also calculated from the binding model (Equations 6.11 and 6.13, respectively). Linear fits (solid lines, Equations 6.12 and 6.14) yield consistent values of \(k_{1}\) and \(k_{-1}\). (D) Slow rates \((k_{\text{slow}}, \text{green data points})\) from fits to the binding model (shown in panel B) saturate at high A3G concentration. Fits to Equation 6.18 (solid green line, see Figure 6.6B) yield forward rate \(k_{2}\) and backward rate \(k_{-2}\). Separate calculations of \(k_{2}\) (purple) and \(k_{-2}\) (red) from the binding model (Equations 6.21 and 6.22) are also concentration-independent. Solid lines reflect weighted averages. Fit parameters and uncertainties (C, D) are reported in Table 6.2.

Binding rates were obtained from fits to this model (Equations 6.6-6.8) at five A3G concentrations (Figure 6.3B). (A3G precipitates at high concentrations [244], so experiments were calibrated using.
force-extension curves at known protein concentrations.) As expected, the observed fast rate $k_{\text{fast}}$ and the on rate $k_1 c$ are both linear with A3G concentration (Figure 6.3C). The bimolecular rate constant $k_1 = 1.5 (\pm 0.1) \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$ and off rate $k_1 = 1.2 (\pm 0.1) \times 10^2 \text{ s}^{-1}$ are consistent with single molecule FRET [261] and fluorescence spectroscopy [259] measurements, considering differences in solution conditions. The observed slow rate $k_{\text{slow}}$ saturates at high A3G concentration (Figure 6.3D), and both the on and off rates for the second, unimolecular step are concentration-independent ($k_2 = 6.7 (\pm 0.6) \times 10^{-3} \text{ s}^{-1}$ and $k_2 = 2.8 (\pm 0.5) \times 10^{-5} \text{ s}^{-1}$). Elementary reaction rates were obtained from the data in several different ways, and agreement of the resulting values (Table 6.2) supports the binding model.

6.3 Discussion

Here we develop a new single molecule method that allows us to precisely quantify two distinct types of A3G binding to ssDNA and characterize the conversion of a fast state into a slow state. These results demonstrate a binding mechanism in which monomers or dimers initially bind ssDNA and rapidly reach equilibrium ($1/k_{\text{fast}} = 24 \pm 1 \text{ s}$ at 200 nM), before slowly converting to oligomers ($1/k_{\text{slow}} = 206 \pm 20 \text{ s}$) (Figure 6.4A). Previous bulk solution experiments have established that A3G oligomerizes in the presence of single-stranded nucleic acids [9, 262], which inhibits efficient deaminase activity [262]. However, our results demonstrate for the first time that A3G oligomerization is the mechanism for deaminase-independent inhibition of viral replication.

**Figure 6.4.** Models for A3G oligomerization (A) in vitro and (B) in virio. (A) Initially monomers or dimers bind ssDNA with on rate $k_1 c$ and off rate $k_1$. These forward and backward rates are on similar timescales ($1/k_1 c = 33 \pm 1 \text{ s}$ at 200 nM A3G, and $1/k_1 = 85 \pm 5 \text{ s}$), so fast binding reaches equilibrium before the monomers or dimers convert to oligomers ($1/k_2 = 149 \pm 13 \text{ s}$) on ssDNA. Oligomer dissociation is significantly slower ($1/k_2 = 10 \pm 2 \text{ h}$) in vitro. (B) Inside the virus, A3G oligomerizes on the RNA genome, blocking minus-strand DNA synthesis by RT (panel 1). Once the oligomer dissociates, the monomers or dimers released bind ssDNA within a second, allowing rapid enzymatic activity (panel 2) until A3G oligomerizes on the ssDNA template in 150 s and blocks plus strand synthesis (panel 3).

In light of these quantitative results in vitro, we propose a model for A3G’s effects on HIV-1 replication. A3G oligomerizes on viral RNA when it is packaged inside the virion [9, 10], stalling RT during minus-strand synthesis. RT then pauses until the oligomer dissociates, or switches to the other RNA template strand, circumventing the A3G roadblock and leading to partial inhibition of reverse transcription [257]. Once the A3G oligomer dissociates during minus-strand synthesis, the monomers or dimers that are released have low affinity for the newly formed RNA-DNA duplex [239, 244]. As RNase H activity exposes the minus strand, all A3G molecules bind the ssDNA template within a second, as indicated by $1/k_2 = 0.7 \pm 0.1 \text{ s}$ when fast binding rates are extrapolated to the estimated $13 \pm 8 \mu$M A3G concentration in the virion [122]. A3G remains bound in a rapid sliding mode [259, 261] for $1/k_1 = 85 \pm 5 \text{ s}$, which allows high deamination rates [259, 260] until oligomerization forms a roadblock to plus-strand synthesis after $1/k_2 = 149 \pm 13 \text{ s}$ (Figure 6.4B).
The in virio model that we propose based on our measurements is consistent with the available data on A3G function. Recent cell-based experiments demonstrate that A3G uniformly blocks minus-strand synthesis in the absence of preferred RT termination sites along the viral template [257]. However, this does not necessarily indicate that A3G binds throughout the length of the genomic RNA. Oligomerization is nucleic acid sequence-independent and therefore we expect the roadblock to form at random sites along the viral genome, as observed in virio. Although A3G may bind RT in an RNA-independent interaction [263], this type of binding is equally probable for any of the 200 RT heterodimers present in the virion [8] and the probability that one of 7 ± 4 A3G molecules binds and inhibits the catalytically active RT molecule is therefore less than 5%. Thus, the reported A3G-RT interaction is unlikely to be primarily responsible for A3G-induced inhibition of reverse transcription.

