SINGLE MOLECULE FORCE SPECTROSCOPY STUDIES OF DNA BINDING AND CHAPERONE PROTEINS

A dissertation presented
by
Fei Wang
to
The Department of Physics
In partial fulfillment of the requirements for the degree of Doctoral Philosophy
In the field of
Physics
Northeastern University
Boston, Massachusetts
December 2008
SINGLE MOLECULE FORCE SPECTROSCOPY STUDIES OF DNA BINDING AND CHAPERONE PROTEINS

by

Fei Wang

ABSTRACT OF DISSERTATION

To the department of Physics
In partial fulfillment of the requirements
for the degree of Doctoral Philosophy
in the field of Physics of Northeastern University
Boston, Massachusetts
December 2008
Abstract

The nucleocapsid protein (NC) plays an important role in retroviral replication, in part, by facilitating numerous nucleic acid rearrangements throughout the reverse transcription process. The nucleic acid chaperone activity of the human immunodeficiency virus type-1 (HIV-1) NC has been extensively studied, and duplex destabilization, nucleic acid aggregation, and rapid protein binding kinetics have been identified as major components of the activity of this highly basic protein (pI ~10). The chaperone activity of other NC proteins is not well understood. We used single molecule DNA stretching to characterize the activity of HIV-1, RSV, and MMLV NC. We found distinct differences in the chaperone activities of each protein, which reflect the requirements for nucleic acid chaperone activity in each retroviral replication system.

HTLV-1 NC exhibited overall poor nucleic acid chaperone activity. This result is explained by its poor aggregating activity and slow dissociation from single-stranded DNA. This NC protein is overall neutral at pH=7.5 and possesses a unique, acidic C-terminal domain. By studying different HTLV-1 NC mutants, the role of C-terminal domains to the chaperone activity was elucidated. The results suggest that the electrostatic interaction between HTLV-1 NC and nucleic acids is the major factor determining the kinetics. We also examine the nucleic acid interaction properties of the Apolipoprotein B mRNA editing enzyme, a catalytic polypeptide-like 3G (APOBEC3G/A3G) that is known to inhibit HIV-1 reverse transcription in absence of viral infectivity factor (Vif). Our stretching experiments suggested a novel mechanism for deaminase-independent inhibition of reverse transcription due to vital differences of nucleic acid
binding kinetics between NC, A3G and reverse transcriptase (RT). Finally, Long
interspersed nucleic elements (LINE) are highly repeated nucleic acids sequences in
mammal genomes. Our single DNA molecule stretching experiments characterized the
nucleic acid chaperone function of ORF1p in the mouse LINE-1 retrotransposon. We
found that a single amino acid substitution altered retrotransposition efficiency by a
factor of 15 due to a reduction in nucleic acid chaperone activity exhibited by ORF1p.
For all of the studies presented here, we used single molecule methods to characterize
the nucleic acid interactions of proteins involved in reverse transcription in retroviruses
or retrotransposons. In each case, complementary bulk experiments were done by
collaborators. The results are presented together in each chapter of the thesis.
Acknowledgements

I want to thank my wife first for her support. She took care of our home and never doubted my goal. And I want to specially thank my advisor, Professor Mark Williams. He not only guided me in my experiments but also taught me how to be prepared in future academic career. This work is results of several exciting collaboration with professors from different universities. I am very glad to have chance working with them on such inspiring projects. The NC project was collaborated with Professor Karin Musier-Forsyth from Ohio State University. Professor Ioulia Rouzina from University of Minnesota helped on building models to explain our experimental results. The A3G project was collaborated with Professor Judith Levin from National Institutes of Health. The ORF project was collaborated with Professor Sandra Martin from University of Colorado School of Medicine. I’m thankful that all projects turned to be successful achievements. I also feel lucky to work in such great lab with wonderful friends. Our super Postdoc Micah McCauley always helped me whenever I asked him questions or had problems in my experiments. His excellent management in our lab made our job much easier. My good friend Thayapara paramanathan is such warm-hearted guy. It’s always pleasure to work with him. Kathy Chaurasiya helped a lot improve my writing. I also want to thank Dr. Margareta Crueanu and Dr. Leila Shokri for training when I started my projects. Dr. Ioana Vladescu also gave a lot of advice and discussion during my research. It was pleasant to share the setup with Chen Lu. Here I also want to thank people from Professor Champion’s lab, Professor Tim Sage’s lab and Professor Armen Sptepanyants’s lab from inspiring discussion in Biophysics group meeting. I do have a lot of people to thank. And I feel sorry not to list everyone’s name here. I wish everyone good luck in their jobs.
Table of Contents

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

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

1. INTRODUCTION ....................................................................................................................... 7
  1.1. OVERVIEW .......................................................................................................................... 8
  1.2. DNA AND RNA STRUCTURE ............................................................................................ 8
  1.3. DNA TRANSCRIPTION AND REVERSE TRANSCRIPTION ................................................. 11
  1.4. DNA MELTING .................................................................................................................. 12
  1.5. OPTICAL TWEEZERS ....................................................................................................... 15

2. COMPARATIVE STUDY OF NUCLEOCAPSID PROTEIN OF RETROVIRUS: HIV-1, RSV, MOMLV AND HTLV-1 ................................................................. 19
  2.1. BACKGROUND ................................................................................................................... 19
    2.1.1. Retrovirus life cycle (14) ............................................................................................. 19
    2.1.2. Nucleocapsid (NC) protein of retrovirus ................................................................. 23
             HIV-1 NCp7 .................................................................................................................. 25
             RSV NCp12 ................................................................................................................... 26
             MoMLV NCp10 ............................................................................................................. 27
             HTLV-1 NCp15 ............................................................................................................. 28
  2.2. MATERIALS AND METHODS ........................................................................................... 29
    2.2.1. Protein and nucleic acid preparation ........................................................................... 29
    2.2.2. Annealing assays ........................................................................................................ 31
    2.2.3. Sedimentation/aggregation assays ............................................................................ 32
    2.2.4. FA experiments ......................................................................................................... 33
    2.2.5. Single-molecule DNA stretching .............................................................................. 34
  2.3. RESULTS ................................................................................................................................ 35
    2.3.1. Effects of different retroviral NC proteins on mini-TAR RNA/DNA annealing ....... 35
             Mini-TAR RNA/DNA annealing kinetics ........................................................................ 35
             Temperature dependence of mini-TAR RNA/DNA annealing ........................................ 40
             Equilibrium stability of the mini-TAR RNA/DNA duplex ........................................... 42
    2.3.2. Effects of different retroviral NC proteins on mini-TAR RNA/DNA aggregation ... 45
    2.3.3. Nucleic acid binding studies of different retroviral NC proteins ............................... 47
    2.3.4. Duplex destabilization by retroviral NC proteins ..................................................... 48
    2.3.5. DNA-stretching studies with different retroviral NC proteins .................................... 51
  2.4. DISCUSSION ....................................................................................................................... 55

3. ELECTROSTATIC INTERACTION BETWEEN THE ANIONIC C-TERMINAL DOMAIN OF HTLV-1 NC WITH ITS CATIONIC N-TERMINAL DOMAIN LEADS TO THE SLOW NUCLEIC ACID - PROTEIN INTERACTION AND POOR CHAPERONE ACTIVITY ................................................................. 59
  3.1. BACKGROUND ................................................................................................................... 59
  3.2. METHODS AND MATERIALS ........................................................................................... 63
    3.2.1. Time-resolved FRET ................................................................................................ 63
  3.3. RESULTS ........................................................................................................................... 64
    3.3.1. Effect of wt HTLV-1 NC and mutant ΔC29 on Mini-TAR DNA/RNA annealing ........... 64
    3.3.2. Chelation of zinc from HTLV-1 NC zinc fingers does not improve its chaperone function 65
3.3.3. High salt makes WT and ΔC29 HTLV NC similarly good chaperones .............................................. 66
3.3.4. Salt effect on the kinetics of mini-TAR RNA/DNA annealing in the presence of HTLV-1 NC and ΔC29 HTLV-1 NC proteins ................................................................. 69
3.3.5. NA aggregating ability of HTLV-1 NC and its mutants correlate with the charge of its C terminal domain ................................................................................................................. 70
3.3.6. The nucleic acid binding and duplex destabilizing activities of HTLV-1 NC are only weakly affected by its CTD mutations ................................................................................. 73
3.3.7. HTLV-1 NC and its CTD mutants are good NA duplex destabilizers ......................................................... 74
3.3.8. Single molecule DNA stretching studies suggest slow kinetics of the WT HTLV-1 NC proteins and much faster dissociation of its CTD deletion mutant ............................................... 75
3.4. DISCUSSION ......................................................................................................................................... 79

4. DEAMINASE-INDEPENDENT INHIBITION OF HIV-1 REVERSE TRANSCRIPTION BY APOBEC3G ....... 85

4.1. BACKGROUND ............................................................................................................................................... 85
4.2. MATERIALS AND METHODS .................................................................................................................... 87
4.2.1. Materials .................................................................................................................................................. 87
4.2.2. Methods .................................................................................................................................................. 88
        Plasmid construction ................................................................................................................................. 88
        Preparation of RNA ................................................................................................................................. 88
        Reverse transcription assays ..................................................................................................................... 89
        tRNA\textsubscript{Lys}\textsuperscript{-} annealing to viral RNA (vRNA) ................................................................. 89
4.3. RESULTS ..................................................................................................................................................... 92
4.3.1. Effect of A3G on primer placement and (−) SSDNA synthesis .............................................................. 92
4.3.2. Effect of A3G on minus-strand transfer reactions .................................................................................. 96
4.3.3. Effect of A3G on (+) SSDNA synthesis and plus-strand transfer ......................................................... 99
4.3.4. Single molecule DNA stretching and FA-binding measurements ......................................................... 101
4.4. DISCUSSION .............................................................................................................................................. 103

5. A SINGLE AMINO ACID SUBSTITUTION IN ORF1 DRAMATICALLY DECREASES L1 RETROTRANSPOSITION AND PROVIDES INSIGHT INTO NUCLEIC ACID CHAPERONE ACTIVITY ........... 109

5.1. BACKGROUND ............................................................................................................................................ 109
5.2. MATERIALS AND METHODS ................................................................................................................ 112
5.2.1. Constructs ............................................................................................................................................. 112
5.2.2. Cell culture and autonomous retrotransposition .................................................................................. 112
5.2.3. Immunofluorescence microscopy of L1 ORF1p .................................................................................... 113
5.2.4. RNA, DNA and protein analyses from transfected cells ....................................................................... 114
5.2.5. Assays using purified proteins .............................................................................................................. 115
5.3. RESULTS .................................................................................................................................................. 117
5.4. DISCUSSION ............................................................................................................................................ 126

REFERENCE: ............................................................................................................................................... 133

1. Introduction
1.1. Overview

In the work described here we will study the interaction of various proteins with DNA and RNA, defined below. In the Introduction, we will describe the biophysical and biochemical characteristics of DNA and RNA. We will then describe DNA stretching with optical tweezers, and discuss how these experiments shed light on DNA-protein interactions. In the subsequent chapters, we will describe experiments to probe specific nucleic acids-protein interactions that are essential for the replication of retrovirus and retrotransposons. In chapter 2, complementary experimental approaches are used to characterize and compare the chaperone activities of NC proteins from four different retroviruses: Human immunodeficiency virus type 1 (HIV-1), Moloney murine leukemia virus (MoMLV), Rous sarcoma virus (RSV), and human T-cell lymphotropic virus type 1 (HTLV-1). In chapter 3, we further investigate the extremely poor nucleic acid (NA) chaperone activity of the nucleocapsid (NC) protein from HTLV-1. In chapter 4, we report the interplay between NC, reverse transcriptase (RT) and A3G, a host protein that inhibits HIV-1 reverse transcription and replication in the absence of viral infectivity factor (Vif). In the final chapter, we examine two chaperone proteins from the LINE-1 retrotransposon. Despite the fact that these chaperone proteins differ by only one amino acid, we find that one protein is 15 fold more efficient as a chaperone relative to the other protein.

1.2. DNA and RNA structure

DNA (Deoxyribonucleic acid) is a directional biopolymer with backbone composed of phosphate and deoxyribose and four possible nucleobases as its side groups (5,6). As shown in figure 1.1a, carbon atoms on deoxyribose ring can be labeled from 1 to 5. Directionality of DNA is due to phosphate groups always connecting between 3’ and 5’ carbon from neighboring deoxyribose. By biological convention DNA sequence can be represented by a letter string starting from 5’ end using nucleoside
Figure 1.1: DNA and RNA structure shown in 2D plane. a) double-stranded DNA linked by hydrogen bonds (dash lines). Carbon atoms in deoxyribose ring of top two nucleotides are labeled from 1 to 5. Four ends of DNA strands are indicated with 3' or 5'. Nucleobases are indicated by letter symbols. b) RNA sequence with 4 different bases. The additional hydroxyl group at 2’ position is shown. U on the top substitutes T in DNA sequence.

symbols, A for adenosine, C for cytidine, G for guanosine, T for thymidine. In the 1950s the double helix DNA structure was postulated based on crystallography data (6). In this structure, known as B form, DNA makes a turn every 3.4 nm, spanning 10 base pairs.

Two DNA strands are linked together by base pairing between its nucleobases, form right-handed double helix. Only AT and GC base pairs are energetically favored between the four nucleobases in the DNA double helix. The AT base pair forms two hydrogen bonds and the GC base pair has three hydrogen bonds, as shown in figure 1.1a. The DNA
sequence of GC-rich base pairs is thermodynamically more stable than sequence of AT-rich base pairs. However, double-stranded DNA (dsDNA) is mainly stabilized by stacking interaction between nearest base pairs, not by hydrogen bonds between base pairs (7). Because of unique base pairing, DNA strands are complementary to each other in the DNA duplex (8). This specific pairing determines the mechanism of RNA transcription from DNA by the RNA polymerase. The strand storing gene information is called coding strand and its complementary strand is called the template strand. We will discuss DNA transcription in next section.

RNA (ribonucleic acid) has ribose instead of deoxyribose in its backbone and Uracil (U) substituting Thymine (T) as its nucleobase. So RNA has similar base paring as DNA. However, because of the additional hydroxyl group on the 2 prime position of the ribose ring, RNA forms an A-form helix rather than DNA’s B-form (9). Most of the time RNA doesn’t form the double helix structure in the cell. DNA is ideal for storing genetic information and RNA serves primarily as intermediate messenger to pass genetic information. This RNA is called messenger RNA (mRNA), which life time can be only seconds or hours differing in different cells. Messenger RNA is relatively long, which can be more than 10k bases, and very flexible. Some complementary sequences of same RNA strand can re-anneal to each other and form short duplex, which is called hairpin or pseudo knots depending on the structure (10). This structure is very important for RNA splicing and enzyme activities. There are also other very important functional RNAs, such as transfer RNA (tRNA) and ribosomal RNA (rRNA).
The Phosphate group in backbone of DNA and RNA is highly negatively charged, which makes DNA and RNA strands negatively charged. In solution counterion concentration in the proximity of DNA/RNA can be 1-2 M, which screens the repulsion force between two strands in the DNA duplex and is independent of solution concentration (11,12). Therefore, the solution salt concentration affects the stability of the DNA duplex.

1.3. DNA transcription and reverse transcription

Prokaryotic and eukaryotic transcription is different in some stages involving different proteins, but both of them use RNA polymerase to transcribe DNA to mRNA (13). Before DNA transcription starts, RNA polymerase binds to DNA and unwinds the DNA double helix to create a DNA bubble so RNA polymerase can access single-stranded DNA. Using the single-stranded DNA template, RNA polymerase makes a new complementary RNA strand, which is a copy of coding strand with U substituting T. DNA transcription can simultaneously start from different position of one gene sequence and cycle many times to produce multiple mRNA. Messenger RNA will eventually be translated to protein.

In retroviruses, genetic information is reserved in viral RNA and the retrovirus integrates its genetic information into the host cell DNA by a mechanism that uses reverse transcription to create a DNA copy of its RNA sequence. Human immunodeficiency virus type 1 (HIV-1), the causative agent of acquired immunodeficiency syndrome (AIDS), is one type of retrovirus that is well studied (14). It
has an RNA dimer, composed of two identical copies of its genomic RNA, inside of its virion. After the virion is fused into host cell during infection, reverse transcriptase (RT) from the virion binds to one viral RNA strand and starts to copy a complementary DNA (cDNA) strand (minus strand) from this viral RNA starting from a vRNA primer at the primer binding site (PBS). After the nascent DNA strand is completed through a complex process involving a strand transfer, most of the viral RNA is degraded by RNase H domain of RT. This nascent DNA is used as a new template and another DNA strand is synthesized. These two new complementary DNA strands form a double helix. With the help of other proteins from the virion this DNA duplex is integrated into the host cell. Then using host cell machinery the viral DNA can be used to transcribe more viral RNA and RNA coding for proteins, which is eventually packed into new virions, and which then buds out. The reverse transcription mechanism is also used in some other biological systems, such as retrotransposons in mammalian genomes (1).

1.4. DNA melting

The double-stranded DNA helix is thermodynamically stable at room temperature. The transition curve in figure 1.2 is typical for DNA melting experiments (9). After DNA
solution temperature reaches to certain point, the DNA duplex is cooperatively
separated to two DNA strands. The cooperativity of DNA melting is indicated by the
narrow transition width. The transition temperature at the midpoint is called the DNA
melting temperature $T_m$.