The data we present resolve seemingly contradictory mechanisms for A3G inhibition of viral replication by demonstrating that A3G can function as both a fast deaminase and a slow nucleic acid binding protein. Regulation of enzymatic activity via protein oligomerization may be a general property of other APOBEC family members that inhibit replication of retroviruses and retrotransposons independent of deaminase activity. The single molecule method described here does not require labeling of DNA or protein and is optimal for measuring the biologically important process of slow protein oligomerization.

6.4 Materials and Methods

6.4.1 A3G preparation and purification

Recombinant, catalytically active A3G was expressed in a baculovirus expression system with an N-terminal glutathione S-transferase (GST) tag [244]. The initial steps in purification were performed essentially as described previously [244]. Briefly, cell lysates treated with DNase I and RNase A were prepared and the soluble fraction was bound to glutathione Sepharose High Performance resin (GE Healthcare) for 3 h at 4 °C. The resin was washed as described [244] and bound GST-A3G was eluted with glutathione buffer (50 mM Tris-HCl, pH 8.0, 1 M NaCl, 40 mM reduced glutathione, 10 μM ZnCl₂, 10% glycerol). The eluate was filtered through a Acrodisc 25 mm syringe filter with a 1.2 µm Versapor® membrane (Pall Corp.) and was then dialyzed overnight at 4 °C in buffer containing 20 mM Tris-HCl, pH 7.4, 1 M NaCl, 5 mM DTT, and 10% glycerol. Samples were stored at -80 °C prior to further purification. The partially purified GST-A3G was loaded onto a gel filtration column (GE Healthcare, Superdex 200 HiLoad 16/60) pre-equilibrated in a buffer containing 20 mM HEPES, pH 7.4, 200 mM NaCl, 5 mM DTT, and 10% glycerol. The protein was eluted with the same buffer and the peak corresponding to monomeric GST-A3G was collected and concentrated to an appropriate volume. The GST tag was removed using a Novagen Enterokinase Cleavage Capture Kit. After the cleavage reaction, enterokinase was removed using the capture resin provided in the kit and the GST tag was removed with glutathione Sepharose 4B (GE Healthcare). The final monomeric A3G was concentrated and the concentration was determined by Bradford assay and UV absorbance (ε₂₈₀ = 105,260 M⁻¹cm⁻¹). The purified A3G was stored at -80 °C.

Deaminase activity of purified A3G was measured using a gel-based uracil DNA glycosylase assay with the 40-nt ssDNA substrate labeled at its 5’ end with Alexa Fluor® 488 (Integrated DNA Technologies, Coralville, IA) instead of ³²P [244].

6.4.2 Single molecule experiments

Single molecule DNA stretching experiments were conducted with dual beam optical tweezers as previously described [159]. Biotin-labeled bacteriophage λ DNA (48.5 kbp, 16.5 μm contour length) was captured between two streptavidin-coated polystyrene beads (Bangs Labs), one fixed on a micropipette tip and one in the optical trap. The flow chamber was rinsed with buffer (50 mM Na⁺, 10 mM HEPES, pH
7.5), and the captured DNA molecule was then stretched and released to obtain a DNA-only force-extension curve (Figure 6.5A, black). End-to-end DNA extension is scaled by the total number of base pairs in the molecule to make force-extension curves independent of DNA length. However, the total time for a stretch-release cycle at a fixed pulling rate does depend on DNA length. Specifically, λ DNA increases 10 μm in length during the force-induced melting transition, so going through the transition takes 100 s at the fixed pulling rate of 100 nm/s. DNA release takes the same amount of time. The WLC (Equation 2.4) and FJC (Equation 2.5) models describe dsDNA (Figure 6.5A, green) and ssDNA (blue), respectively. A linear combination of these models describes the measured length \( b \) as a function of the fraction of ssDNA generated upon DNA extension \( f_{ss} \) (Equation 2.6) [159]. After three force-extension cycles of DNA alone, the buffer was replaced with protein dilutions ranging from 1 nM to 200 nM A3G.