The double-stranded DNA is an ordered helix, while single-stranded DNA is very
flexible. The dsDNA thermal melting can be described as one-dimensional helix-coil
transition. If $s = \exp\left(\frac{\Delta G}{k_B T}\right)$ is probability ratio of helix to coil and $\sigma = \exp\left(\frac{2\Delta G_s}{k_B T}\right)$ is
the nucleation factor or co-operativity parameter, the fraction of base pairs is
determined by formula (15):

$$\Theta(T) = \frac{1}{2} \left(1 + \frac{s(T) - 1}{\sqrt{(s(T) - 1)^2 + 4s(T)\sigma(T)}}\right), \quad (0.1)$$

where $\Delta G$ is free energy difference between dsDNA and ssDNA and $\Delta G_s$ is determined by the
free energy of the helix-coil boundary. In this model, dsDNA is considered as a long
homopolymer and self avoidance is ignored. At the melting temperature, the helical
fraction is equal to $\frac{1}{2}$, which means $\Delta G = \Delta H - T_m \Delta S$ and

$$T_m = \frac{\Delta H}{\Delta S}. \quad (0.2)$$

Here $\Delta H$ and $\Delta S$ is enthalpy and entropy change of DNA from separated strands to double
helix.
In experiments DNA melting temperature depends on fraction of GC base pairs and solution condition. The following formula gives a good estimation of $T_m$ for long dsDNAs in solution.(16,17)

$$T_m = 360.31 + 34.47 \times x_{GC} + \left( 20.15 - 6.52 \times x_{GC} \right) \log(I)$$  \hspace{1cm} (0.3)

In above formula, $x_{GC}$ is the fraction of GC base pairs, and $I = \frac{1}{2} \sum_{i=1}^{n} c_i q_i^2$ is the ionic strength of the solution with ion of net charge $q_i$ and concentration $c_i$. In our experiments we use lambda DNA with half GC base pairs and half AT base pairs in $100\text{mM Na}^+$, pH7.5. The estimated melting temperature is about 82°C, which agrees with experiments very well. There is another model to predict the melting temperature where nearest base pair stacking interaction is considered more significant than hydrogen bonds of base pairing. The formula is shown in formula (0.4) (7):

$$T_m = \frac{\Delta H}{\Delta S + R \ln \left( C_1 - \frac{C_2}{2} \right)} - 273.15\degree C$$  \hspace{1cm} (0.4)

where $R$ is ideal gas constant ($1.9865 \text{cal.mol}^{-1}.\text{K}^{-1}$) and $C_1$ and $C_2$ is concentration for two complementary strands assuming $C_1$ is greater than $C_2$.

Thermal melting causes DNA bubbles in the DNA duplex, which simulates a structure that may occur in DNA replication and transcription. This structure can be used to study the binding kinetics of proteins involved in DNA replication or transcription. However, in this traditional DNA melting assay proteins are easily
denatured at the transition temperature $T_m$. Our recently developed single molecule method described below can be performed at room temperature, which provides a unique way to study proteins involved in DNA replication and transcription.

1.5. Optical tweezers

In the 1970s Arthur Ashkin reported optical scattering and gradient forces on micrometer-sized particles (18). Later in the 1980s he applied this technique to trap a virus and a bacterium (19,20). After him other scientists such as Carlos Bustamante and Steve Block extended his technique and pioneered the application in biological systems (21,22). Optical trapping (or optical tweezers) provides a valuable technique for studying molecular motors with piconewton (pN) force resolution and nanometer (nm) spatial

---

**Figure 1.3:** dual bean optical tweezers schematics. a) two inferred laser beams pass through microscopic objectives and co-focus on a polystyrene bead. Another bead is stuck to a glass pipette. A lambda DNA is caught between two beads. b) polystyrene beads and glass pipette is seen in CCD camera. Shining points indicates the laser focal point (the laser trap). c) schematics shows stretching DNA while DNA is free to rotate.
resolution at the single molecule level. Researchers have also applied optical tweezers to other biological systems, such as DNA replication and transcription, to study binding kinetics and thermodynamics (23-26). Scientists have also modified and developed other force trapping techniques, for example magnetic tweezers and atomic force microscopies (AFM), to characterize biological system (27,28).

In the work described here we use dual beam laser optical tweezers as shown in figure 1.3 (29). Two infrared lasers of 200mW power co-focus on a 5-micrometer diameter, streptavidin-coated polystyrene bead. Another streptavidin-coated bead is hold to a glass pipette which can be moved by manual control knobs or by a piezo-electric stage at nanometer step size with a range of 20 micrometers. Bacteriophage

**Figure 1.4:** dsDNA stretching curve using optical tweezers. Stretching curves are shown in solid-dot lines and relaxing curve is shown in open dot. Theoretical fitting for dsDNA is shown in black solid line and fitting for ssDNA is shown in purple solid line. Light blue area indicates free energy required to melt dsDNA to ssDNA.
lambda DNA of 48,500 base pairs or 16.5 micrometers (µm) contour length is biotinylated on both 3’ ends. Biotins on both DNA ends can form very strong non-covalent bonds with streptavidin on polystyrene beads, which allows us to stretch the DNA up to 180pN by moving the bead on the glass pipette tip. Since only the 3’ end is labeled, dsDNA is free to rotate during stretching.

Figure 1.4 illustrates typical stretching curves for lambda DNA in solution of pH 7.5 and 100 mM Na⁺ at 18°C (24,25,29-31). When the extension of DNA strands is shorter than 16 µm (0.34nm/bp), the tension in DNA is not more than 30 pN. The tension is mainly due to thermal fluctuations in its extension, or entropy. Tension in DNA increases dramatically when we stretch the DNA further. Fitting this linear region gives the stretch modulus of double-stranded DNA (dsDNA). Until now the whole stretching curve can fit worm-like chain (WLC) model very well, which is described by equation (0.5) (21,32):

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

This model gives contour length per base-pair \( b_{ds}^{\text{max}} = 0.34 \text{nm} \), persistence length \( P_{ds} = 50 \text{nm} \), stretch modulus \( K_{ds} = 1000 \text{pN} \). After the stretching force reaches about 60pN, the DNA extension increases to 1.7 times of contour length with only 3pN additional stretching force. If we move the pipette back to starting point after the force plateau, the relaxation curve overlaps the previous stretching curve. This means the stretching including the force plateau is at equilibrium. If we stretch DNA further after
this plateau, stretching force goes up dramatically again. This part can fit the freely-jointed chain model for single-stranded DNA (ssDNA), which is indicated in figure 1.4 and described by equation (0.6) (32):

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

The fitting gives \( b_{sss}^{\text{max}} = 0.58 \text{nm} \), \( P_{ss} = 0.7 \text{nm} \) and \( K_{ds} = 900 \text{pN} \). The persistence length is consistent with our freely-joint chain model. So the force plateau is actually a phase transition from dsDNA to ssDNA due to external stretching force. The transition is at equilibrium, so the area between the stretching curve and the curve for ssDNA gives the free energy which is required to melt dsDNA to two parallel ssDNA (32).

\[
\Delta G (F) = \int_{0}^{f} \left[ b_{ss} (f) - b_{ds} (f) \right] df
\]  

This melting free energy depends on experimental conditions, such as pH value, salt concentration, and temperature (17,30,31,33,34). In the shown stretching curve the free energy to melt dsDNA is about 2.3k_B T/bp (17,31,34,35).

If we have proteins or chemicals in experimental solution, which bind to nucleic acids, the stretching curves will change due to interaction between nucleic acids and protein or chemicals. This gives us opportunities to study the binding thermodynamics of those proteins regulating DNA transcription or replication, and the mechanism of drugs which may have anti-cancer capabilities.
2. Comparative study of Nucleocapsid protein of retrovirus: HIV-1, RSV, MoMLV and HTLV-1

This work was the result of a collaboration with Karin Musier-Forsyth’s lab and was published in the Journal of Virology, October 2008, p. 10129-10142, Vol. 82. Bulk experiments were performed by Karin Musier-Forsyth’s lab and single molecule experiments were by Fei Wang. It has been reformatted to fit this chapter.

2.1. Background

Human immunodeficiency virus type-1 (HIV-1) is the cause of acquired immunodeficiency syndrome (AIDS) (36). Because of the aggressiveness of this virus in modern society, it has been intensely studied (14,37). HIV-1 belongs to a large family of viruses, called retroviruses, which integrates its genetic sequence into the host DNA using the reverse transcription mechanism. During reverse transcription, the retrovirus does not have a precise proofreading mechanism, as does DNA replication. Therefore, retroviruses mutate very often, which helps to avoid drug attack and makes it difficult to develop inhibitors or vaccines (14).

2.1.1. Retrovirus life cycle (14)

Infection starts when the mature retrovirus approaches the host cell and the envelope protein of the virus binds to the receptors of the host cell. Then the viral envelope protein fuses with the membrane of the host cell, which releases viral RNA and associated proteins into the cytoplasm. Reverse transcriptase from the retrovirus
starts to copy one of the viral RNAs to DNA. Later, this nascent DNA strand is used as

![Figure 2.1: life circle of retrovirus. RNA is indicated in red and DNA is in green. Viral proteins are marked as blue dots, while proteins from the host cell are illustrated in yellow.](image)

template to make a new complementary DNA strand. These two DNA strands form a
double helix, which is transported into nucleus and inserted into host cell chromatin.

Then, using host cell machinery, the integrated proviral DNA can be transcribed to RNA
many times. These RNAs may serves as mRNA to translate viral proteins or they may
serve as the viral genome of next generation of virus. The new viral RNA and proteins
will be packed and bud out. After the virus matures, a new virus life circle will start as
shown in figure 2.1.

Transcription was classically thought to only occur from DNA to RNA. However,
the retrovirus transfers its genetic information from RNA to DNA with help of reverse
transcriptase (RT). We focus on the reverse transcription step of retroviral life cycle. The detailed reverse transcription process is shown in figure 2.2. The retroviral genome encodes its protein at the middle. The gag gene encodes a protein polypeptide which may be cleaved to several important proteins in mature virion. The pol and env gene encode reverse transcriptase and viral envelope protein, respectively. Neglecting the poly-A tail, there is a repeat region (R) on both ends. Next to the R sequence there is a unique sequence for either the 5’ or 3’ end, named U5 or U3, respectively. Following the U5 sequence down the strand, a piece of sequence is important for initiating reverse
transcription, which is called primer binding site (PBS). A tRNA molecule (corresponding to lys3 for HIV-1) binds the PBS of the retroviral genome. The hydroxyl group of this tRNA initiates reverse transcriptase to start making a new DNA strand using viral RNA as template. Because of the RNase H activity of reverse transcriptase, the RNA template is degraded after new DNA nucleotides are added to the nascent DNA sequence, as shown in second step in figure 2.2. This DNA strand is called minus strand strong stop DNA (-ssDNA) and synthesis of this ssDNA is called minus strand transfer. This step pauses when reverse transcription meets the 5’ end of viral RNA. The R region of the new DNA strand is complementary to the R region on the 3’ end of the viral genome. These two sequences can anneal with help of the nucleocapsid (NC) protein from the retrovirus. After the annealing process -ssDNA can be further extended to the PBS of the viral genome shown in step 4 and 5 of figure 2.2. The R and U5 region have jumped from the left to the right side. Although most of the viral RNA sequence is degraded during this step, there is a sequence called polypurine tract (PPT) which is relatively resistant to RNA degradation. The PPT sequence is used as a primer to initiate synthesis of the complementary DNA strand, plus strand strong stop DNA (+ssDNA), using the nascent DNA as template. This step halts after +ssDNA transcribes the tRNA primer, which is degraded after this step. The PBS on +ssDNA is complementary to 3’ end of minus strand DNA, which can anneal to each other as shown in step 7. After annealing, the U3 and R domain on +ssDNA are moved to left side. Elongation of both minus strand DNA and plus strand DNA will continue and finally finish. The two PPT sequences are eventually degraded during this step. On both ends of newly synthesized DNA duplex
there are U3, R, and U5 domains, which is called the long terminal repeat (LTR). This new double stranded DNA will be transported to the nucleus and inserted into the host chromatin with the help of integrase from the retrovirus. Notice that the viral DNA, called provirus, has an extra U3 or U5 on either end. In some literature the retrovirus genome includes the complete LTR on both 5’ and 3’ ends.

During this critical reverse transcription in retrovirus life cycle, there are several important annealing steps, which require NC protein from the retrovirus. NC proteins are normally small and very sensitive to mutation. Therefore, studying the activities and properties of the NC protein, which facilitates these annealing steps, helps us to understand retroviral replication and develop inhibitors targeting the NC protein.

2.1.2. Nucleocapsid (NC) protein of retrovirus

The NC protein is encoded in the gag domain of viral RNA. NC comes from cleavage of gag polypeptides in mature virion or direct expression of the NC gene (14). NC proteins are small and have similar structures for different retroviruses. It normally has one or two CCHC-type zinc figures and one or two aromatic residues on each zinc finger, which interacts with nucleotides (38,39). On the N-terminus there are some basic residues which are important for protein-protein interactions and protein-nucleic acid interactions.

NC plays important roles in different steps of the retroviral life cycle. NC helps tRNA to bind PBS of viral RNA (40,41) and facilitates minus strand transfer by annealing the R domain of –sssDNA with R domain on 3’ end of viral RNA. NC protein is also one of the important co-factors for provirus integration. And the NC domain of the Gag protein
is very important for new viral RNA dimerization and packaging into new virion (40,42). All of the activities of the NC protein are related to its chaperone property, in which NC facilitates the rearrangement of nucleic acids into thermodynamically more stable conformations (43-47). Intensive study of the chaperone activities of HIV-1 NC revealed that chaperone activities of NC protein include nucleic acid aggregation (48,49) and weak duplex destabilization (24,25,50-61). Recent research also showed that rapid kinetics of nucleic acid dissociation was another key feature of NC chaperone properties (55).

In this study we focus on chaperone activities of NC proteins in minus strand transfer for several retroviruses, including HIV-1, Rous sarcoma virus (RSV), Moloney murine leukemia virus (MoMLV), and human T-cell lymphototropic virus type 1 (HTLV-1).

Secondary structures of four NC proteins are illustrated in figure 2.3.

**Figure 2.3:** secondary structure of R region in retrovirus: HIV-1, RSV, MoMLV, and HTLV-1. Zinc ion and zinc finger structure are indicated with coordinating amino acid residues circled. Aromatic residues are pointed to by arrows.

Nucleic acid chaperone activity of HIV-1 NC has been intensively studied. However, relatively little is known about the chaperone activity of other retroviral NCs. These four
retroviruses are chosen for a number of reasons. First, there are distinct structural differences between them, including differences in the number of zinc fingers, the number and location of aromatic and basic residues, and the overall charge of the proteins. Second, the R terminal-repeat regions among the genomes of these retroviruses differ significantly in length and structural complexity (shown in figure 2.4). The complementary R regions present in minus-strand DNA and viral RNA are annealed following minus-strand strong-stop DNA synthesis. Based on the structural differences in the R regions, the roles of the other retroviral NC proteins in minus-strand transfer and prevention of self-priming may differ significantly. Finally, the NC proteins studied in this work come from retroviruses belonging to four different genera, each of which is only distantly related to the other three.

**HIV-1 NCp7**

HIV-1 NCp7 is a 55-amino acid polypeptide with net charge of ~9, theoretical pl of ~10). NCp7 has two zinc fingers (ZF), each of which contains one aromatic residue (indicated in Fig. 2.3) interacting with

![Secondary structure of R region in retrovirus: HIV-1, RSV, HTLV-1](image_url)

Figure 2.4: secondary structure of R region in retrovirus: HIV-1, RSV, HTLV-1
nucleic bases. Basic residues on the N terminus account for nucleic acids aggregation and aromatic residues destabilize duplex (24,62).

NCp7 prefers binding to ssDNA 100 fold than dsDNA (63). Although it doesn’t bind to specific sequence, NCp7 prefers binding to TG rich sequence (60,64,65). The binding affinity of NC for ssDNA varies from $7\times10^3$ M$^{-1}$ for d(AAAAAA) to $7.1\times10^6$ M$^{-1}$ for d(TGTGCC) in 100 NaCl buffer (65). Strong binding of NCp7 to exposed guanines in stem-loop (SL) is critical for RNA packaging (66-68). Binding affinity of NCp7 to nucleic acids strongly depends on salt concentration due to its electrostatic interaction with nucleic acids (60,65).

HIV-1 NCp7 is essential to destabilize and anneal the complementary R domains of viral RNA (TAR-1) and the nascent ssDNA (-R sssDNA) (68). In the absence of HIV-1 NCp7, the efficiency of the specific strand transfer is significantly lowered. Using hydroxyl radical footprinting, gel-shift and fluorescence anisotropy binding assays, studies have shown that NC remains bound to the TAR RNA even after the heteroduplex formation (69). And the specific zinc-finger architecture is required for optimal binding to TAR-1, where ZF1 is critical for NC-TAR RNA aggregate formation (69).

**RSV NCp12**

As a retrovirus, RSV is the first described oncovirus, which caused sarcoma in chickens. Besides of gag, pol and env, RSV includes gene src, which is oncogenic and triggers uncontrolled cell growth

RSV NCp12 is longer than HIV-1 NCp7 and also highly positive with 16 positive residues and net charge 10 (pI~10). The first ZF contains 2 aromatic residues, while ZF2
has none (shown in Fig. 2.3). As NCp7, the second ZF is not as important as the first one for RNA binding and packaging (70-72). Basic residues in the central region and Arg61 (right after ZF2) are essential for RNA binding and efficient packaging (73). The N and C termini are dispensable if one of the zinc fingers is properly folded. RSV NCp12 doesn’t appear to have special affinity for specific sequences or helix destabilizing activities (74,75). It is the NC domain of unprocessed Gag precursor and not individual NC protein that recognizes the ψ packaging signal (75).