After the initial stretch-release cycle in the presence of A3G, the DNA molecule was stretched again. Incubation was performed at fixed DNA extension above the melting transition (Figure 6.2) before the release curve was obtained. The following stretch-release cycle was a control experiment performed without incubation to ensure that the additional binding effect was no longer observed. A 30 s wait time was introduced between each DNA release and the subsequent stretch to ensure complete dissociation of the A3G component exhibiting fast binding kinetics.

Total A3G-ssDNA binding saturated at 200 nM A3G after incubation for 500 s or longer. The force versus extension curve \( F(b_{A3G}) \) for A3G-saturated ssDNA was phenomenologically described by a second-order polynomial:

\[
F(b_{A3G}) = Ab_{A3G}^2 + Bb_{A3G} + C \tag{6.2}
\]

or, in terms of extension \( b \) versus force \( F \):

\[
b_{A3G}(F) = -\frac{B + \sqrt{B^2 - 4AC}}{2A} \tag{6.3}
\]

The parameters of the fit to this polynomial are \( A = 1713, B = 1028, C = 155 \) (Figure 6.1). The A3G-saturated ssDNA curve was used to analyze fractional A3G binding. Fits above the melting force \( F_m = 61.0 \pm 0.5 \) pN, the force-extension curve in the presence of A3G was fit to a linear combination of the dsDNA curve \( b_{ds}(F) \) (Equation 2.4) and the A3G-saturated curve \( b_{A3G}(F) \) (Equation 6.3), shown in Figure 6.5A:

\[
b(F < F_m) = b_{ds}(1 - f_{total}) + b_{A3G}f_{total} \tag{6.4}
\]

where \( f_{total} \) is fraction of A3G-bound ssDNA. Fits above the melting transition were to a linear combination of \( b_{A3G}(F) \) (Equation 6.3) and the ssDNA curve \( b_{ss}(F) \) (Equation 2.5):

\[
b(F > F_m) = b_{A3G}f_{total} + b_{ss}(1 - f_{total}) \tag{6.5}
\]

Measurements at forces above the transition were used to confirm that the results of both independent methods agree within uncertainty (Figure 6.5B). Values above and below the melting transition were averaged (data points with propagated error bars not shown for clarity), and fraction bound \( f_{total} \) as a function of protein concentration \( c \) was fit to a simple DNA binding isotherm (Equation 4.1 where \( \theta \) is \( f_{total} \)) to obtain is equilibrium dissociation constant \( K_d \).
Figure 6.5. DNA force-extension curves in the presence of A3G are used to quantify A3G binding to ssDNA. (A) DNA release curve (brown circles, data also shown in Figure 6.2D) after stretching dsDNA in the presence of 50 nM A3G followed by 1000 s incubation at 0.53 nm/bp (indicated by the brown arrow). Below the melting transition, the force-extension curve is a linear combination of dsDNA (green, Equation 2.4) and A3G-saturated ssDNA (purple, Equation 6.3) because any ssDNA generated by force-induced melting that is not bound to A3G will immediately reanneal into dsDNA. The fit yields the fraction bound $f_{\text{total}} = 0.90 \pm 0.03$ (red, Equation 6.4). Uncertainty from the fits is small relative to variations across multiple measurements, so only standard errors are reported here. This data may also be fit above the melting transition as a linear combination of A3G-saturated ssDNA (purple, Equation 6.3) and ssDNA (Equation 2.5) because the molecule is primarily ssDNA regardless of protein binding. A3G binding during incubation shifts the release curve toward the saturated A3G curve (Equation 6.5, fit not shown for clarity). (B) Total A3G bound without incubation ($t = 50$ s) as a function of concentration. Fits above (green) and below (blue) the melting transition yield similar results, as expected. Data points are averages from at least three fits, and error bars reflect standard error. The values from two methods were averaged (points and propagated error bars not shown for clarity) and fit to a simple DNA binding isotherm (solid red line, Equation 4.1). The fit yields $K_d = 125 \pm 25$ nM, which is similar to $K_d = 76 \pm 21$ nM [244] and $K_d = 160$ nM [259] obtained from ensemble measurements. This measured $K_d$ value agrees with $K_{d1} = 127 \pm 6$ nM for fast binding obtained from $K_1$ (Equation 6.17, Figure 6.6A).

### 6.4.3 Two-step binding model

The kinetics of A3G-ssDNA binding was described as a two-step process, with an initial fast bimolecular binding step that reaches pre-equilibrium before a slow, unimolecular conversion to a more stable complex (Equation 6.1). A complete analytical description of this process [221] yields explicit expressions for the fast, slow, and total fractions of A3G bound as a function of time:

\[
\begin{align*}
    f_{\text{fast}}(t) &= P_{\text{total}} P_{\text{fast}} (e^{-k_{\text{slow}} t} - e^{-k_{\text{fast}} t}) \\
    f_{\text{slow}}(t) &= P_{\text{total}} (1 - e^{-k_{\text{slow}} t}) \\
    f_{\text{total}}(t) &= P_{\text{total}} (1 - P_{\text{fast}} e^{-k_{\text{fast}} t} - (1 - P_{\text{fast}}) e^{-k_{\text{slow}} t})
\end{align*}
\]

where $k_{\text{fast}}$ and $k_{\text{slow}}$ are binding rates, and $P_{\text{fast}}$ and $P_{\text{total}}$ are the equilibrium fractions of fast and total complex bound. $P_{\text{fast}}$ is the ratio of the forward rate and the total rate for the first step:

\[
P_{\text{fast}} = \frac{k_1 c}{k_1 c + k_{-1}}
\]

Fits to Equations 6.6-6.8 (Figure 6.3A-B) yield the four parameters $k_{\text{fast}}$, $k_{\text{slow}}$, $P_{\text{fast}}$, and $P_{\text{total}}$ (reported in Table 6.1.), which are used to determine the four elementary reaction rates $k_1 c$, $k_{-1}$, $k_2$ and $k_{-2}$, as well as the two equilibrium binding constants $K_1$ and $K_2$, for the first and the second binding steps, respectively.

### 6.4.4 Fast on and off rates
The fast rate $k_{fast}$ is a sum of the forward and backward elementary reaction rates:

$$k_{fast} = k_1 c + k_{-1} \tag{6.10}$$

where $k_1 c$ is the on rate and $k_{-1}$ is the off rate for the first step, defined in Equation 6.1. The model assumes the fast process is bimolecular, so the on rate is proportional to A3G concentration $c$ with rate constant $k_1$. The data confirms that $k_{fast}(c)$ is linear (Figure 6.3C), with bimolecular rate constant $k_1 = 2.0 \ (\pm 0.1) \times 10^5 \text{ M}^{-1}\text{s}^{-1}$ and off rate $k_{-1} = 1.0 \ (\pm 0.1) \times 10^{-2} \text{ s}^{-1}$ (Table 6.2, method A1).

The fast on and off rates were also calculated as follows:

$$k_1 c = k_{fast} P_{fast} \tag{6.11}$$

The slope of a linear fit to the calculated on rate versus $c$ (Figure 6.3C) is an additional measurement of $k_1$:

$$k_1 c = k_1 c + b \tag{6.12}$$

where the measured value of $k_1 = 1.4 \ (\pm 0.1) \times 10^5 \text{ M}^{-1}\text{s}^{-1}$ (Table 6.2, method A2) is consistent with the value from the first method. As expected, the y-intercept of the fit is nearly zero ($b = -8 \ (\pm 5) \times 10^{-4} \text{ s}^{-1}$). The binding model also defines the off rate $k_{-1}$:

$$k_{-1} = k_{fast} (1 - P_{fast}) \tag{6.13}$$

which is expected to be concentration-independent (Figure 6.3D). Although the linear fit:

$$k_{-1} = m c + k_1 \tag{6.14}$$

yields a non-zero slope ($m = 6.1 \ (\pm 0.3) \times 10^4 \text{ M}^{-1}\text{s}^{-1}$), the linear dependence is an order of magnitude smaller than that of $k_{fast}(c)$. This measurement of $k_{-1} = 1.1 \ (\pm 0.1) \times 10^{-2} \text{ s}^{-1}$ (Table 6.2, method A2) agrees with the value from the first method within uncertainty.

The bimolecular rate constant calculated from Equation 6.11 is concentration-independent as expected, with a weighted average of $k_1 = 1.2 \ (\pm 0.1) \times 10^5 \text{ M}^{-1}\text{s}^{-1}$. The $k_1$ values from the second and third methods agree within uncertainty. The off rates calculated from Equation 6.13 are also nearly concentration-independent, and the weighted average of the five concentrations is $k_{-1} = 2.0 \ (\pm 0.2) \times 10^{-2} \text{ s}^{-1}$ (Table 6.2, method A3).