The R domain of viral RNA and nascent ssDNA are short and lack of extensive secondary structure. So the self-priming of R domain isn’t a critical issue for RSV. However, RSV NCp12 is not strictly required for the minus strand transfer.

**MoMLV NCp10**

MoMLV NCp10, like HIV-1 NCp7, is very small and basic (net charge ~9). It has only one zinc finger with two aromatic residues.

Only are the basic residues important when NCp10 binds poly(eA) and poly(U) non-cooperatively (76). The two aromatic residues cannot be interchanged, suggesting that the folding of the zinc finger is essential and only the Trp at position 35 can stack within the base pairs (77).

Studies show that MoMLV NCp10 can mediate the strand transfer of other viral R structures in this order: MoMLV~RSV>>HIV-1>>artificial R (78), where the R region was replaced by sequence of viral RNA from RSV RNA, HIV-1 or a 21-nt artificial R structure. This implies that the minus strand transfer of more stable structures involves more than just alignment of the two sequences. Later works showed that only the basic
residues were important for the annealing of R sequences in the minus strand transfer
*in vitro*, but the zinc finger was required *in vivo* (79,80).

**HTLV-1 NCp15**

HTLV-1, opposite to HIV-1, is very less aggressive and affects mainly two regions in the
world. There is little information available for HTLV-1 and its NCp15. Unlike its counterpart in
other retrovirus, NCp15 was found to bind weakly to both ss and ds DNA. The weak
binding of the NC protein is probably related to low infectivity of the virus.

HTLV-1 NCp15 is long and neutral (at pH 7.5) with two zinc fingers. There are two
aromatic residues in ZF1 and one in ZF2 (indicated in Fig.2.3). It was suggested that the
two zinc finger motifs interact with each other and with the acidic C-terminal domain
(CTD) through electrostatic interactions (81), which may be responsible for the low
cooperative binding of HTLV-1 NCp15 to ssDNA in low salt (1.3 mM Na\(^+\)). The salt
dependence of affinity to ssDNA showed a slope of only one, which meant only one Na\(^+\)
ion was released to surrounding solution. NCp15 prefers binding to ssDNA than ds DNA
and has slightly higher affinity for T (or U) rich sequence (81,82), which makes it a
possibly modest helix destabilizer.

While HIV-1 NCp7 can form polymers (63), HTLV-1 NCp15 forms only dimmers at
high concentrations (82). Maximal effect of HTLV-1 NCp15 on the viral RT activity was
obtained in the endogenous RT assay (using viral RNA), where NC was preincubated
with viral RNA (83). It suggests ability of NCp15 to arrange the secondary structures of
cDNA and R region of viral RNA during the minus strand transfer. The R region of HTLV-1
is much longer than those of other retroviruses and can form a very stable structure if
self-priming is not prevented (83,84). Study showed HIV-1 NCp7 was more important for preventing the nonspecific strand transfer than preventing self-priming. However, HTLV-1 in vivo may require the NCp15 to hold the viral RNA and the cDNA unfolded until the minus strand transfer is complete. So HTLV-1 NC tends to stabilize ssDNA.

It’s assumed that the catalytic activities of an NC protein during annealing and strand exchange are independent of its origin. We can substitute NC proteins in same retrovirus system to compare their chaperone activities. In this work we applied all four NC proteins in same assays, such as mini-TAR RNA annealing assay or aggregation assay.

2.2. Materials and Methods

2.2.1. Protein and nucleic acid preparation

HIV-1 NC was prepared either by solid-phase synthesis (85) or by recombinant methods, as described previously (56,86). The gene for RSV NC was PCR amplified from the RCASBP(B) vector, a generous gift from Stephen H. Hughes (National Cancer Institute-Frederick, Frederick, MD) (87) and cloned into pET32a (Novagen, Madison, WI) using standard molecular biology methods. The resulting clone, pDB1023, expressed RSV NCp12 as a thioredoxin fusion and used an enterokinase (DDDK) cleavage site to liberate authentic RSV NCp12. The expression plasmid for MOMLV NCp10, pDB1056, was prepared by insertion of the MOMLV NC sequence (GenBank accession no. J02255) into pET32a with an enterokinase site encoded between thioredoxin and NC. The gene encoding HTLV-1 NCp15 was PCR amplified from a full-length HTLV-1 proviral plasmid, pCS-HTLV-1 (88) (a generous gift from David Derse, National Cancer Institute-Frederick),
and cloned into pET32a to generate the plasmid pDR2559, which expresses HTLV NCp15 as a thioredoxin fusion with a tobacco etch virus protease cleavage site (ENLYFQ) (89,90) to liberate authentic HTLV-1 NC. RSV, MOMLV, and HTLV-1 NCs were expressed, isolated, and purified essentially as described previously (56,86). The NCs were stored in lyophilized forms at –80°C. Prior to use, the proteins were resuspended in diethylpyrocarbonate (DEPC)-treated water or NC storage buffer [20 mM HEPES, 5 mM β-mercaptoethanol, 0.1 mM Tris(2-carboxyethyl)phosphine hydrochloride, pH 7.5]. NC concentrations were determined by measuring the absorbance at 280 nm (A280) and using the following extinction coefficients: HIV-1 NC, 6,050 M⁻¹ cm⁻¹; RSV NC, 8,430 M⁻¹ cm⁻¹; MOMLV NC, 6,970 M⁻¹ cm⁻¹; HTLV-1 NC, 11,740 M⁻¹ cm⁻¹.

The mini-TAR RNA construct was obtained from Dharmacon (Lafayette, CO), and the mini-TAR DNA construct was obtained from Integrated DNA Technologies (Coralville, IA). The mini-TAR RNA and DNA oligonucleotides were purified on 16% denaturing polyacrylamide gels, dissolved in DEPC-treated water, and stored at –20°C. Oligonucleotide concentrations were determined by the A260 using the following extinction coefficients: mini-TAR RNA, 2.82 x 10⁵ M⁻¹ cm⁻¹; mini-TAR DNA, 3.06 x 10⁵ M⁻¹ cm⁻¹.

Mini-TAR RNA was radiolabeled with [γ-32P]ATP (Perkin-Elmer Life and Analytical Sciences, Waltham, MA) and T4 polynucleotide kinase (New England Biolabs, Ipswich, MA) using standard protocols. The radiolabeled TAR constructs were purified on 16% denaturing polyacrylamide gels, dissolved in DEPC-treated water, and stored at –20°C.
Prior to use, the mini-TAR oligonucleotides were refolded in 25 mM HEPES, pH 7.5, and 100 mM NaCl at a concentration that was 100-fold greater than the final assay concentration. The oligonucleotides were incubated at 80°C for 2 min and cooled to 60°C for 2 min, followed by the addition of MgCl2 to 10 mM and placement on ice for at least 5 min.

2.2.2. Annealing assays

Steady-state annealing assays were performed to determine the amount of NC required to reach saturating annealing levels. Solutions containing 15 nM refolded 32P-labeled mini-TAR RNA and 45 nM refolded mini-TAR DNA in reaction buffer (20 mM HEPES, pH 7.5, 20 mM NaCl, 5 mM dithiothreitol, and 0.2 MgCl2) were incubated for 5 min at 37°C. NC was then added to final concentrations varying from 0.5 to 5 µM, and the reaction mixtures were incubated for 30 min at 37°C. The reactions were quenched by incubation with 1% (vol/vol) sodium dodecyl sulfate (SDS) for 5 min at room temperature, followed by placement on ice. Samples were extracted twice with 4:1 phenol-chloroform and then analyzed on 15% SDS-polyacrylamide gels (375 mM Tris-HCl, pH 8.8, 0.1% SDS, 19:1 [wt/vol] acrylamide-bisacrylamide) run at 25°C in Tris-glycine (25 mM Tris, 250 mM glycine, pH 8.3) running buffer. Following overnight exposure to phosphor screens, the gels were visualized using a Bio-Rad Molecular Imager FX and analyzed using Bio-Rad Quantity One software.

Annealing assays to determine the kinetics of mini-TAR RNA and DNA annealing were performed in the presence of 1.5 µM HIV-1 NC (1.2 nt-to-NC ratio), 1.0 µM RSV NC (1.8 nt/NC), 3.0 µM MOMLV NC (0.6 nt/NC), or 3.0 µM HTLV-1 NC (0.6 nt/NC).
assays carried out over a 30-min period, 5 nM or 15 nM mini-TAR RNA and various concentrations of mini-TAR DNA (indicated in the figure legends) were incubated for 5 min at 37°C in reaction buffer prior to the addition of NC. Following the addition of NC, aliquots from the reaction mixture were removed and quenched by the addition of SDS to a final concentration of 1% (vol/vol), incubated at room temperature for 5 min, and then placed on ice. Samples were analyzed as described above. For assays carried out over a 120-min period, separate reaction mixtures containing mini-TAR RNA and mini- TAR DNA in reaction buffer were prepared and preincubated at 37°C (unless otherwise noted in the figure legends) prior to the addition of NC. The reactions were quenched at the specified time points and analyzed as described previously.

2.2.3. Sedimentation/aggregation assays

Solutions containing refolded 32P-labeled mini-TAR RNA (15 nM) and mini-TAR DNA (45 nM) were prepared in reaction buffer. NC was added to concentrations of 0.25, 0.5, 1.25, 2.5, and 5.0 µM (7.2, 3.6, 1.4, 0.7, and 0.4 nt/NC, respectively), and samples were incubated at 37°C for 30 min. The samples were then centrifuged at 13,400 x g (12,000 rpm; IEC Micromax RF) at 4°C for 20 min, and an aliquot of the supernatant (5 µl) was analyzed by scintillation counting. The percent radioactivity remaining in solution relative to a sample measured in the absence of NC (set to 100%) was plotted as a function of the NC concentration.
2.2.4. FA experiments

Equilibrium binding of HIV-1, RSV, MOMLV, and HTLV-1 NCs to a 6-carboxyfluorescein (FAM)-labeled 20-nt single-stranded DNA (ssDNA) oligonucleotide (5'-FAM-CTTCTTTGGAGTGAATTAG-3') was examined using FA. The high-performance liquid chromatography-purified DNA oligomer (5'-FAM DNA20) was from TriLink Biotechnologies (San Diego, CA). Binding to the corresponding 20-nt RNA/DNA hybrid duplex was also measured. The high-performance liquid chromatography-purified RNA oligomer (5'-CUAAUUCACUCCCAAAGAAG-3') was from Dharmacon. FA measurements were performed on an Analyst AD plate reader system (Molecular Devices, Sunnyvale, CA) using Corning 3676 low-volume 384-well black nonbinding-surface polystyrene plates. The reaction mixtures contained 20 nM 5'-FAM DNA20 or the RNA/DNA duplex, various concentrations of NC (0 to 10.2 µM, i.e., to 0.04 or 0.08 nt/NC for DNA or DNA/RNA experiments, respectively), and a final buffer consisting of 20 mM HEPES, pH 7.5, 50 mM NaCl, 5 mM β-mercaptoethanol, 1 µM ZnCl2, 100 µM Tris(2-carboxyethyl)phosphine, and 1 mM dithiothreitol. Samples were excited at 485 nm, and the emission intensities at 530 nm from the parallel and perpendicular planes were measured. Equilibrium dissociation constants (Kd) were determined by fitting the FA signal, A, as a function of the protein concentration, C, with respect to Kd using the following expression (62,91,92):

\[
A(C) = \frac{A_p + \Theta \cdot (A_p R - A_p)}{\Theta \cdot (R - 1) + 1}
\]  

(2.1)
where \( \Theta = \frac{1}{2D} \left[ D + C + K_d - \sqrt{(D + C + K_d)^2 - 4C \cdot D} \right] \) is the fraction of oligonucleotides bound, D is the oligonucleotide strand concentration, and AB and AF are the anisotropy values of the fully bound and unbound oligonucleotides, respectively. R is the ratio of the fluorescence intensity of saturated bound oligonucleotide relative to free oligonucleotide, which accounts for changes in fluorescence intensity upon NC binding (62,92,93). Equation (2.1) assumes 1:1 oligomer-protein binding.

2.2.5. Single-molecule DNA stretching

The optical-tweezers instrument, preparation of biotinylated \( \lambda \)-DNA, and data acquisition were described previously (24,25). The DNA was labeled at the 3' ends so that it would be free to rotate when stretched. The buffer used for the force-induced melting experiments contained 50 mM Na\(^+\) (10 mM HEPES, 45 mM NaCl, and 5 mM NaOH, pH 7.5). The protein equilibrium dissociation constant, \( K_d \), was estimated as described previously (62) using the following equation:

\[
\frac{1}{K_d} = \frac{1}{\delta F_{sat}} \cdot \frac{\delta F(c)}{c} = \frac{1}{\delta F_{sat}} \cdot \frac{d(\Delta F(c))}{dc}
\]

(2.2)

Where \( \delta F(c) \) is the change in the DNA force-induced melting transition width \( \Delta F(c) \) due to the addition of NC at the concentration \( c \), \( \delta F_{sat} \) is \( \delta F(c) \) at saturating protein binding, and \( \frac{d(\Delta F(c))}{dc} \) is the slope of \( \Delta F(c) \) versus \( c \) in the low-concentration (linear) region. The DNA force-induced melting transition width, \( \Delta F(c) \), is the force range.
where the force extension curve has the smallest slope corresponding to DNA elongation from its double-stranded to single-stranded form (62).

2.3. Results

2.3.1. Effects of different retroviral NC proteins on mini-TAR RNA/DNA annealing

Mini-TAR RNA/DNA annealing kinetics

The annealing of mini-TAR RNA and DNA hairpins in the presence of HIV-1, MOMLV, RSV, and HTLV-1 NCs (Fig. 2.4) was studied by gel shift annealing assays using 32P-labeled RNA (Fig. 2.5). The predicted secondary structures of HIV-1 TAR RNA (NL4-3 isolate) and TAR DNA hairpins are shown in Fig. 2.5. The mini-TAR constructs comprise the top portions of the TAR hairpins and contain the main structural features of TAR. Quantification of the band intensities provides the percentage of RNA molecules annealed as a function of time, P(t). In the absence of a protein chaperone, annealing of mini-TAR RNA and DNA involves the fast formation of an extended kissing-loop intermediate, followed by slow conversion to the product duplex (94). The apparent kinetics of mini-TAR RNA/DNA annealing can be described by the following equation (94):

$$P(t) = P_x \cdot \left( f \cdot \left(1 - e^{-k_1 t}\right) + (1 - f) \cdot \left(1 - e^{-k_2 t}\right) \right)$$  \hspace{1cm} (2.3)
Here, $k_f$ and $k_s$ are the high and low annealing rates, corresponding to formation of

the intermediate and the fully annealed duplex, respectively, and $f$ is the probability of intermediate formation. $P_e$ is the equilibrium final percentage of RNA annealed. Under some experimental conditions (e.g., in the presence of saturating levels of HIV, RSV, or MOMLV NC), the hairpin annealing kinetics becomes effectively single exponential due to fast formation of the stable intermediate. In these cases, the data are fitted to a single-exponential equation:

Figure 2.5: Predicted secondary structures of full-length TAR and mini-TAR RNA and DNA hairpins. The sequences are derived from the HIV-1 NL4-3 isolate. The mini-TAR constructs are derived from the top part of the hairpin (dotted boxes). Mini-TAR DNA contains an additional 5-nt single-stranded overhang. The secondary structures were predicted by m-fold analysis. (Inset) Typical mini-TAR gel shift analysis performed as a function of time. The arrow labeled F indicates the position of free mini-TAR RNA, and the arrow labeled B points to the mini-TAR RNA/DNA binary complex.
\[ P(t) = P_\infty \cdot (1 - e^{-kt}) \]  

(2.4)

where \( k \) is the annealing rate.

Figure 2.6: Annealing time courses for mini-TAR RNA/DNA hairpins. Annealing was performed in the presence of 15 nM RNA and 90 nM DNA at 37°C. The concentrations of NCs used were 1.5 \( \mu M \) HIV-1, 1 \( \mu M \) RSV, 3 \( \mu M \) MoMLV, and 3 \( \mu M \) HTLV-1. The curves represent single-exponential fits to equation 4 with variable final percent annealed product. The error bars indicate standard deviations.

Figure 2.6 compares the annealing of mini-TAR RNA/DNA hairpins in the presence of saturating concentrations of HIV-1, RSV, MOMLV, and HTLV-1 NCs. The amount of NC required to achieve saturated annealing is indicated in the legend to Fig. 2.6 and was determined by steady-state annealing assays, as described in Materials and Methods. These data suggest that the chaperone activity decreases in the order HIV-1 NC \( \cdot \)/RSV NC > MoMLV NC > HTLV-1 NC. The effective annealing rates and final percentages of RNA annealed differ considerably for the four NCs studied. Annealing in
the presence of RSV NC results in higher plateau levels of duplex product than for HIV-1 NC, and an annealing rate, $k$, of $\sim 1.0 \text{ min}^{-1}$, which is lower than the $k$ observed for HIV-1, $\sim 1.6 \text{ min}^{-1}$ (Table 2.1). For MoMLV NC, the annealing rate, $0.15 \text{ min}^{-1}$, final annealing levels were markedly reduced relative to those observed for HIV-1 NC and RSV NC. Addition of HTLV-1 NC to the annealing reaction resulted in only minor stimulation of annealing relative to the reaction performed in the absence of protein, with an annealing rate of $0.008 \text{ min}^{-1}$. Since HTLV-1 NC is a very inefficient nucleic acid chaperone, the annealing kinetics was not investigated further using this method.