### 6.4.5 Fast equilibrium binding constant

The equilibrium binding constant for the bimolecular reaction step, $K_1$, is a ratio of the forward and backward rates:

$$K_1 = \frac{k_1 c}{k_{-1}} \tag{6.15}$$

and was calculated from values of $k_1 c$ and $k_{-1}$ obtained from fits. It may also be calculated from $P_{fast}(c)$:

$$K_1 = P_{fast} (1 - P_{fast}) \tag{6.16}$$

The fast equilibrium binding constant depends linearly upon A3G concentration, as expected (Figure 6.6A), and the slope was used to calculate the equilibrium association constant $K_1/c$ (Table 6.2, method C1).
Experimentally determined values of $k_1$ and $k_2$ were also used to calculate $K_d/c$ directly from Equation 6.15 (Table 6.2, method C2). This was used to obtain the equilibrium dissociation constant $K_{d1}$ for the first step:

$$K_{d1} = \frac{c}{K_1} = \frac{k_1}{k_1}$$

6.17

The $K_d$ value obtained from this method (127 ± 6 nM) agrees with the $K_d$ of the total fraction bound at $t = 50$ s of A3G exposure to ssDNA (125 ± 25 nM) fit to a simple DNA binding isotherm (Figure 6.5B). This is a reasonable observation, in light of the fact that total A3G binding measured after only 50 s is primarily fast binding, which has equilibrated ($1/k_{fast} = 24 ± 1$ s) while slow binding has not ($1/k_{slow} = 206 ± 20$ s).

**Figure 6.6.** (A) The equilibrium binding constant $K_1$ is linearly dependent on A3G concentration, as expected for a bimolecular process. The solid line is a linear fit to Equation 6.15 with slope $k_1/k_1 = 6.6 (± 0.4) \times 10^6$ M$^{-1}$, which yields a dissociation constant for the fast step ($K_{d1} = 127 ± 6$ nM) that agrees with the one obtained from a simple DNA binding isotherm without incubation ($K_d = 125 ± 25$ nM, see Figure 6.5B). (B) Fast portion of A3G binding $P_{fast}$ saturates at high protein concentration. Data points and their uncertainties are from fits to the binding model (see Table 6.1), solid line is a fit to Equation 6.9 that yields $k_1 = 1.4 (± 0.1) \times 10^5$ M$^{-1}$ s$^{-1}$ and $k_1 = 2.1 (± 0.9) \times 10^{-2}$ s$^{-1}$. These values are fixed in the $k_{slow}$ versus $c$ fit (Figure 6.3D, solid green line) to limit the fit to two free parameters.

<table>
<thead>
<tr>
<th>A3G (nM)</th>
<th>$k_{fast}$ (x 10$^{-2}$ s$^{-1}$)</th>
<th>$k_{slow}$ (x 10$^{-3}$ s$^{-1}$)</th>
<th>$P_{fast}$</th>
<th>$P_{total}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1.2 ± 0.2</td>
<td>0.5 ± 0.1</td>
<td>0.05 ± 0.03</td>
<td>0.90 ± 0.01</td>
</tr>
<tr>
<td>20</td>
<td>1.4 ± 0.2</td>
<td>0.7 ± 0.2</td>
<td>0.12 ± 0.03</td>
<td>0.88 ± 0.02</td>
</tr>
<tr>
<td>50</td>
<td>1.9 ± 0.2</td>
<td>1.6 ± 0.3</td>
<td>0.25 ± 0.07</td>
<td>0.95 ± 0.02</td>
</tr>
<tr>
<td>100</td>
<td>3.0 ± 0.2</td>
<td>2.9 ± 0.3</td>
<td>0.45 ± 0.05</td>
<td>0.97 ± 0.05</td>
</tr>
<tr>
<td>200</td>
<td>5.0 ± 0.5</td>
<td>5.0 ± 1.0</td>
<td>0.55 ± 0.06</td>
<td>0.99 ± 0.15</td>
</tr>
</tbody>
</table>

**Table 6.1.** Values for fitting parameters in the steady-state binding model. Total binding at a given A3G concentration ($f_{total}$) was separated into its fast ($f_{fast}$) and slow ($f_{slow}$) fractions and monitored as a function of ssDNA-A3G incubation time, as illustrated in Figure 6.2. The three curves obtained at each protein concentration (shown in Figure 6.3A for 50 nM A3G) were fit to Equations 6.6-6.8. The parameters from each fit were used to calculate weighted averages for the observed binding rates $k_{fast}$ and $k_{slow}$ and the equilibrium fractions of fast and total binding $P_{fast}$ and $P_{total}$.