To gain further insights into the effects of HIV-1, RSV, and MoMLV NCs on mini-TAR RNA/DNA annealing, assays were performed in the presence of various DNA concentrations ($D$) (Figure 2.7a to c). Rates of mini-TAR RNA/DNA annealing were plotted as a function of $D$ (Fig. 2.8), and the effective bimolecular rate constant, $k_{\text{eff}}$, was

<table>
<thead>
<tr>
<th></th>
<th>$k$</th>
<th>$k_{\text{eff}}$</th>
<th>$\Delta H^\ddagger$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(min$^{-1}$)</td>
<td>(M$^{-1}$s$^{-1}$)</td>
<td>(kcal/mol)</td>
</tr>
<tr>
<td>no NC</td>
<td>0.001</td>
<td>$(1 \pm 0.5) \times 10^2$</td>
<td>19$\pm$3</td>
</tr>
<tr>
<td>HIV-1 NC</td>
<td>1.6$\pm$0.3</td>
<td>$(3 \pm 2) \times 10^5$</td>
<td>8.3$\pm$2</td>
</tr>
<tr>
<td>RSV NC</td>
<td>1.0$\pm$0.2</td>
<td>$(6 \pm 2) \times 10^4$</td>
<td>12.8$\pm$2</td>
</tr>
<tr>
<td>MoMLV NC</td>
<td>0.15$\pm$0.05</td>
<td>$(3 \pm 2) \times 10^4$</td>
<td>10.5$\pm$3</td>
</tr>
<tr>
<td>HTLV-1 NC</td>
<td>0.008</td>
<td>ND</td>
<td>ND</td>
</tr>
</tbody>
</table>

Table 2.1: Kinetic parameters for annealing of mini-TAR RNA/DNA in the absence and presence of different retroviral NCs
obtained from the initial slopes of these plots.

For mini-TAR RNA/DNA annealing in the absence of NC (94), the dominant low rate of annealing was proportional to D, and the effective bimolecular association rate constant was \( \sim 100 \text{ M}^{-1} \cdot \text{s}^{-1} \) (Table 2.1). Addition of saturating amounts of HIV-1 NC resulted in \( \sim 1,000 \)-fold overall rate enhancement (Table 2.1), leading to a \( k_{\text{eff}} \) of \( \sim 3 \times 10^5 \text{ M}^{-1} \cdot \text{s}^{-1} \). At low D (<60 nM), the annealing rate with RSV NC was higher than with HIV-1 NC (Fig. 2.8). However, at higher D, the opposite was true, resulting in a \( k_{\text{eff}} \) of \( \sim 6 \times 10^4 \text{ M}^{-1} \cdot \text{s}^{-1} \) for RSV NC (Table 2.1).

For MoMLV NC, the annealing at all D values was slow compared to annealing with either HIV-1 or RSV NC (Fig. 2.7c). Both the apparent reaction rate, \( k \), and the \( k_{\text{eff}} \)
were significantly smaller (Table 2.1 and Fig. 2.8), suggesting that MoMLV NC is a relatively poor nucleic acid chaperone compared to HIV-1 or RSV NC.

Figure 2.8: Rates of mini-TAR RNA/DNA annealing as a function of the mini-TAR DNA concentration in the absence and presence of 1.5 µM HIV-1 NC, 1 µM RSV NC, and 3 µM MOMLV NC. The lines are guides for the eye, emphasizing linear k-versus-D relationship at lower D values. The rates were derived from annealing time courses similar to the ones presented in Fig. 2.7.

Temperature dependence of mini-TAR RNA/DNA annealing

Mini-TAR hairpin annealing rates were enhanced with increasing temperature in both the absence and presence of HIV-1 NC, suggesting that melting of some portion of the mini-TAR hairpin secondary structure is in pre-equilibrium to bimolecular association (104). A similar enhancement of annealing rates with increased temperature was observed in the presence of RSV and MOMLV NCs. The effect of increased temperature on mini-TAR RNA/DNA annealing in the presence of MoMLV NC is shown
Fig. 2.9: Temperature dependence of mini-TAR RNA/DNA annealing kinetics. a) Annealing time course of 15 nM mini-TAR RNA and 45 nM mini-TAR DNA in the presence of 3 µM MoMLV NC at four different temperatures, as indicated in the figure. b) Annealing rates plotted as a function of the reciprocal temperature in the absence and presence of 1.5 µM HIV-1 NC, 1 µM RSV NC, and 3 µM MoMLV NC. The lines are the fits of the data points to the Arrhenius expression (equation 2.5) with slopes equal to the enthalpies of the rate-limiting step of annealing.

Fig. 2.9a. The enthalpy values for the rate-limiting step of annealing, $\Delta H^\ddagger$, were obtained according to the Arrhenius expression:

$$\Delta H^\ddagger = -RT \frac{d\left(\ln k\right)}{d\left(\frac{1}{T}\right)} \tag{2.5}$$

where $R$ is the molar gas constant and $T$ is temperature in Kelvin. The Arrhenius analysis is shown in Fig. 2.9b. $\Delta H^\ddagger$ is the enthalpy associated with the melting of base pairs that must occur in order for the binary complex to nucleate. The data indicate that all retroviral NCs produce similar reductions in melting enthalpy relative to the no-NC reaction (Table 2.1). These results are consistent with the prediction that all retroviral NCs bind with more favorable enthalpy to single-stranded nucleic acids than to double-
stranded nucleic acids.

**Figure 2.10**: Free energy of mini-TAR RNA/DNA hairpin annealing in the presence of HIV-1, RSV, and MoMLV NCs. The data points were calculated from the measured equilibrium annealing levels according to equation 2.6. a) D dependence of annealing free energy at 37°C. The lines are \( \Delta G(D) \) dependencies calculated according to the second equality in equation 2.6, with \( K_{m}^{\text{TAR}} \) as a fitting parameter. The fitted \( K_{m}^{\text{TAR}} \) values are reported in Table 2.2. b) Temperature dependence of annealing free energy. The lines are linear fits of the experimental data points to equation 2.7. The line for MoMLV NC is dashed. The free energy of annealing of mini-TAR RNA/DNA hairpins in the absence of NC was calculated for the given solution conditions using DINAmelt (72). The fitted values of \( \Delta H \) and \( \Delta S \), along with \( \Delta G^{37} \), are summarized in Table 2.2, where \( \Delta G^{37} \) is the annealing free energy at 37°C.

**Equilibrium stability of the mini-TAR RNA/DNA duplex**

The annealing of mini-TAR RNA/DNA to form a complete duplex is driven by formation of nine new base pairs but is opposed by the entropy loss of one of the hairpins. Decreased stability of the annealed duplex, or lowering of the hairpin concentration, leads to a shift in equilibrium toward the reactants. In the absence of NC at long reaction times (i.e., hours or even days) (94), equilibrium plateau levels of annealing are achieved that approach 100%. Therefore, the time courses for mini-TAR RNA/mini-TAR DNA annealing in the absence of NC are always fitted to equation (2.3),
with $P_\infty$ equal to 100. In contrast, in the presence of NC proteins, annealing remains incomplete at equilibrium due to NC-induced destabilization of the annealed duplex. In other words, NC proteins facilitate nucleic acid dissociation to an even greater extent than annealing. Therefore, while the equilibrium in the presence of NC proteins is achieved much faster than in the absence of protein, the amount of product is smaller.

In this case, the percentage of mini-TAR RNA annealed at equilibrium, $P_\infty$, can be used to estimate the free energy of annealing:

$$\Delta G = -RT \cdot \ln \left( \frac{P_\infty}{1 - P_\infty} \right) = -RT \cdot \ln \left( \frac{D}{K_{d,mTAR}} \right),$$

where $K_{d,mTAR}$ is the dissociation constant of the fully annealed duplex.

The free-energy values for mini-TAR annealing calculated from measured $P_\infty$ values according to equation (2.6) are presented in Fig. 2.10a as a function of $D$. The lines in the figure represent fits of these data using the second equality in equation (2.6) and $K_{d,mTAR}$ as the fitting parameter. The $K_{d,mTAR}$ values obtained suggest that the annealed duplex in the presence of the retroviral NCs is relatively stable, with the highest stability observed for RSV NC (Table 2.2). The theoretical $\Delta G(D)$ dependence (i.e., the lines in Fig. 2.10a) describes the measured $\Delta G(D)$ values (i.e., the symbols in Fig. 2.10a) quite well. This result supports the hypothesis that the change in equilibrium annealing with $D$ in the presence of NC is a result of changing stability of the annealed duplex. The accuracy of the $\Delta G(D)$ determination is not very high due to variability in
measured $P_\infty$ levels between individual assays. Nevertheless, the $\Delta G$ values provide insights into the relative duplex-destabilizing abilities of HIV-1, RSV, and MoMLV NCs. The annealed mini-TAR duplex is most stable in the presence of RSV NC, suggesting that RSV NC does not destabilize nucleic acid base pairs as efficiently as HIV-1 NC. HIV-1 NC and MoMLV NC appear to be very similar in their duplex-destabilizing abilities.

The effects of the different NCs on the enthalpy, $\Delta H$, and entropy, $\Delta S$, of annealing were determined by measuring the equilibrium annealing levels, $P_\infty$, of mini-TAR as a function of temperature. The $P_\infty$ values were used to calculate the annealing free energies, $\Delta G$, according to equation 2.6, and then $\Delta G$ was fitted as a function of temperature to the general thermodynamic relationship:

$$\Delta G(T) = \Delta H - T\Delta S$$

(2.7)

The estimated $\Delta G$ values, along with the fits to equation (2.7), are presented in Fig. 2.10b, and the resulting thermodynamic parameters are summarized in Table 2.2. As observed for HIV-1 NC, MoMLV NC strongly reduced the melting enthalpy and entropy of the annealed duplex. The effects of HIV-1 NC and MoMLV NC on the thermodynamic parameters of duplex melting were almost indistinguishable from each other within the accuracy of these measurements (Fig. 2.10b and Table 2.2). While RSV NC also significantly reduced the melting enthalpy, it had less of an effect on the melting entropy than HIV-1 or MoMLV NC. As discussed previously, the physical reason for the reduction in $\Delta H$ and $\Delta S$ in the presence of NC is likely due to the favorable enthalpy of NC interaction with single-stranded nucleic acids, which stabilizes the melted state,
coupled with a loss of entropy upon NC binding to single-stranded nucleic acids (94).

The combination of these two effects results in moderate duplex destabilization, since the protein's effect on the melting enthalpy is slightly stronger than its effect on the entropic component. Thus, the results obtained here suggest that RSV NC may not stabilize the melted state as effectively as HIV-1 NC and that RSV NC-bound single-stranded nucleic acids are more mobile in solution. In summary, analysis of the equilibrium mini-TAR RNA/DNA annealing levels in the presence of HIV, RSV, and MoMLV NCs suggests that the duplex-destabilizing abilities of NC proteins decrease in the following order: HIV ~MoMLV > RSV.

2.3.2. Effects of different retroviral NC proteins on mini-TAR RNA/DNA aggregation

<table>
<thead>
<tr>
<th></th>
<th>( P_\infty (%) )</th>
<th>( \Delta G^{37} (\text{kcal/mol}) )</th>
<th>( K_{d}^{\text{mini-TAR}} (\text{M}) )</th>
<th>( \Delta H (\text{kcal/mol}) )</th>
<th>( \Delta S (\text{kcal/mol}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>no NC</td>
<td>100</td>
<td>-6.7</td>
<td>(&lt;10^{-7})</td>
<td>-77.6</td>
<td>-232</td>
</tr>
<tr>
<td>HIV-1 NC</td>
<td>76±5</td>
<td>-0.7±2</td>
<td>((1.6±0.1)\times10^{-8})</td>
<td>-8.5±4</td>
<td>-25±12</td>
</tr>
<tr>
<td>RSV NC</td>
<td>95±5</td>
<td>-1.5±2</td>
<td>((0.5±0.1)\times10^{-8})</td>
<td>-14±8</td>
<td>-92±15</td>
</tr>
<tr>
<td>MLV NC</td>
<td>66±5</td>
<td>-0.8±2</td>
<td>((2.0±0.2)\times10^{-8})</td>
<td>-5.2±3</td>
<td>-14±18</td>
</tr>
</tbody>
</table>

Table 2.2: Thermodynamic parameters for annealing of mini-TAR RNA/DNA in the absence and presence of HIV-1, RSV, and MLV NCs.
To directly examine the aggregation abilities of the four retroviral NCs, a sedimentation assay was used to measure the fraction of aggregated mini-TAR molecules as a function of the NC concentration under the same conditions used in the gel shift annealing assays (Fig. 2.11). The results suggest that all NC proteins, except for HTLV-1 NC, effectively aggregate nucleic acids upon saturated binding at ~1 µM. The results obtained for HIV-1 NC are consistent with previous findings (48,49,94). RSV NC appears to be slightly more effective at aggregating nucleic acids than HIV-1 NC, and these results are consistent with its ability to effectively facilitate the bimolecular-association step (k values in Table 2.1). MoMLV NC also exhibits good aggregation activity at concentrations of ~1 µM, despite the fact that this NC is significantly less effective at facilitating the bimolecular-association step than HIV-1 NC or RSV NC (compare the k values in Table 2.1). Interestingly, a previous study showed that the morphologies of nucleic acid aggregates formed in the presence of HIV-1 NC differ from those of MoMLV NC (95).

Strikingly, HTLV-1 NC lacks any detectable

![Figure 2.11](image.png)

**Figure 2.11:** Percents of mini-TAR RNA aggregated from a solution of 15 nM 32P-labeled mini-TAR RNA and 45 nM mini-TAR DNA at 37°C as a function of the NC concentration.
nucleic-acid-aggregating activity in the range of concentrations studied (i.e., up to 5 µM NC). In contrast to the other NC proteins, which are highly charged cationic proteins, HTLV-1 NC is neutral (pI, ~7) at physiological pH. The C-terminal domain of HTLV-1 NC contains a high density of acidic residues, whereas the majority of the basic residues are concentrated in the N terminus and zinc finger domains (Fig. 2.3). Therefore, HTLV-1 NC’s inability to aggregate nucleic acids is consistent with the electrostatic model of NC-induced nucleic acid aggregation (96,97).

2.3.3. Nucleic acid binding studies of different retroviral NC proteins

FA was used to examine the binding of HIV-1, RSV, MoMLV, and HTLV-1 NCs to a 20-nt ssDNA oligonucleotide and a 20-bp double-stranded RNA/DNA (dsRNA/DNA) hybrid duplex (Fig. 2.12). The \( K_d \) values determined from these measurements are listed in Table 2.3. The \( K_d^{ss} \) and \( K_d^{ds} \) values for the NC proteins differ from each other by less than \(~4\)-fold and \(~7\)-fold, respectively. In addition, for each NC studied, the \( K_d^{ss} \) is two- to threefold smaller than \( K_d^{ds} \), with the exception of HTLV-1 NC, where the
difference is closer to four fold. The binding strengths of the proteins decrease in the order HIV-1 NC > RSV NC > MoMLV NC > HTLV-1 NC for both the ssDNA and the dsRNA/DNA duplex. These binding measurements were carried out using a fairly random nucleic acid sequence in order to minimize sequence-specific binding effects.

Although NCs generally bind nucleic acids nonspecifically, HIV-1 NC (64-67,93,98,99) and MoMLV NC (100-102) have an ~10-fold preference for single-stranded TG- or UG-rich sequences, and RSV NC has a preference for G residues (65,100,103-106).

2.3.4. Duplex destabilization by retroviral NC proteins

Stronger binding to single-stranded versus double-stranded nucleic acids is characteristic of proteins that induce duplex destabilization (107). Assuming a similar site size of 6 nt (or 3 bp) (60-62,64,65,81,93,108-111) for binding to single-stranded and double-stranded sequences, the maximum duplex destabilization free energy by NC protein, \( \delta G \), per base pair can be approximated according to equation 2.8 (107):

\[
\delta G = \frac{2}{3} \cdot RT \cdot \ln \left( \frac{K_{ss}^{d}}{K_{ds}^{d}} \right)
\]

(2.8)

Values for the base pair destabilization free energies were calculated using equation 2.8, and the \( K_{ss}^{d} \) and \( K_{ds}^{d} \) values are given in Table 2.3. The destabilization free energy per base pair, \( \delta G \), values for different retroviral NC proteins range between \(-0.3\) and \(-0.5\) kcal/mol/bp (Table 2.3). For HIV-1 NC, this estimate is consistent with other studies (when extrapolated to the present salt conditions) (24,25,50-52,54,55,58,61,62,112). To
the best of our knowledge, there are no reported estimates of this quantity for the other retroviral NCs.

Interestingly, the NC-induced destabilization free energies calculated for the four retroviral NCs studied here are rather similar, with the \( \delta G \) value for HIV-1 NC at the lower end and that for HTLV-1 NC at the higher end of the range. The accuracy of \( \delta G \) is limited by errors in the \( K_d \) measurements. The FA curves obtained were fitted assuming a single binding site per oligonucleotide (see Materials and Methods), although it is likely that there are multiple NC binding sites, especially considering that 20-nt single-stranded and double-stranded sequences were used. Another source of uncertainty in \( \delta G \) is the assumption in equation 2.8 that the NC binding site sizes for single-stranded and double-stranded nucleic acids are the same. Nevertheless, these

<table>
<thead>
<tr>
<th>NC</th>
<th>( K_{ds}^{ss} ), nM (FA)</th>
<th>( K_{ds}^{ds} ), nM (FA)</th>
<th>( \delta G ) (kcal/mol/bp)</th>
<th>( K_d ), nM (DNA stretching)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIV-1</td>
<td>123±9</td>
<td>233±11</td>
<td>-0.26±0.06</td>
<td>10±5</td>
</tr>
<tr>
<td>RSV</td>
<td>200±49</td>
<td>660±47</td>
<td>-0.46±0.06</td>
<td>80±10</td>
</tr>
<tr>
<td>MLV</td>
<td>334±34</td>
<td>1092±28</td>
<td>-0.46±0.06</td>
<td>150±20</td>
</tr>
<tr>
<td>HTLV-1</td>
<td>431±87</td>
<td>1625±270</td>
<td>-0.52±0.06</td>
<td>300±50</td>
</tr>
</tbody>
</table>

Table 2.3: Duplex destabilization free energies and apparent equilibrium dissociation constants for binding of retroviral NC proteins to single-stranded and double-stranded nucleic acids
estimates of NC-induced duplex destabilization are generally consistent with the results of the annealing assays (Fig. 2.10). Due to its weak overall annealing activity, a detailed kinetic analysis of HTLV-1 NC could not be carried out using the gel shift assay, yet interestingly, this retroviral NC protein has the lowest $\delta G$ value and is therefore expected to display stronger duplex destabilization activity than HIV-1 NC (Table 2.3). Ensemble and single-molecule fluorescence resonance energy transfer-based studies have been carried out to directly probe the destabilization activity of HTLV-1 NC, and the results are consistent with this hypothesis (M. Mitra, I. Rouzina, and K. Musier-Forsyth, unpublished observations; Q. Darugar, H. Kim, R. J. Gorelick, and C. Landes, unpublished data).

Taken together, the average duplex-destabilizing effect measured for all four retroviral NC proteins was as follows: $\delta G = -0.43 \pm 0.1$ kcal/mol/bp. This is a rather moderate effect, considering that the average melting free energy per base pair at 37°C is ~2 kcal/mol/bp for DNA and ~3 kcal/mol/bp for RNA duplexes (113,114). However, it is sufficient to lead to the strong effects of some NC proteins on both the kinetics and equilibrium annealing levels of the TAR hairpins. The overall effect of a chaperone protein on duplex stability is expected to be similar to the effect of increasing the solution temperature according to equation 2.9:

$$\delta T = \frac{\delta G}{\Delta S} = \frac{0.43 \text{kcal/mol/bp}}{0.025 \text{kcal/mol/bp/K}} = 18^\circ C$$

where the entropy of melting, $\Delta S$, is assumed to be 0.025 kcal/mol/K/bp (114,115). Thus, the addition of saturating NC has the same effect on duplex stability as heating
the solution by ~18°C, which is an alternative way to facilitate nucleic acid refolding without melting the final lowest energy state (116).

2.3.5. DNA-stretching studies with different retroviral NC proteins

The small differences in the duplex-destabilizing abilities of the four NC proteins cannot account for the marked differences in their chaperone activities (Fig. 2.6). In addition, with the exception of HTLV-1 NC, the NC proteins studied here exhibited similar aggregation abilities. To gain further insights into the differences in the thermodynamics and kinetics of NC-nucleic acid interactions, we performed single-molecule DNA-stretching experiments (23-25,29-31,55,62,97,117-122). In these studies, double-stranded λ-DNA is tethered between two beads, and force is applied to stretch the DNA through the helix-coil transition. The effect of HIV-1 NC on the force-induced melting transition of λ-DNA has been extensively studied (24,25,55,62,117). Briefly, binding of HIV-1 NC results in significant broadening of the melting transition (compared to the DNA-only curve, which displays a sharp plateau at ~60 pN) and a reduction in the force required to stretch λ-DNA (Fig. 13a). The small hysteresis, that is, the lack of an exact match between the stretching and relaxation curves, is a sign of the fast kinetics of HIV-1 NC binding and dissociation to ds- and ssDNA on the time scale of the experiment. The high reproducibility of the stretching curves with repeated cycles and various pulling rates is indicative of an equilibrium process in which HIV-1 NC is able to rapidly adjust to changes in DNA structure (55).

The DNA stretching curves in the absence and presence of saturating amounts of each of the four NC proteins examined here are presented in Fig. 2.13. RSV NC binding
results in a dsDNA melting transition that remains highly cooperative and occurs at a force that exceeds the melting force measured in the absence of protein by \(-5\) pN (Fig. 2.13b). A possible mechanism for this apparent duplex stabilization may involve changes in the elastic properties of ssDNA upon RSV NC binding. Similar to HIV-1 NC, RSV NC produces little hysteresis in the stretching cycle upon saturated DNA binding.

Figure 2.13: \(\lambda\)-DNA stretching and relaxation curves in the presence of saturating levels of HIV-1 (top left), RSV (top right), MOMLV (bottom left), and HTLV-1 (bottom right) NCs. In each panel, the thin lines represent the curves in the absence of protein. Experiments in the presence of NC are shown by the thick black lines. The experiments were performed in 45 mM NaCl and 10 mM HEPES, pH 7.5, for a total of 50 mM Na\(^+\). In each cycle shown, the upper curve represents the stretching and the lower curve is the relaxation.

In addition, RSV NC also produces a small but reproducible increase in the force at DNA extensions below the \(\lambda\)-DNA contour length (0.34 nm/bp). These two features signify
fast kinetics of protein-DNA interaction and the ability to very effectively aggregate DNA, respectively, and are consistent with RSV NC's efficient facilitation of mini-TAR RNA/DNA annealing (Table 2.1 and Fig. 2.9 and 2.12).

Saturation with MoMLV NC results in broadening of the DNA melting transition and a large hysteresis that increases at higher NC concentrations (Fig. 13c). This large hysteresis indicates slow reannealing of the DNA strands after the release of tension. MoMLV NC destabilizes duplexes (Fig. 2.1 and Table 2.1) and aggregates nucleic acids quite efficiently (Fig. 2.12). Based on the results of the DNA-stretching assay, the reduced chaperone activity of MoMLV NC appears to be due to the relatively slow kinetics of MoMLV NC dissociation from single-stranded nucleic acids, which would be expected to inhibit duplex-closing events.

The effect of HTLV-1 NC on the DNA-stretching profile differs considerably from those of HIV-1, RSV, and MoMLV NCs (Fig. 13d). HTLV-1 NC has little effect on the transition width but reduces the force required for DNA melting. Strikingly, the HTLV-1 NC-stretching curve exhibits the greatest amount of hysteresis observed for any NC we have studied. This result implies that HTLV-1 NC preferentially binds ssDNA and, once bound, dissociates slowly, thereby completely preventing strand reannealing on the time scale of the experiment. The features of the HTLV-1 NC DNA stretching and relaxation curves are very similar to the DNA-stretching behavior in the presence of ssDNA binding proteins (SSBs), such as bacteriophage T4 gene 32 protein (gp32) (118-121) and bacteriophage T7 gene 2.5 protein (gp2.5) (23). DNA-stretching studies have demonstrated that SSBs bind dsDNA very poorly and dissociate from ssDNA very slowly.
In the case of gp32, its slow dissociation is due to the cooperativity of protein binding to ssDNA resulting from "head-to-tail" protein-protein interactions (119).

In the case of HTLV-1 NC, it is plausible that the bound proteins interact in a similar fashion, with the cationic N-terminal domain of one NC binding to the anionic C-terminal domain of an adjacent NC molecule. Previously, fluorescence studies demonstrated that HTLV-1 NC binds nucleic acids with modest cooperativity (81). Such binding cooperativity might explain the slow dissociation of HTLV-1 NC from ssDNA.

Taken together, these data suggest that the SSB-type properties of HTLV-1 NC are responsible for its poor performance as a nucleic acid chaperone.

DNA-stretching results were also used to estimate equilibrium dissociation constants (Table 2.3, column 5). These values are determined by measuring the slope of the linear region of the force transition width as a function of the concentration (equation 2.2) (see Materials and Methods) (1,55,62) and reflect a composite of affinities of binding to ssDNA and dsDNA. When the protein binds ssDNA more strongly than dsDNA, the $K_d$ values primarily reflect protein interactions with ssDNA. In accordance with the preference of all retroviral NCs for binding to ssDNA, the $K_d$ values obtained from the stretching curves are more closely correlated with the values measured by FA analysis for single-stranded binding (Table 2.3, column 2) than for double-stranded binding (column 3). Interestingly, although the absolute values differ, the relative trends in strengths of binding to polymeric DNA between the NC proteins is similar to the relative affinities measured for binding to a 20-nt ssDNA oligonucleotide (i.e., HIV-1 > RSV > MoMLV > HTLV-1 NC).
2.4. Discussion

In this work, the chaperone activities of HIV-1, RSV, MoMLV, and HTLV-1 NCs were characterized using complementary experimental approaches. Gel shift annealing assays demonstrated that the overall chaperone activities of the NC proteins studied decrease in the order HIV-1 NC > RSV NC > MoMLV NC >> HTLV-1 NC (Fig. 2.6). Furthermore, detailed analysis of hairpin-annealing time courses as a function of the DNA concentration and temperature provide additional insights into how NC proteins differ in their nucleic acid chaperone activities. RSV NC effectively facilitates the bimolecular-association step of mini-TAR RNA/DNA annealing, suggesting that it is a better aggregating agent than HIV-1 NC. Overall, MoMLV NC is less efficient in facilitating both bimolecular association and conversion to the extended duplex. Thermodynamic analysis suggests that HIV-1 and MoMLV NCs are similar to each other and only slightly better than RSV NC at duplex destabilization, while all three reduce the melting enthalpy and entropy of the duplex.

Sedimentation assays demonstrated that HIV-1, RSV, and MoMLV NCs efficiently aggregate nucleic acids under saturating conditions (>1 μM protein), whereas HTLV-1 NC lacks nucleic-acid-aggregating capability. We hypothesize that HTLV-1 NC is unable to aggregate nucleic acids due to its overall neutral charge. Experiments in our laboratory have demonstrated that removal of the C-terminal cationic domain of HTLV-1 NC restores aggregation and chaperone activity (K. Stewart-Maynard, I. Rouzina, and K. Musier-Forsyth, unpublished data). Furthermore, measurements of equilibrium binding of the four NC proteins to 20-nt ssDNA and a dsRNA/DNA hybrid duplex suggest that the
duplex-destabilizing abilities of the NC proteins studied here are similar. Interestingly, despite its poor overall chaperone activity, HTLV-1 NC appears to be a very effective duplex-destabilizing agent.

The observed differences in duplex destabilization and nucleic acid aggregation cannot fully account for the differences in nucleic acid chaperone activities between the four retroviral NC proteins. For instance, saturating concentrations of MoMLV NC aggregate nucleic acids as effectively as HIV-1 and RSV NC, and MoMLV’s duplex destabilization capability is also similar to those of HIV-1 and RSV NCs. However, gel shift assays demonstrated that MoMLV NC does not facilitate mini-TAR RNA/DNA annealing as efficiently as HIV-1 or RSV NC. The rates of protein dissociation from ssDNA, revealed by single-molecule DNA-stretching experiments, decrease in the order RSV NC $\geq$ HIV-1 NC $\gg$ MoMLV NC $\gg$ HTLV-1 NC, which closely parallels their chaperone activities (Fig. 2.6). These results support the notion that the fast kinetics of protein-nucleic acid interaction is another major component of the NC nucleic acid chaperone function (55).

NC’s ATP-independent chaperone function requires stoichiometric binding to facilitate nucleic acid aggregation and maximal duplex destabilization (94). High-affinity binding, in combination with fast kinetics, is a feature of cationic binding to nucleic acids in a predominately electrostatic mode (11,12,123-129). Such electrostatic binding leads to the formation of protein-induced nucleic acid aggregates that are characterized by high mobility of nucleic acid and protein (97,123,124,126,127,130-134), thereby strongly facilitating the bimolecular steps of nucleic acid annealing. However, purely electrostatic
binding results in stabilization of the double-stranded, rather than the single-stranded, form of nucleic acids (113), whereas duplex destabilization requires preferential binding to the single-stranded form. For HIV-1 NC, destabilization is facilitated, in part, by partial stacking of two conserved aromatic residues within the zinc finger domains (Fig. 2.4) with nucleic acid bases (66,67,98,99). RSV (105,106) and MoMLV (100,102) NC proteins also engage in similar interactions. While this binding leads to duplex destabilization, it is also expected to result in reduced dissociation kinetics. HIV-1 NC appears to have optimized these contradictory properties; it is a moderate duplex destabilizer, displaying rapid nucleic acid interaction kinetics and high aggregating ability. RSV NC is an even better aggregating agent and exhibits rapid binding kinetics, but it appears to be a relatively poor duplex destabilizer (Fig. 2.10, 2.11, and 2.13 and Table 2.1). MoMLV NC aggregates and destabilizes duplexes effectively (Fig. 2.10). However, due to its slow dissociation kinetics (Fig. 2.13), the protein is a less effective chaperone (Fig. 2.6).

Recent studies support the conclusion that MoMLV NC is a less effective annealing agent than HIV-1 NC, despite the fact that it has similar binding and destabilization capabilities (135). MoMLV NC was also shown to reduce the rate of HIV-1 reverse transcription (136) and to induce DNA aggregates of lower density than HIV-1 NC (95).

The chaperone properties of HTLV-1 NC differ substantially from those of the other retroviral NCs. Despite the fact that this deltaretroviral NC is a relatively strong duplex destabilizer, the neutral protein does not cause nucleic acid aggregation. Our DNA-stretching data suggest that HTLV-1 NC's extremely slow nucleic acid dissociation kinetics is likely the major factor leading to the poor chaperone properties observed in
vitro (Fig. 2.6). The implications of these findings for RNA packaging and reverse transcription in vivo remain to be explored, and additional studies using HTLV-1 NC protein variants are currently under way.

In this work, we have investigated the effects of different retroviral NC proteins on a reaction that models the TAR RNA/DNA annealing that occurs during the minus-strand transfer step of HIV-1 reverse transcription. However, the R regions of the four retroviruses differ in length, sequence, and structural stability. For example, the R region of RSV is only 21 nt long and lacks significant secondary structure (78). Therefore, one can speculate that the somewhat weaker duplex-destabilizing ability of RSV NC coevolved with the unstructured R region of the retrovirus. The 68-nt R region of MoMLV is comparable in length to the 98-nt R region of HIV-1 and folds into several hairpins with stability comparable to that of HIV-1 TAR. This observation is consistent with the similar duplex-destabilizing capabilities of these NC proteins.

Given the poor chaperone activity of HTLV-1 NC relative to the other retroviral NCs examined here, it is surprising that the HTLV-1 R region contains 228 nt and folds into a complex secondary structure (84). The chaperone function of HTLV-1 NC may be regulated by other factors in vivo, possibly involving a conformational change of the protein's unique acidic C-terminal domain. Alternatively, other host or viral proteins may be involved in chaperoning key steps in reverse transcription in this retrovirus. Clearly, additional studies of the chaperone activities of different NC proteins using substrates that mimic their respective R regions are needed to better understand the biological roles of the observed differences in chaperone properties.
3. Electrostatic interaction between the anionic C-terminal domain of HTLV-1 NC with its cationic N-terminal domain leads to the slow nucleic acid-protein interaction and poor chaperone activity

This work was the result of a collaboration with Karin Musier-Forsyth’s lab. Bulk experiments were performed by Karin Musier-Forsyth’s lab and single molecule experiments were by Fei Wang. It has been reformatted to fit this chapter.

3.1. Background

In the last chapter we showed that retroviral nucleocapsid proteins (NC) from different retroviruses have wide range of nucleic acid (NA) chaperone activities, i.e. differ greatly in their ability to facilitate the nucleic acid remodeling reactions. Specifically, we have used the annealing reaction of complementary DNA and RNA hairpins derived from the R region of the HIV-1 genome, TAR or mini-TAR hairpins (see fig 2.4 and 2.5) to characterize the facilitating effect on annealing of four different retroviral NC proteins: HIV-1 NC, RSV NC, MoMLV NC and HTLV-1 NC. It was shown that the annealing activity of these NC proteins varies by five orders of magnitude with HIV-1 NC being the best and HTLV-1 NC being the worst annealing agent. We have also investigated the physical origin of these differences in NC chaperone functioning. Thus, we have characterized the ability of these proteins to aggregate NA, which was
previously proven to be one of the major factors contributing to the protein’s NA chaperone activity (137-141). All studied NC proteins except for HTLV-1 NC were able to fully aggregate NA at comparable concentrations, while even much higher HTLV-1 NC concentrations were unable to do so. Thus, the aggregation data were unable to explain the relatively poor annealing activity of MoMLV NC compared to HIV or RSV NC. The NC protein’s ability to destabilize all NA duplexes was shown to be another important property that contributes to its chaperone function. To characterize the NA duplex destabilizing abilities of these different retroviral NC proteins we have used three complementary experimental approaches. In particular, we have measured the equilibrium dissociation constants ($K_{d}$s) of these proteins to single- and double-stranded DNA (ssDNA or dsDNA), and have shown that all of these proteins bind stronger to ssDNA, thereby slightly destabilizing the NA duplex (138). The duplex destabilizing ability of all NC proteins studies appeared to be comparable (138), and therefore could not explain major defect in the chaperone function of the later. Finally, the single molecule DNA stretching experiment in the presence of these NC proteins have shown that the major difference in the physical properties of these protein/DNA interaction that correlates best with their chaperone function is their ability to rapidly dissociate from single-stranded nucleic acids (ssNA). Thus the MoMLV NC was much slower in dissociating than HIV or RSV NC, while HTLV-1 NC behaved similarly to the prototypical single stranded binding (SSB) protein gp32 (142-145) that is known to polymerize on ssDNA and not to unbind from it on the time scale of minutes. We concluded that the rapid kinetics of protein/NA association and dissociation is a major
component of its chaperone function, with the less critical, but also very important role of the protein’s NA aggregating ability, and still less important, but also contributing role of duplex destabilizing activities. Recently single molecule FRET studies (146) have confirmed our conclusion of the very poor chaperone activity of the HTLV-1 NC protein. However, a rather strong ability of the HTLV-1 NC to destabilize the DNA duplex was observed (146).