### 6.4.6 Slow on and off rates

The slow rate $k_{slow}$ is the sum of the forward rate $k_2$ and backward rate $k_{-2}$:

$$k_{slow} = k_2 P_{fast} + k_{-2}$$

6.18

76
where the effective forward rate is the product of the conversion rate \( k_2 \) and the equilibrium fraction of the fast bound complex \( P_{fast} \). A fit of \( P_{fast}(c) \) to Equation 6.9 (Figure 6.6B) is a fourth method of measuring the fast on and off rates, \( k_1 = 1.4 \pm 0.1 \times 10^5 \text{M}^{-1}\text{s}^{-1} \) and \( k_1 = 2.1 \pm 0.9 \times 10^{-2} \text{s}^{-1} \) (Table 6.2, method A4). These values are fixed in the fit of \( k_{slow}(c) \) to Equation 6.18, which yields \( k_2 = 7.0 \pm 1.0 \times 10^{-3} \text{s}^{-1} \) and \( k_2 = 3.0 \pm 2.0 \times 10^{-5} \text{s}^{-1} \) (Table 6.2, method B1).

An alternative way to calculate these elementary reaction rates involves the slow equilibrium binding constant \( K_2 \):

\[
K_2 = \frac{k_2}{k_1}
\]

which can also be expressed as:

\[
K_2 = \frac{1 + \frac{1}{K_1}}{1 - \frac{1}{P_{total}}}
\]

Calculations from Equations 6.19 and 6.20 yield a weighted average of \( K_2 = 174 \pm 16 \), as reported in Table 6.2 (method D), suggesting that the second binding step is so highly forward-driven that it is nearly irreversible.

The value of \( K_2 \) may be used to calculate the forward rate \( k_2 \):

\[
k_2 = \frac{k_{slow}K_2}{K_2P_{fast} + 1}
\]

and the backward rate \( k_2 \):

\[
k_2 = \frac{k_{slow}}{K_2P_{fast} + 1}
\]

The weighted average of these values is \( k_2 = 6.6 \pm 0.7 \times 10^{-3} \text{s}^{-1} \) and \( k_2 = 2.8 \pm 0.5 \times 10^{-5} \text{s}^{-1} \) (Table 6.2, method B2), which is consistent with the estimate of these rates based on the fit of \( k_{slow} \) as discussed above. Both on and off rates are independent of protein concentration within uncertainty (Figure 6.3D), consistent with a model in which slow binding is unimolecular. Therefore, \( k_{slow} \) increases with A3G concentration as fast binding grows, and then saturates with on rate \( k_2 \) at high protein concentrations such as those relevant in the virus.

<table>
<thead>
<tr>
<th>Method</th>
<th>( k_1 \times 10^5 \text{M}^{-1}\text{s}^{-1} )</th>
<th>( k_2 \times 10^{-3} \text{s}^{-1} )</th>
<th>( k_2 \times 10^{-3} \text{s}^{-1} )</th>
<th>( K_2/c \times 10^8 \text{M}^{-1} )</th>
<th>Method</th>
<th>( K_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>2.0 ± 0.1</td>
<td>1.0 ± 0.1</td>
<td>7.0 ± 1.0</td>
<td>3.0 ± 2.0</td>
<td>B1</td>
<td>6.6 ± 0.4</td>
</tr>
<tr>
<td>A2</td>
<td>1.4 ± 0.1</td>
<td>1.1 ± 0.1</td>
<td>6.6 ± 0.7</td>
<td>2.8 ± 0.5</td>
<td>B2</td>
<td>12.9 ± 0.8</td>
</tr>
<tr>
<td>A3</td>
<td>1.2 ± 0.1</td>
<td>2.0 ± 0.2</td>
<td>7.0 ± 1.0</td>
<td>3.0 ± 2.0</td>
<td>C1</td>
<td>6.6 ± 0.4</td>
</tr>
<tr>
<td>A4</td>
<td>1.4 ± 0.1</td>
<td>2.1 ± 0.9</td>
<td>6.6 ± 1.0</td>
<td>3.0 ± 2.0</td>
<td>D1</td>
<td>166 ± 17</td>
</tr>
<tr>
<td>Average</td>
<td>1.5 ± 0.1</td>
<td>1.2 ± 0.1</td>
<td>6.7 ± 0.6</td>
<td>2.8 ± 0.5</td>
<td>C2</td>
<td>7.9 ± 0.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>D2</td>
<td>251 ± 54</td>
</tr>
</tbody>
</table>

Table 6.2. Concentration-independent reaction parameters are similar across multiple methods, which supports the binding model. Weighted averages and uncertainty are reported in the main text. Methods A1 and A2 are linear fits of \( k_{slow}, k_1 \), and \( k_3 \) vs. \( c \) (Figure 6.3A). Calculations from Equations 6.11 and 6.13 were averaged across protein concentration in method A3. Method A4 is the \( P_{fast} \) vs. \( c \) fit (Figure 6.6B), and method B1 is the \( k_{slow} \) vs. \( c \) fit (Figure 6.3B). Calculations from Equations 6.21
and 6.22 were averaged across A3G concentration in method B2. Method C1 is the $K_1$ vs. $c$ fit (Figure 6.6A) and method C2 uses the weighted averages of $k_1$ and $k_{-1}$ in Equation 6.15. The weighted average of calculations from Equation 6.20 over all measured protein concentrations is method D1. Method D2 uses experimentally-determined parameters in Equation 6.19.