**Figure 3.1:** secondary structure of a) HTLV-1 NC and mutants, b) HTLV-2 NC, and c) HIV-1 NC. Aromatic residues are indicated by arrows. Basic residues are in blue and acidic residues in red. d) primary structure of HTLV-1 NC, HTLV-2 NC, BLV NC and HIV-1 NC are lined up with matching amino acids sequences.
Comparing the amino acid sequence of HTLV-1 NC (shown in Fig. 3.1a) with the sequences of the three other NC proteins studied (138), we have hypothesized that it is the presence in the former of its highly anionic C-terminal domain (CTD), absent in all other NC proteins, that probably leads to the very poor chaperone function of this protein. Here we investigate this hypothesis by comparing the chaperone properties of the WT and ΔC29 HTLV-1 NC mutant lacking 29 amino acids from its C-terminal domain, including 8 anionic residues (shown in Fig. 3.1a), as well as three other mutants (shown in Fig. 3.1a) lacking different numbers of anionic and neutral residues. The negative charge of the anionic CTD of these mutants is diminished to a different extent, such that they can be expected to show variable chaperone activity. Our nomenclature of these HTLV-1 NC mutants is similar to the one used in the recent study (147) where these proteins were shown in vivo to have a significant effect on the APOBEC3G packaging in HTLV-1 virions.

In addition to studies of the TAR RNA/DNA annealing kinetics in the presence of HTLV-1 NC and its mutants, we also use a number of other biophysical approaches to characterize these proteins interaction with nucleic acids, developed previously for HIV-1 NC (55,62,139-141) and other NC proteins (138). In particular, we investigate the NA aggregating ability of HTLV-1 NC and its CTD mutants, as well as their ability to destabilize NA duplexes. We also use single molecule DNA stretching in the presence of WT and ΔC29 HTLV-1 NC proteins to characterize their kinetics of interaction with DNA(55,62).
3.2. Methods and materials

NC proteins and mini-TAR RNA were prepared as described in chapter 2. Annealing assays is the same as that used in chapter 2. Fluorescence anisotropy experiments and single-molecule stretching experiments have also been described in chapter 2.

3.2.1. Time-resolved FRET

64-nt c-TAR DNA labeled with Alexafluor 488 at the 5’-end and 4-((4’-dimethylaminophenylazo) benzoic acid (DABCYL) at the 3’-end was obtained from TriLink Biotechnologies (San Diego CA). c-TAR DNA was refolded in 25 mM HEPES pH 7.5 and 20 mM NaCl by heating at 80°C for 2 min, then cooling to 60°C for 2 min, followed by addition of MgCl₂ (to 10 mM final) and placement on ice. For the time-resolved FRET experiments, folded TAR DNA at a final concentration of 100 nM was incubated with and without NC in a buffer containing 20 mM HEPES pH 7.5, 50 mM NaCl, 10 µM TCEP, 5 mM β-mercaptoethanol and 1 µM zinc acetate for 30 min. Time-resolved fluorescence experiments were performed using time-correlated single photon counting technique at room temperature on Life-spec(Red) time-resolved spectrometer (Edinburgh Instruments, Livingston UK) equipped with EPL-475 picosecond diode laser (wavelength = 475 nm, repetition rate of ). Samples were excited using 475 nm vertically polarized pulses from the laser and the emission at 520 nm was detected at 900 with respect to the excitation axis with miniature-PMT detector (Hamamatsu H7422) and a Glan Thompson polarizer set at the magic angle. The emission decay was fit using non
iterative nonlinear least squares fitting procedure by reconvolution fit analysis module of T900 software (Edinburgh Instruments), in which decay data was fit to a sum of exponential decays convoluted with the instrumental response function.

3.3. Results

3.3.1. Effect of wt HTLV-1 NC and mutant ΔC29 on Mini-TAR DNA/RNA annealing

As was shown before (139,140) in the absence of NC proteins the annealing is extremely slow, i.e. in low salt (20 mM Na) it takes hours, and the corresponding bimolecular rate constant of mini-TAR RNA/DNA annealing is \( \sim 100 \text{ M}^{-1}\text{s}^{-1} \). Addition of HIV-1 NC increases this rate by \( \sim 10^3 \)-fold (139,140). Presented in Figure 3.2 is the annealing time course of this reaction in solution of 20 mM NaCl in the presence of HTLV-1 NC and its four mutants introduced in Figure 3.1: WT

![Figure 3.2](image-url)

**Figure 3.2:** Annealing time courses for mini-TAR RNA/DNA hairpins annealing without NC protein or in presence of wt HTLV-1 NC, EED, D69A, LLD and ΔC29 at 37 °C. Concentrations of NCs used were 1.5 μM HIV-1, 1 μM RSV, 3 μM MLV, and 3 μM HTLV-1. Curves represent single-exponential fits to equation 2.4.
HTLV-1, ∆C29, EED, D69A and LLD NC. We see that the ability of these proteins to facilitate annealing reactions varies greatly among these proteins. The overall neutral WT HTLV-1 NC has practically no facilitating effect even at 5 µM concentration, where its binding to NA is well-saturated according to our titration (Fig. 3.3a). At the same time, its CTD deletion mutant ∆C29 that lacks 8 out of its total 11 negative charges facilitates annealing almost as well as HIV-1 NC (see Fig. 2.6a), the so far best-known NA chaperone (138). The next best chaperone HTLV-1 NC EED facilitates annealing just a little less than the ∆C29 mutant, suggesting that the three subsequent charges that were deleted in this mutant provide the major contribution to the inhibition of HTLV-1 NC chaperone activity. According to the annealing time course presented in Figure 3.2, the chaperone activity of the next two mutants D69A and LLD is almost indistinguishable within our experimental error. This result is consistent with both of these mutants lacking just a single negatively charged amino acid D at position 69 (Figure 3.1).

3.3.2. Chelation of zinc from HTLV-1 NC zinc fingers does not improve its chaperone function

We have performed the mini-TAR RNA/DNA annealing experiments in the presence of saturating amounts of the WT HTLV-1 NC pre-incubated for 30 min with 1 mM EDTA. The latter is expected to fully chelate the zinc from its zinc fingers, thereby destroying this protein’s ability to position its aromatic residues, and most likely, to destabilize NA duplexes. This effect of EDTA chelation of zinc on HIV-1 NC was recently characterized ((55) and unpublished data by Musier-Forsyth et al.). We have observed
(data not shown) that the annealing time course is practically unaffected by EDTA addition, implying that the zinc finger structures of HTLV-1 NC are not responsible for the poor chaperone function of this protein. This is an important conclusion, suggesting that the three aromatic residues of HTLV-1 NC zinc fingers are not participating in the interactions that make this protein a poor NA chaperone. This result is consistent with the major facilitating effect of the HTLV-1 NC CTD deletion on its chaperone function.

3.3.3. High salt makes WT and ΔC29 HTLV NC similarly good chaperones

Based on the results presented in Figure 3.2, it is the CTD of HTLV-1 NC, and more specifically, its anionic residues, that are responsible for the poor chaperone function of the WT protein. We further hypothesize that the electrostatic interaction between this protein’s anionic CTD and cationic NTD is responsible for its poor chaperone function. If true, the higher salt is expected to screen this electrostatic interaction, thereby rendering WT HTLV-1 NC and ΔC29 HTLV-1 NC more similar to each other. We test this hypothesis by comparing the mini-TAR RNA/DNA annealing at saturating amounts of these proteins in low (20 mM NaCl) and high (100 mM NaCl) salt, as presented in Figure 3.3b. In order to find the saturating conditions for the WT and ΔC29 HTLV-1 NC proteins for these two salt conditions we performed titration experiments presented in Figure 3.3a. Here the percent of mini-TAR RNA annealed to mini-TAR DNA at 30 minutes after mixing was monitored as a function of protein concentration at 20 and 100 mM NaCl. We see that in the absence of protein (i.e. at [HTLV-1 NC]=0) the annealing after 30 min is significantly higher at 100 mM compared to 20 mM NaCl. This result is in agreement with the facilitating effect of mono- and
divalent cations on annealing (139). The annealing in the presence of ΔC29 is always higher than with the WT, but the behavior of these two proteins is much more similar at higher salt. In 100 mM NaCl the effect of either protein on annealing saturates at ~5 µM. Interestingly, for both proteins in 20 mM NaCl the annealing first increases with [NC] up to [NC]~1-2 µM, but starts to decline upon further protein increase. This phenomenon can be explained by the significant NA duplex destabilization by the saturating amounts of both WT and ΔC29 HTLV-1 NC proteins, leading to the destabilization of the final annealed RNA/DNA duplex.

Figure 3.3: A) titration of mini-TAR RNA for wt HTLV-1 NC and mutant ΔC29 at 20mM NaCl and 100mM NaCl. B) annealing time courses of wt HTLV-1 NC and mutant ΔC29 at 20mM NaCl and 100mM NaCl.
effect shifts the mini-TAR RNA/DNA annealing equilibrium towards reactants, thereby leading to less product formation at equilibrium, as was previously shown for the HIV-1 NC (140) and other retroviral NC proteins (138). This effect is rather strong for HTLV-1 NC protein and its CTD mutants, in accord with their relatively strong duplex destabilizing ability, as estimated based on these protein’s $K_d$s to ss and ds DNA (see Table 3.1). It is also consistent with the results of the recent single molecule FRET study (146) that showed HTLV-1 NC to be a good duplex destabilizer, but a poor NA chaperone. This destabilization of the annealed mini-TAR RNA/DNA duplex by the NC proteins is only apparent in lower salt, when this duplex in the absence of protein is much less stable than it is in higher salt. The fact that in 20 mM NaCl the annealing saturates at $[\text{HTLV-1 NC}] \geq 4 \, \mu\text{M}$ is consistent with the high $K_d$~4 $\mu\text{M}$ of these proteins to dsDNA, measured at 20 mM NaCl (see Table 3.1). These results also suggest that the $K_d$s for the WT and ΔC29 HTLV-1 NC are similar to each other, also in accord with the data in Table 3.1. In addition, the fact that the curves in Figure 3.3a saturate at comparable [NC] in both low and high salt hints at the fact that binding of these proteins to NA is rather salt-independent. This conclusion is consistent with the weak salt dependence of HTLV-1 NC binding to RNA measured by Morcock et.al. (81). Based on these steady state results we further perform the mini-TAR RNA/DNA annealing kinetic studies in low and high salt in the presence of 3 $\mu\text{M}$ and 6 $\mu\text{M}$ of either protein, respectively. While both proteins are at about saturation at these concentrations, no strong decline in the amount of mini-TAR RNA annealed due to duplex destabilization in low salt at 3 $\mu\text{M}$ of either protein is observed yet (Figure 3.3a).
3.3.4. Salt effect on the kinetics of mini-TAR RNA/DNA annealing in the presence of HTLV-1 NC and ΔC29 HTLV-1 NC proteins

Presented in Figure 3.3b are the mini-TAR annealing time courses at low and high salt in the presence of either WT or ΔC29 HTLV-1 NC. Also, presented for the comparison are the annealing time courses in 20 and 100 mM NaCl in the absence of protein. We see that in the absence of protein the annealing is very slow in both: low or high salt, even though it is slightly facilitated by higher salt. Interestingly, the annealing in the presence of ΔC29 HTLV-1 NC is equally efficient either at low or high salt. At the same time, annealing in the presence of the WT HTLV-1 NC is very poor in low salt (10 mM NaCl), but is greatly facilitated in higher salt (100 mM NaCl), which brings it almost to the level of annealing with ΔC29. It is clearly not just the salt-induced facilitation of annealing (compare 20 and 100 mM salt in no protein case), but rather the direct effect of salt on the WT HTLV-1 NC protein, which makes it almost as good of a chaperone as ΔC29. We expect still higher salt to further screen the electrostatic attraction between the anionic CTD and cationic NTD of the WT HTLV-1 NC, thereby making the WT and ΔC29 HTLV-1 NC proteins even more similar to each other. This result strongly supports our hypothesis that it is the electrostatic attraction between the anionic CTD and cationic NTD of HTLV-1 NC that makes it a poor chaperone. The nature of this phenomenon will become clearer upon its further investigation with complementary approaches in the rest of this study. The schematic representation of the physical model (Fig. 3.8) of HTVL-1 NC interaction with NA is consistent with all of the results of this and other(81,146) relevant studies.
All of the assays presented in Figure 3.3 were also performed with the closely related HTLV-2 NC protein and its CTD truncation mutant shown Figure 3.1b. Interestingly, the results for the HTLV-1 and -2 are indistinguishable within the experimental error. This result again suggests that it is the overall cationic/anionic structure of the two HTLV NC domains, rather than the exact amino acid sequence of these proteins (see Figure 3.1a and b), that determines their chaperone activity.

3.3.5. NA aggregating ability of HTLV-1 NC and its mutants correlate with the charge of its C terminal domain

Presented in Fig. 3.4 is the percent mini-TAR RNA aggregated from solution under conditions similar to the ones used in annealing experiments as a function of concentration of HTLV-1 NC or its mutant. Comparing data in Fig. 3.2 and 3.4, we see that the NA aggregating ability of these proteins correlate with their NA chaperone activity. Indeed, while the RNA aggregation of HTLV-1 ΔC29 NC is almost as efficient as that of the HIV-1 NC, the WT HTLV-1 NC and its two mutants with just a single negative charge removed from its CTD: LLD and D69A, cannot aggregate RNA at any concentration. At the same time

![Figure 3.4](image)
the HTLV-1 NC EED protein with the three subsequent negative charges removed has intermediate NA aggregating ability.

In chapter 2 we observed that HIV-1 NC, RSV NC and MoMLV NC all aggregate RNA very well under the similar solution conditions, where saturating concentration of each is required for NA aggregation. This requirement was further investigated in our other work, where the HIV-1 NC competition with Mg$^{2+}$ and Na$^+$ for NA binding was studied (139). It was shown that the HIV-1 NC competes with Mg$^{2+}$ and Na$^+$ as an effectively trivalent cation. So depending on Mg$^{2+}$ and Na$^+$ concentrations, different amounts of HIV-1 NC protein are required to saturate NA, and therefore, to produce NA aggregation. We further speculated that HIV-1 NC protein aggregates NA via the same electrostatic mechanism as the other multivalent cations, such as polyamines, polylysines, Cobalt Hexamine etc. (96,97,113,148,149). This electrostatic aggregation requires that the cationic protein binds NA in the delocalized i.e. polyelectrolyte mode, characterized by the high mobility of the protein in its NA-bound state. This high mobility of the cationic protein in its NA-bound state is expected to correlate with its fast on/off rate, which in turns seems to be required for the high chaperone activity of the protein (55,62). Indeed, we have shown(138) that in contrast to HIV-1 NC and RSV NC, the MoMLV NC, which dissociates slowly from ssDNA, does not chaperone mini-TAR DNA/RNA annealing well.

The wt HTLV-1 NC is overall neutral, and therefore cannot be expected to aggregate NA as multivalent cations do. Later we will show HTLV-1 NC is slow to dissociate ssDNA, such that even if it did aggregate NA, it would most likely not be an
efficient chaperone. Deletion of its CTD makes this protein both:

<table>
<thead>
<tr>
<th>NC protein</th>
<th>$K_{ss}^{ns}$, nM</th>
<th>$K_{ds}^{ds}$, nM</th>
<th>$\delta G$, cal/mol/bp</th>
</tr>
</thead>
<tbody>
<tr>
<td>HTLV-1 NC ΔC29</td>
<td>276±16</td>
<td>4366±777</td>
<td>- (0.4 ± 0.1)</td>
</tr>
<tr>
<td>HTLV-1 NC EED</td>
<td>380±15</td>
<td>4213±1353</td>
<td>- (0.3 ± 0.1)</td>
</tr>
<tr>
<td>HTLV-1 NC LLD</td>
<td>252±19</td>
<td>6059±1453</td>
<td>- (0.5 ± 0.05)</td>
</tr>
<tr>
<td>HTLV-1 NC D69A</td>
<td>291±30</td>
<td>4596±1863</td>
<td>- (0.4 ± 0.05)</td>
</tr>
<tr>
<td>HTLV-1 NC</td>
<td>431±87</td>
<td>1840±400</td>
<td>- (0.2 ± 0.05)</td>
</tr>
<tr>
<td>HIV-1 NC</td>
<td>123±9</td>
<td>233±11</td>
<td>- (0.1 ± 0.02)</td>
</tr>
<tr>
<td>RSV NC</td>
<td>200±49</td>
<td>660±47</td>
<td>- (0.2± 0.05)</td>
</tr>
<tr>
<td>MoMLV NC</td>
<td>334±34</td>
<td>1092±28</td>
<td>- (0.2 ± 0.05)</td>
</tr>
<tr>
<td>SSB gp2.5</td>
<td>3162</td>
<td>$1\times10^7$</td>
<td>- (1.4± 0.1)</td>
</tr>
<tr>
<td>SSB gp2.5D26C</td>
<td>20</td>
<td>$4\times10^4$</td>
<td>- (1.3± 0.1)</td>
</tr>
<tr>
<td>SSB gp32</td>
<td>32</td>
<td>$1\times10^5$</td>
<td>- (1.4± 0.1)</td>
</tr>
<tr>
<td>SSB gp32DC</td>
<td>0.2</td>
<td>$2.5\times10^3$</td>
<td>- (1.6± 0.1)</td>
</tr>
</tbody>
</table>

Table 3.1: Dissociation constants of NC proteins from single- and double- stranded DNA, and change of free energy (per base pair) caused by the NC protein destabilization.

Overall cationic and fast dissociating, consistent with its efficient NA aggregating and
chaperone activities. Partial CTD charge removal in the EED mutant leads to the partial restoration of these activities.

3.3.6. The nucleic acid binding and duplex destabilizing activities of HTLV-1 NC are only weakly affected by its CTD mutations

Collected in Table 3.1 are the dissociation constants, $K_d$, for HTLV-1 NC and its four mutants binding to 20-nt ssDNA and 20-bp dsDNA along with $K_d$s measured previously with the same approach for the other retroviral NC proteins from chapter 2. Most notably, the HTLV-1 NC binds ssDNA a few-fold weaker and dsDNA ~10-fold weaker than the other retroviral NC proteins. The measured $K_d$ values are in good agreement with the results of fluorescent quenching binding study (81), which obtained the following values for the WT HTLV-1 NC: $K_{d,ss}^{ss} = 400\pm200$ nM and $K_{d,ds}^{ds} = 3000\pm2000$ nM. Even though the later studies were conducted in much lower salt (1 mM of Na phosphate buffer, instead of 50 mM of NaCl in our case) these $K_d$ values are very close to ours (see Table 3.1). This result is in agreement with the weak salt dependence of HTLV-1 NC binding to NA observed previously (81). Interestingly, HTLV-1 NC binds DNA more weakly than other NC proteins, although it contains more aromatic residues in its zinc fingers than most other retroviral NC proteins (see Fig. 3.1a and c). In the case of HIV-1 NC it was shown that the stacking interactions of its aromatic residues with the unpaired NA bases are responsible for the preferential binding of this protein to ssNA (45,55,56,59,150). In Morcork et al. (81) the double zinc finger mutant (aa11-51) of HTLV-1 NC protein bound ssDNA ~5-fold weaker than the WT. This is in contrast to our
\( \Delta C29 \) HTLV-1 NC (lacking residues 57-85), which binds both ss and ds DNA \( \sim 1.2 \)-fold better than the WT (see Table 2.1). The reason for the significantly weaker binding of the HTLV-1 NC (11-51) is probably the lack of the two N-terminal cationic residues. Thus, the weak binding of HTLV-1 NC protein to NA cannot be attributed to the presence of the anionic CTD. It is possible that the excessive proportion of the “rigid” proline residues in this protein leads to its weak binding.

3.3.7. HTLV-1 NC and its CTD mutants are good NA duplex destabilizers

The fact that the HTLV-1 NC and all of its CTD mutants bind better to ss- then to ds- DNA implies that these proteins destabilize the DNA duplexes. The values of \( K_{ds}^{ss} \) and \( K_{ds}^{ds} \), and the base pair number that binds a single protein, \( n=4 \) (or 8 nucleotides) (81) can be used to estimate the destabilization free energy per base pair, \( \delta G \), according to (107,138,140)

\[
\delta G = \frac{RT}{n} \ln \left( \frac{K_{ds}^{ss}}{K_{ds}^{ds}} \right) \quad (2.10)
\]

\( \delta G \) values obtained in this way are collected in Table 3.1. We see that the HTLV-1 NC is at least as good or better destabilizer as the rest of the retroviral proteins. As for all other NC proteins, HTLV-1 NC and its mutants partially destabilize DNA duplexes without completely melting them. While this activity definitely contributes to HTLV-1 NC’s chaperone function, is unable to make it a good chaperone by itself, as is clear from the very poor chaperone properties of the WT HTLV-1 NC. The CTD mutants of HTLV-1 NC destabilize the DNA duplex even slightly better than the WT. This stronger
duplex destabilization by HTLV-1 NC and its mutants compared to other retroviral NC proteins comes largely from the HTLV-1 NC is weaker binding to dsDNA, rather than from its stronger binding to ssDNA. The observed stronger DNA duplex destabilization by HTLV-1 NC is also in full agreement with a recent study (146), which also draws the conclusion that HTLV-1 NC is a better duplex destabilizer than HIV-1 NC as previously studied with the same single molecule FRET approach (151,152). At the same time, the authors (146) also observed that WT HTLV-1 NC was unable to facilitate the NA annealing reaction.

3.3.8. Single molecule DNA stretching studies suggest slow kinetics of the WT HTLV-1 NC proteins and much faster dissociation of its CTD deletion mutant.

**DNA force-extension experiments**

Figure 3.5a shows a stretching and relaxation cycle for bacteriophage λ DNA in the absence of protein and in the presence of a saturating concentration of wild type HTLV-1 NC using optical tweezers. In the absence of protein, at extensions much less than the B-form contour length of dsDNA, very little force is required to stretch the DNA. As the contour length is approached, the force increases dramatically, reflecting the elasticity of the double helix. The dramatic transition observed as DNA is stretched from 0.34 nm/bp to 0.6 nm/bp at a constant force of about 60 pN represents a transition from dsDNA to ssDNA, or a force-induced melting transition (25,31,32,34). In the absence of protein, the transition is cooperative, and as the DNA is relaxed back to
Fig 3.5: a) DNA stretching (solid line) and relaxation (dotted line) in the absence (black) and presence (red) of HTLV-1 NC (700 nM). b) DNA stretching (solid line) and relaxation (dotted line) in the absence (black) and presence (red) of HTLV-1 ΔC29 NC (200 nM).

lower extensions, the force-extension curve is almost completely reversible, showing little hysteresis. In contrast, in the presence of wild type HTLV-1 NC, the DNA relaxation happens at significantly lower forces than extension. In other words, upon relaxation the two strands do not re-anneal, but rather relax as ssDNA bound by HTLV-1 NC, indicating that the proteins do not unbind ssDNA on the time scale of the relaxation cycle, which is a few minutes. The stretching part of the DNA cycle is non-equilibrium, and therefore, depends on the rate of DNA pulling, as was observed before for SSB proteins (142-144,153).

To test the hypothesis that the large hysteresis, i.e. slow HTLV-1 NC unbinding from ssDNA, was due to its anionic CTD, we also performed DNA stretching in the presence of its CTD deletion mutant, HTLV-1 NC ΔC29. Figure 3.5b shows a representative DNA stretching cycle in the presence of saturated HTLV-1 NC ΔC29. At low force there were always jumps observed on the stretching curve, in which the force
increases at extensions below the B-form contour length of 0.34 nm, but then drops suddenly. These jumps represent dsDNA attraction to itself, or self-aggregation of the single DNA molecule induced by the protein. This result is consistent with the observation that in contrast to the wt HTLV-1 NC, its CTD deletion mutant does aggregate NA (shown in Fig. 3.4). This type of aggregation has been observed in the presence of several nucleic acid chaperone proteins, including HIV-1 NC (24,25,62), the retrotransposon chaperone LINE-1 ORF1p (1,154), and other nucleocapsid proteins (138). The single stranded DNA stretching curve above the strand separation transition shifted to shorter contour lengths, which indicated that protein-bound ssDNA is shorter than ssDNA in the absence of protein. The hysteresis observed upon relaxation in the presence of HTLV-1 NCΔC29 is much smaller than that observed in the presence of wild type HTLV-1 NC. Also, the DNA relaxation curve is similar in shape to the stretching curve, showing that protein was able to unbind ssDNA, and the DNA was able to almost completely re-anneal on the time scale of the relaxation cycle.

**DNA re-annealing time constant measurement**

As discussed above, the different amounts of hysteresis observed upon DNA relaxation after force-induced melting reflects the difference in the DNA re-annealing time in the presence of the HTLV-1 NC and its CTD deletion mutant. This difference in DNA re-annealing time, in turns, reflects the difference in the protein dissociation time from ssDNA.

To measure the DNA re-annealing time in the presence of HTLV-1 NC and its CTD deletion mutant we stretched the DNA completely through the melting plateau until
reaching the single stranded stretching region, about 0.6 nm/bp. During the process of stretching, DNA base pairs are opened and ssDNA is exposed for protein binding. Then we relaxed the DNA until halfway through the transition, at an extension of 0.46 nm/bp. At this point, we held the DNA extension constant and measured the force as a function of time. As shown in Figure 3.6, the force increased exponentially with time until saturation, reflecting the protein’s dissociation from ssDNA and the subsequent re-annealing of the two protein-free complementary DNA strands. The single-exponential fit of the force relaxation yields the DNA re-annealing time constant $T_\uparrow = 44.8 \pm 6.6$ s and 5.3 ± 2 s for the wild type HTLV-1 NC and the HTLV-1 NCΔC29 mutant, respectively. As will be discussed elsewhere, the relaxation time $T_\uparrow$ depends on the stretched DNA length, and is not equivalent, but related to the protein’s off time $\tau_{\text{off}}$ as follows:

$$T_\uparrow \propto \frac{\tau_{\text{off}}}{K_d}.$$  

The dissociation constants $K_d$ for these two proteins are less than 50%
different (see Table 2.1). Therefore, $\tau_{off}$ for these two proteins are expected to differ by the factor of $\sim 10$, i.e. about as much as their $T_1$ do. These results are consistent with the hypothesis that the electrostatic interaction of the anionic CTD of HTLV-1 NC with the cationic NTD of the neighboring protein bound to ssNA is responsible for the slow dissociation of the WT HTLV-1 NC, as presented in Fig. 3.7. Further discussion of this model, which involves CTD regulation of the HTLV-1 NC/NA on/off kinetics is presented below.

3.4. Discussion

In this work we have investigated the physical nature of the extremely poor chaperone activity of HTLV-1 NC protein. We have found that the deletion of its anionic C terminal domain (see Figure 3.2) or the increasing salt (see Figure 3.3) improve the chaperone activity of HTLV-1 NC significantly. At the same time, elimination of HTLV-1 NC proteins’s zinc fingers by chelation of zinc with EDTA does not improve this protein’s chaperone function. Moreover, the poor chaperone activity of the WT HTLV-1 NC in low salt (20 mM NaCl) is accompanied by the inability of this protein to aggregate NA (Fig. 3.4) or to unbind from the ssDNA on the time scale of seconds (Figures 3.5 and 3.6). This result is consistent with both of these activities being critical for the protein functioning as a nucleic acid chaperone (55,62,139-141). Interestingly, both of these defects are relieved by deletion of the HTLV-1 NC CTD (see Figure 3.1a). In addition, mutations of the CTD of HTLV-1 NC that either neutralize or delete variable number of its anionic charges lead to the protein properties intermediate between the WT HTLV-1
NC and HTLV-1 ΔC29 (see Figures 3.2, 3.3, 3.4). More specifically, mutation of the three continuous anionic EED residues (62-64) into the neutral AAA residues have the largest effect, leading to the protein properties approaching these of HTLV-1 ΔC29. At the same time, the two other tested HTLV-1 NC mutants D69A and LLD with just a single anionic residue D69 eliminated have properties similar to each other, and close to the WT HTLV-1NC protein.

In contrasts, mutations and deletions within this protein’s CTD were found to have very limited effect on this protein’s NA binding strength to either ss- or ds- DNA (see Table 3.1). Overall, binding of HTLV-1 NC and its CTD mutants to NA is weaker than of the NC proteins from other retroviruses (see Table 3.1). This weaker binding is especially pronounced for binding to the ds NA, leading to the stronger duplex destabilization by HTLV-1 NC and its CTD mutants compared to other NC proteins (see Table 3.1).

Our conclusions are in full accord and complemented by the results of the HTLV-1 NC/NA Trp quenching binding studies by Morcock et al. (81). Specifically, these authors found $K_d$ values for WT HTLV-1 NC to either ss or ds DNA that are very similar to our values obtained by fluorescence polarization (see Table 2.1). In addition, they determined the moderate cooperativity of the WT HTLV-1 NC binding to polymeric ssDNA with the cooperativity parameter ($\omega$) decreasing with salt from $\omega \sim 10$ at 1 mM NaCl to $\omega \sim 1$ at 100 mM Na. This result is consistent with our preliminary data from the gel-based (EMSA) assay, yielding $\omega \sim 5$ at 50 mM Na (unpublished data by D. Qualley and K. Musier Forsyth). In addition, Morcock et al. observed that the HTLV-1 NC (11-51)
(lacking CTD and a short fragment at the N-terminal end) binds NA with negligible cooperativity, i.e. $\omega \sim 1$. These authors have also studied the Trp fluorescence of the unbound HTLV-1 NC protein and found interactions between its zinc fingers and its own CTD, that is eliminated by higher salt, and absent in the HTLV-1 NC (11-51) peptide.

Taken together the current results as well as Morcock’s data suggest that the electrostatic interaction between the anionic CTD and cationic NTD (including its zinc fingers) of HTLV-1 NC controls this protein’s binding to nucleic acids, as presented schematically in Figure 3.7. Indeed, such electrostatic interactions between the NA-bound protein neighbors lead to the apparent salt-dependent binding cooperativity (81), and result in the HTLV-1 NC polymerization on NA, thereby preventing its aggregation (Fig. 3.4) and fast protein unbinding (Figures 3.5 and 3.6), and also, leading to its extremely poor chaperone function (Figures 3.2 and 3.3). The fact that the CTD deletion makes dissociation of the HTLV-1 NC protein faster (Figure 7 and 8 and Table 2.2), but does not significantly affect its $K_d$ (see Table 2.1), just as the salt has no effect on its $K_d$ (81), implies that the NA association of the HTLV-1 NC also becomes faster upon reduction (or screening) of the negative charge of its CTD. This, in turns, implies that the electrostatic interaction between the anionic CTD and cationic NTD within the single unbound HTLV-1 NC protein (81) slows down this protein’s NA association rate, probably by interfering with its NTD binding to NA. As CTD is removed (or CTD/NTD electrostatic attraction become screened by higher salt), both protein association with and dissociation from NA become similarly facilitated, such that the protein’s $K_d$ involving the ratio of the two rates remain unaffected.
Comparing HTLV-1 NC with NC proteins from other retroviruses and with the single stranded binding (SSB) proteins

It is interesting to compare HTLV-1 NC protein with other NC and single stranded binding (SSB) proteins. These two protein classes have both common and divergent features. Thus, both NC and SSB proteins bind predominantly sequence nonspecifically to all nucleic acids (ss and ds, RNA and DNA) with the significant preference for ssNA, thereby destabilizing all NA duplexes. However, NC proteins destabilize NA duplexes only very slightly, while SSB proteins are able to melt dsNA completely, i.e. to stabilize ssNA under physiological conditions. These differences reflect the divergent biological roles of these two groups of proteins. Indeed, one of the major functions of the retroviral NC proteins is to assist reverse transcription by facilitating nucleic acid remodeling (45). This chaperone activity of NC proteins requires slight destabilization of NA duplexes, to help NA avoid misfolded conformations without destabilizing its final most stable state. This slight duplex destabilization by NC proteins is reflected in the small free energy of duplex destabilization per bp of 0.1-0.5 kcal/mol/bp (see column 4 in Table 3.1) compared to the average free energy of dsDNA melting per bp at physiological conditions ~1.36 kcal/mol/bp (113). In contrast, the primary function of the bacteriophage SSB proteins is to support phage replication via stabilization of the helicase-unwound ssDNA(155). This function is consistent with the duplex destabilizing free energy per bp of 1.4 - 1.6 kcal/mol/bp for gp32 from T4 (142-144) and gp2.5 from T7 (153) bacteriophages (see Table 3.1) comparable to or larger than the average DNA bp melting free energy.
While thermodynamically HTLV-1 NC protein belongs to mild NA duplex destabilizers (Table 3.1), from the kinetic point of view it behaves more like much slower SSB protein. Indeed, according to our preliminary estimates of the protein off rates based on our single molecule DNA stretching data, WT HTLV NC has an off rate from ssDNA $\sim 10^2$ s$^{-1}$, that is similar to the slow off rate of the gp32 SSB protein. The HTLV-1 NC ΔC29 protein has about 10-fold higher off rate, that is similar to the off rate of the faster gp2.5 SSB protein. At the same time, the rest of the NC proteins from other retroviruses have still much faster off rates that are too high to be measured by this approach (i.e. off rate $\gg 10^3$ s$^{-1}$).

Our proposed model of NA binding regulation by the CTD of HTLV-1 NC presented in Figure 3.7 is reminiscent of the CTD-regulated binding of gp32 (142-145) and gp2.5 (23,153). For all three proteins their association with NA requires prior unbinding of the anionic CTD from their cationic NA binding site. However, in contrast to HTLV-1 NC, the CTD of gp32 does not participate in the cooperative interaction between
the NA-bound proteins, such that this interaction is salt-independent, and is not eliminated by its CTD truncation. Also, there is no evidence of any attraction between the NA-bound gp2.5 proteins, such that the off rate of this protein is also not affected by its CTD deletion. As a result, only the on rates and the equilibrium binding constants, but not the off rates of these SSB proteins are increased by their CTD truncation. In contrast, for HTLV-1 NC both its on and off rates are increased by its CTD deletion (or high salt), while its $K_d$ remains largely unaffected. These specific protein features have probably evolved for some specific biological function that is yet to be understood.
4. Deaminase-independent inhibition of HIV-1 reverse transcription by APOBEC3G

This work was a collaboration with Judith Levin’s lab and originally published in Nucleic Acids Research, 2007, Vol. 35, No. 21 7096-7108. It’s reformatted to fit in this thesis. Single molecule experiments were done by Fei Wang and bulk experiments by other co-authors.