**Acknowledgements**

We thank Micah McCauley for discussions. This work was supported in part by NIH [GM072462] and NSF [MCB-0744456] (MCW), NIH [GM065056] (KMF), and the NIH Intramural Research Program (NICHD) (JGL). KRC was supported by the NSF IGERT Program [DGE-0504331].
Conclusions

I never see what has been done; I only see what remains to be done.  

Marie Curie

Optical tweezers are a powerful quantitative technique for investigating the biophysics of DNA binding proteins. We develop single molecule methods to characterize proteins involved in replication of *E. coli*, retrotransposons Ty3 and LINE-1 endogenous to eukaryotic cells, and the retrovirus HIV in human cells.

DNA polymerases replicate the cellular genome under conditions of DNA damage in order to survive under stress, and we investigate DNA damage tolerance in the model system *E. coli*. UmuD is a DNA damage response protein that regulates bacterial mutagenesis through its interactions with DNA polymerases [264]. We use single molecule biophysics to determine that UmuD disrupts the interaction between the replicate polymerase DNA pol III α and ssDNA. This suggests that UmuD removes α from its ssDNA template, which may allow DNA repair proteins to access the damaged DNA for non-mutagenic repair. It may also be one mechanism by which UmuD facilitates exchange of α for an error-prone TLS polymerase to allow DNA replication to proceed despite the mutagenic cost. Understanding DNA damage tolerance mechanisms in *E. coli* contributes to an understanding of the multiple systems cells use to survive under stress.

We also use optical tweezers to investigate the function of nucleic acid chaperone proteins from endogenous retrotransposons, which cause insertional mutagenesis in eukaryotic cells [16]. We show that Ty3 NC and LINE-1 ORF1p both balance binding to dsDNA and ssDNA to facilitate nucleic acid rearrangements during retrotransposition. This nucleic acid chaperone activity is required for successful retrotransposition, and understanding replication of Ty3 and LINE-1 provides insight into eukaryotic mutagenesis and the evolution of retroviruses such as HIV.

The human cellular protein A3G inhibits HIV-1 replication, and the virus encodes Vif to combat packaging of A3G inside virions [265]. A3G may partially inhibit HIV-1 infectivity in vivo even in the presence of Vif, since HIV patients exhibit A3G-induced mutations of the viral genome [265]. The mechanism by which A3G inhibits HIV-1 replication is not yet fully understood, and we develop a method to quantify two seemingly contradictory molecular mechanisms. We demonstrate that oligomerization transforms A3G from a fast enzyme to a slow nucleic acid binding protein, and report all of the fundamental rates that describe this process. Our results suggest a model for A3G’s effects on viral replication in which A3G oligomers initially stall RT during viral DNA synthesis. Upon dissociation from the RNA genome, the oligomer becomes monomers or dimers that rapidly mutate the newly-
synthesized viral ssDNA strand. Over time, A3G oligomerizes again and blocks RT from synthesizing the second viral DNA strand. The biophysical mechanism we illustrate may also be how other APOBEC proteins regulate their enzymatic activity in order to inhibit retroviral replication in multiple ways. A complete understanding of this mechanism is essential for new antiretroviral therapies such as the rational design of drugs that mimic aspects of A3G’s function, protect A3G from Vif, or efficiently deliver cellular A3 proteins to the viral core [265].

In this work, we describe new approaches for studying DNA binding proteins with single molecule force spectroscopy. Our findings illustrate that basic biophysical mechanisms regulate the DNA interaction properties of proteins involved in complex biological processes. Quantitative characterization of these binding mechanisms provides fundamental insight into protein function that is crucial for treatment of human disease.
Publications

6. **Kathy R. Chaurasiya**, Clarissa Ruslie, Michelle C. Silva, Philip Nevin, Samer Lone, Penny J. Beuning, and Mark C. Williams. Polymerase manager protein UmuD regulates *E. coli* DNA polymerase III α binding to ssDNA. *Submitted.*


## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td>DNA pol III α, DNA polymerase III subunit α</td>
</tr>
<tr>
<td>A</td>
<td>adenine (nucleobase)</td>
</tr>
<tr>
<td>A</td>
<td>alanine (amino acid)</td>
</tr>
<tr>
<td>A3G</td>
<td>APOBEC3G, apolipoprotein B mRNA editing enzyme</td>
</tr>
<tr>
<td>AFM</td>
<td>atomic force microscopy</td>
</tr>
<tr>
<td>C</td>
<td>cytosine</td>
</tr>
<tr>
<td>CTD</td>
<td>C-terminal domain</td>
</tr>
<tr>
<td>D</td>
<td>aspartic acid</td>
</tr>
<tr>
<td>cDNA</td>
<td>complementary DNA</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>dsDNA</td>
<td>double-stranded DNA</td>
</tr>
<tr>
<td>E. coli</td>
<td><em>Escherichia coli</em></td>
</tr>
<tr>
<td>F</td>
<td>phenylalanine</td>
</tr>
<tr>
<td>FJC</td>
<td>freely-jointed chain</td>
</tr>
<tr>
<td>G</td>
<td>guanine</td>
</tr>
<tr>
<td>GST</td>
<td>glutathione S-transferase</td>
</tr>
<tr>
<td>H</td>
<td>histidine</td>
</tr>
<tr>
<td>HhH</td>
<td>helix-hairpin-helix motif</td>
</tr>
<tr>
<td>HIV-1</td>
<td>human immunodeficiency virus type 1</td>
</tr>
<tr>
<td>HTLV-1</td>
<td>human T-cell lymphotropic virus type 1</td>
</tr>
<tr>
<td>IGERT</td>
<td>Integrative Graduate Education and Research Traineeship</td>
</tr>
<tr>
<td>K</td>
<td>lysine</td>
</tr>
<tr>
<td>L1</td>
<td>LINE-1, long interspersed nuclear element type 1</td>
</tr>
<tr>
<td>LTR</td>
<td>long-terminal repeat</td>
</tr>
<tr>
<td>Met</td>
<td>methionine</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>mtSSB</td>
<td>mitochondrial single-stranded binding protein</td>
</tr>
<tr>
<td>MuLV</td>
<td>Moloney murine leukemia virus</td>
</tr>
<tr>
<td>N</td>
<td>nitrogen</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>NC</td>
<td>nucleocapsid protein</td>
</tr>
<tr>
<td>NIH</td>
<td>National Institutes of Health</td>
</tr>
<tr>
<td>nm</td>
<td>nanometer ($1 \times 10^{-9}$ meters)</td>
</tr>
<tr>
<td>nM</td>
<td>nanomolar ($1 \times 10^{-9}$ molar)</td>
</tr>
<tr>
<td>NSF</td>
<td>National Science Foundation</td>
</tr>
<tr>
<td>NTD</td>
<td>N-terminal domain</td>
</tr>
<tr>
<td>O</td>
<td>oxygen</td>
</tr>
<tr>
<td>ORF1p</td>
<td>open reading frame 1 protein</td>
</tr>
<tr>
<td>P</td>
<td>phosphorous</td>
</tr>
<tr>
<td>PBS</td>
<td>primer binding site</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein Data Bank</td>
</tr>
<tr>
<td>pN</td>
<td>piconewton ($1 \times 10^{-12}$ Newtons)</td>
</tr>
<tr>
<td>R</td>
<td>arginine</td>
</tr>
<tr>
<td>RPA</td>
<td>replication protein A</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>RRM</td>
<td>RNA recognition domain</td>
</tr>
<tr>
<td>RSV</td>
<td>Rous sarcoma virus</td>
</tr>
<tr>
<td>RT</td>
<td>reverse transcriptase</td>
</tr>
<tr>
<td>SINE</td>
<td>short interspersed nuclear element</td>
</tr>
<tr>
<td>SSB</td>
<td>single-stranded binding protein</td>
</tr>
<tr>
<td>ssDNA</td>
<td>single-stranded DNA</td>
</tr>
<tr>
<td>T</td>
<td>thymine</td>
</tr>
<tr>
<td>TAR</td>
<td>trans-activation response element</td>
</tr>
<tr>
<td>TLS</td>
<td>translesion synthesis</td>
</tr>
<tr>
<td>tRNA</td>
<td>transfer RNA</td>
</tr>
<tr>
<td>TPRT</td>
<td>target site-primed reverse transcription</td>
</tr>
<tr>
<td>U</td>
<td>uracil</td>
</tr>
<tr>
<td>UTR</td>
<td>untranslated region</td>
</tr>
<tr>
<td>Vif</td>
<td>virion infectivity factor</td>
</tr>
<tr>
<td>WCL</td>
<td>whole cell lysates</td>
</tr>
<tr>
<td>WLC</td>
<td>worm-like chain</td>
</tr>
<tr>
<td>wt</td>
<td>wild-type</td>
</tr>
<tr>
<td>Y</td>
<td>tyrosine</td>
</tr>
</tbody>
</table>
References


160. R. Sharma, Northeastern University (2010).


170. S. Jergic et al., The unstructured C-terminus of the τ subunit of *Escherichia coli* DNA polymerase III holoenzyme is the site of interaction with the α subunit. *Nucleic Acids Research* **35**, 2813-2824 (2007).


100