4.1. Background

Human APOBEC3G (A3G) is a host cytidine deaminase, which was first identified by Sheehy et al (156) as the cellular factor that blocks HIV-1 replication in the absence of the viral infectivity factor (Vif) protein. Cellular expression of A3G results in its incorporation into vif-deficient HIV-1 particles, whereas its presence in WT virions is dramatically reduced by Vif-induced degradation via the ubiquitination-proteosome pathway (157).

Initial studies suggested that the deamination activity of A3G contributes to its antiviral activity and is associated with G to A hypermutation (158-161). However, more recent results indicate that a deaminase-independent mechanism might also be involved in A3G’s antiviral activity: (i) Deamination activity is not absolutely correlated with antiviral activity against HIV-1 (162-170); (ii) Several reports (though not all) indicate that hepatitis B virus may be inhibited by A3G without significant detection of G to A hypermutation (171-173) and (iii) Other APOBEC proteins block replication of
mouse mammary tumor virus (174) and several retrotransposons (169,175-179) in the absence of editing activity.

Prior to the discovery of A3G, it was already known that Δvif viruses produced in ‘nonpermissive’ cells (PBMCs or certain T-cell lines, e.g. H9) were 100- to 1000-fold less infectious than WT (180-182) and were deficient in their ability to complete reverse transcription (183-186). Moreover, similar results were obtained in endogenous reverse transcription assays (185,187,188). More recent analysis of HIV-1-infected cells expressing A3G or A3F has confirmed the initial observations (160,165,168,170,189-193).

Efficient and specific reverse transcription depends on the viral nucleocapsid protein (NC), which functions as a nucleic acid chaperone (43-46,194). This means that NC can catalyze nucleic acid conformational rearrangements that lead to the most thermodynamically stable structures (47). Like NC, A3G has two zinc finger domains and binds nucleic acids (195). However, whereas A3G has a strong preference for binding single-stranded (ss) nucleic acids (6,12), NC binds ss or double-stranded (ds) RNA and DNA (62) [for further information on NC’s nucleic acid binding properties, see Ref. (93)]. The two proteins also differ in their effects on virus replication: NC acts as a positive factor, whereas A3G is an inhibitor in the absence of Vif.

We previously reported the successful preparation of highly purified, catalytically active A3G expressed in a baculovirus system and demonstrated that the availability of a pure protein (without contamination by other proteins, either host or viral) was invaluable for rigorous analysis of the biochemical properties of A3G (167). In the course of this study, we unexpectedly found that A3G does not interfere with NC binding to
ssRNA (and vice versa) (167). This suggested that inhibition of reverse transcription by A3G is likely to be unrelated to an effect on NC chaperone function.

To test this hypothesis and to probe the mechanism that might be involved, we took advantage of defined biochemical assay systems that we have developed over the years for studies on viral DNA synthesis (196-199). Thus, using our highly purified A3G as well as purified NC and RT, we investigated the effect of A3G on a series of reconstituted reactions that occur during reverse transcription. This allowed us to perform an independent analysis of individual steps in the pathway, which is not possible in cell-based systems. We found that A3G inhibited all reverse transcriptase (RT)-catalyzed DNA elongation reactions, but not RNase H activity or NC's ability to promote annealing. These observations could be explained by critical differences in the binding properties of NC, A3G and RT, as measured by single-molecule DNA stretching and fluorescence anisotropy (FA). Our findings are unique and provide strong support for a novel mechanism that could account for the observed deaminase-independent A3G-mediated antiviral activity.

4.2. Materials and methods

4.2.1. Materials

Purified tRNA\textsubscript{Lys} from human placenta was obtained from Bio S&T (Lachine, Quebec, Canada). DNA and RNA oligonucleotides were purchased from Lofstrand (Gaithersburg, MD), Integrated DNA Technologies (Coralville, IA), Oligos Etc., Inc. (Wilsonville, OR). [γ-32P]ATP (3000 Ci/mmol) and [α-32P]dCTP (6000 Ci/mmol) were
purchased from GE Healthcare (Piscataway, NJ). HIV-1 RT was obtained from Worthington Biochemical Corp. (Lakewood, NJ). Calf intestinal phosphatase, T4 polynucleotide kinase, and Vent DNA polymerase were obtained from New England Biolabs (Beverly, MA). SUPERaseIn, an RNase inhibitor, was purchased from Ambion, Inc. (Austin, TX). Recombinant wild-type HIV-1 NC (55-amino-acid form) was a generous gift from Dr Robert Gorelick and was prepared as described previously (69,200).

Recombinant enzymatically active A3G and the deaminase-deficient A3G mutant (C291S) were expressed in a baculovirus expression system and purified as previously described (167). A3G preparations were confirmed to be free from contamination with RNases (data not shown) and no RNA degradation was apparent in any of the experiments (e.g. see Figure 4.1a).

4.2.2. Methods

**Plasmid construction**

All plasmid sequences were derived from the HIV-1 pNL4-3 clone (201). Plasmid pUL (viral insert from nt 566 to 1419) was constructed from the previously described pRUG plasmid (199).

**Preparation of RNA**

The DNA templates for *in vitro* RNA transcription were derived from plasmid pRUG [for transactivation response element (TAR) RNA, nucleotide(s) (nt) 1–59] or pUL (for RNA UL244, nt 113–244). DNA fragments containing both the T7 promoter and the DNA sequence equivalent to the desired vRNA were amplified by PCR using Vent DNA
polymerase and the following primers: forward primer (5’-ccaatgcttaatcagtgaggc), located at the start of the amp gene in pUL; reverse primers, 5’-gtcctgcgtcagagatc (RNA UL244) and 5’-gggttccctagttagccaga (TAR RNA). The DNA fragments were gel-purified and transcribed using an Ambion MEGAscript kit (Ambion Inc., Austin, TX). Gel-purified RNAs were dephosphorylated by calf intestinal phosphatase and were then labeled at their 5’ ends with [γ-32P]ATP, using T4 polynucleotide kinase (202). Unincorporated nucleotides were removed by passing the reaction mixture through NucAway Spin Columns (Ambion). RNA 244 (197,199) and acceptor RNA 148 (197) were prepared as described.

**Reverse transcription assays**

Reaction components are given for 20 µl unit reactions, which were scaled up as needed. SUPERaseIn at a final concentration of 0.5 U/µl was added to all RNA-containing reaction mixtures. Incubation/preincubation was always at 37°C. Six percent native or denaturing polyacrylamide gels were used for annealing or extension assays, respectively, except as noted. Radioactivity was quantified by using a Typhoon PhosphorImager and ImageQuant software.

**tRNA\textsubscript{Lys} \textsuperscript{Lys} annealing to viral RNA (vRNA)**

Reaction mixtures contained buffer (50 mM Tris–HCl (pH 8.0), 75 mM KCl, 0.1 mM MgCl\textsubscript{2}, 1 mM DTT), 0.2 pmol of purified human placental tRNA\textsubscript{Lys} and 0.1 pmol of 5’ 32P-labeled RNA UL244 and were incubated in the absence or presence of A3G (80
nM) with or without HIV-1 NC (7 nt/NC, 0.2 µM). Prior to loading on the gel, the samples were treated with Proteinase K (0.5 mg/ml).

**(-)Strong-Stop DNA [(-) SSDNA] synthesis**

Template RNA 244 (0.1 pmol) and 0.1 pmol each of human \( tRNA_{\text{Lys}}^{39} \) or 32P-labeled D18 primer, [complementary to the 18-nt primer-binding site (PBS)] were heat-annealed as described (199). A3G or heat-denatured (hd)A3G was then added to buffer designated as ‘reaction buffer’ (50 mM Tris-HCl (pH 8.0), 75 mM KCl, 7 mM MgCl2, 1 mM DTT) and the mixture was preincubated for 5 min. Primer extension was initiated by adding HIV-1 RT (0.1 pmol) and 50 µM each of dATP, dTTP and dGTP plus 10 µCi of \([\alpha-32P]dCTP\) (\( tRNA_{\text{Lys}}^{39} \)) or 50 µM each of all four dNTPs (D18).

**Annealing and minus-strand transfer**

Minus-strand annealing was performed as described previously (203) in the presence or absence of A3G (80 nM) with or without HIV-1 NC (3.5 nt/NC, 0.4 µM). For minus-strand transfer assays, reaction mixtures contained the components present in annealing reactions as well as HIV-1 RT (0.4 pmol), all four dNTPs (each at 50 µM), and 1 mM MgCl₂.

**Initiation of plus-strand DNA synthesis**

The assay was performed using a modified version of a previous protocol (204). Briefly, the 15-nt polypurine tract (PPT) primer (0.2 pmol) was heat-annealed to a 35-nt minus-strand DNA template (0.1 pmol) and the hybrid was preincubated with or without A3G (80 nM) for 5 min in reaction buffer (see above). DNA synthesis was initiated by adding HIV-1 RT (0.4 pmol), all four dNTPs (each at 0.5 µM), and \([\alpha-32P]dCTP\) (20 µCi),
which results in an internally labeled DNA product. Samples were loaded on a 15% denaturing gel. Since RNase H cleavage removes the annealed 15-nt PPT RNA, the DNA product is 20 nt.

**Plus-strand transfer**

The assay was performed as described previously (198) except that the final concentrations of 5' 32P-labeled (+) SSDNA donor (50 nt) and minus-strand DNA acceptor (48 nt) were 5 nM, all four dNTPs were at 50 µM each, and HIV-1 RT was 20 nM. Reactions were incubated in the presence or absence of NC (3.5 nt/NC, 0.14 µM), with or without A3G (80 nM).

**Single molecule DNA stretching**

Purified bacteriophage lambda DNA (48 500 bp) was labeled on its 3' ends with biotin and single DNA molecules were captured between two streptavidin-coated, 5 µm diameter polystyrene beads using a dual beam optical tweezers instrument (29,205). The 16.5 µm contour length DNA molecule was stretched in 100 nm steps using a piezoelectric flexure translation stage (Melles Griot, Carlsbad, CA) to reveal the DNA force-extension curve, as described previously (205). After stretching a single DNA molecule in DNA stretching buffer (10 mM HEPES, pH 7.5, 50 mM Na⁺, at 20°C), and verifying that a single molecule was present, the buffer solution was exchanged for a solution containing the same buffer with a fixed protein concentration. The protein exchange procedure was then repeated for different protein concentrations in order to determine the effect of protein on DNA stretching behavior. The transition width and hysteresis were analyzed as described previously (55) for 3 or more DNA molecules.
4.3. Results

4.3.1. Effect of A3G on primer placement and (−) SSDNA synthesis

To determine the mechanism by which A3G inhibits HIV-1 DNA synthesis, we analyzed the effect of A3G on single steps in the reverse transcription pathway (Fig. 2.2) (45). Annealing and DNA synthesis were assayed in a series of reconstituted model systems. Our goal was to determine whether A3G interfered with the nucleic acid annealing to RNA UL244. Reactions were performed in the absence or presence of NC and A3G, as indicated by the headings at the top of the gel. The positions of the RNA UL244 template and the annealed RNA duplex are shown on the right. (B) The percentage of annealed product was calculated by dividing the amount of annealed RNA by the sum of annealed plus unannealed RNA, multiplied by 100. Symbols: no NC/no A3G (filled circles); +NC/no A3G (open squares); +NC/+hdA3G (open circles); and +NC/+A3G (open triangles). (C) A tRNA<sub>Lys</sub> /RNA 244 complex was extended by HIV-1 RT in the absence (lane 1) or presence of hdA3G (lanes 2–4) or A3G (lanes 5–7). The positions of (−) SSDNA and initial pause products at bases +1, +3 and +5 are shown on the right. A3G concentrations: lane 1, 0 nM; lanes 2 and 5, 20 nM; lanes 3 and 6, 40 nM; lanes 4 and 7, 80 nM.
chaperone activity of NC, the catalytic activity of RT, or both during these reactions. Note that in most of the assays, the concentration of A3G did not exceed 80 nM, since A3G precipitates at high concentrations (167).

To evaluate the effect of A3G on the first step in reverse transcription (primer placement), we investigated the time course of NC-dependent annealing of human \( tRNA_{\text{Lys}}^{\text{UL244}} \) to the PBS in a short vRNA template (RNA UL244) (Figure 4.1a and b). In the absence of NC and A3G, no annealing was detected. As shown in Figure 4.1b, the rates of annealing in the presence of NC were similar for reactions with and without A3G and the end point values at 64 min were 70% and 66%, respectively. Addition of NC and hdA3G resulted in a slightly enhanced rate of annealing, but the end point value was very close to the values with and without native A3G. These results demonstrate that A3G did not interfere with NC-mediated formation of the vRNA–tRNA complex.

Primer placement is followed by extension of \( tRNA_{\text{Lys}}^{\text{UL244}} \) and synthesis of (−) SSDNA (Figure 2.2, steps 1 and 2). To measure extension alone, the tRNA was first heat annealed to the vRNA 244 template (199); the (−) SSDNA product formed by addition of RT was internally labeled (Figure 4.1c) (199). In the absence of A3G (lane 1) or in the presence of hdA3G (lanes 2–4), equivalent amounts of the 258-nt full-length product [(−) SSDNA attached to \( tRNA_{\text{Lys}}^{\text{UL244}} \)] and pause products (including the initial +1, +3 and +5 DNAs) were detected. However, in the presence of increasing amounts of native A3G (20–80 nM) (lanes 5–7), the amounts of fully extended product and pause products were greatly reduced; with 80 nM A3G, only the +1 and +5 DNAs could be detected (lane 7). Furthermore, some of the pause products observed in the presence of A3G
differed from those made in the absence of the inhibitor (compare lanes 2–4 with lanes 5 and 6). Taken together, these results indicate that native A3G strongly inhibited tRNA-primed (−) SSDNA synthesis.

![Figure 4.2](image-url): Effect of A3G on (−) SSDNA synthesis primed by D18. (A) Time course of (−) SSDNA synthesis in reactions containing 32P-labeled D18 and RNA 244 in the presence of increasing concentrations of A3G. Positions of (−) SSDNA and D18 are shown on the right. (B) Graph of percent (−) SSDNA formed plotted versus incubation time. The percentage of (−) SSDNA product was calculated by dividing the amount of (−) SSDNA by the total amount of DNA, multiplied by 100. Symbols: 0 nM (filled circles); 20 nM (open squares); 40 nM (open circles); and 80 nM A3G (open triangles). (C) Mapping of pause sites on the RNA 244 template in A3G-containing reactions. The arrows point to the pause sites.

To investigate the kinetics of (−) SSDNA synthesis in the presence of increasing amounts of A3G, we used 5' 32P-labeled D18 in place of tRNA (Figure 4.2a). This allowed us to obtain quantitative data, which were plotted as the percent of (−) SSDNA in total
DNA products [% (−) SSDNA] versus Time (Figure 4.2b). In accord with the tRNA
experiment Figure 4.1c), A3G reduced (−) SSDNA synthesis in a dose-dependent manner
(Figure 4.2a and b). For example, at 64 min, the relative amount of (−) SSDNA
synthesized in the 80 nM reaction was decreased by 24-fold compared with the minus
A3G value. In addition, as also shown in Figure 4.1c, we observed changes in the pausing
pattern in A3G-containing reactions, which were accentuated with increasing
concentrations of A3G (Figure 4.2a). Interestingly, mapping these sites on the RNA 244
template structure (199) indicated that the pause sites occurred near or within ss
regions in the template (see Figure 4.2c). This is consistent with A3G’s well-documented,
strong preference for binding to ss nucleic acids (161,167).

The data in Figures 4.1 and 4.2 demonstrate that A3G dramatically suppressed (−)
SSDNA synthesis primed by either tRNA\textsubscript{Lys}\textsuperscript{244} or D18. The results with D18 also suggest
that the reduction of tRNA-primed (−) sssDNA synthesis by A3G was not due to an
altered configuration of the tRNA/vRNA initiation complex. Interestingly, when 6-fold
less RNA template was used, 20 nM of A3G could completely inhibit (−) SSDNA synthesis.
This indicates that A3G inhibition of (−) SSDNA synthesis is dependent on the ratio of
A3G:RNA.

It was also of interest to determine whether deaminase activity is required for
the A3G inhibitory effect on (−) SSDNA synthesis. To address this question, we used
purified deaminase-deficient A3G C291S protein (167) and then compared the effect of
WT A3G or C291S on D18-primed (−) SSDNA synthesis. As shown in Figure 4.3, the rates
and extents of (−) SSDNA synthesis in the presence of WT or mutant A3G were virtually
the same, in accord with the observation that WT and C291S proteins have similar $K_d$ values for binding to ss nucleic acids (167). The data of Figure 4.4 are significant and demonstrate that the inhibitory effect of A3G on (−) SSDNA synthesis is independent of A3G deaminase activity in our assay system.

![Figure 4.3: A3G inhibition of (−) SSDNA synthesis in the absence of deaminase activity. Symbols: no A3G (filled circles); WT A3G (80 nM) (open squares) and A3G C291S (80 nM) (open triangles).](image)

**Figure 4.3**: A3G inhibition of (−) SSDNA synthesis in the absence of deaminase activity. Symbols: no A3G (filled circles); WT A3G (80 nM) (open squares) and A3G C291S (80 nM) (open triangles).

4.3.2. Effect of A3G on minus-strand transfer reactions

Synthesis of a full-length minus-strand copy of the vRNA genome is achieved by transfer of (−) SSDNA to the 3' end of vRNA (‘acceptor RNA’) followed by RT-catalyzed extension of the annealed DNA (Figure 2.2, step 3). This process (minus-strand transfer) is facilitated by base pairing of the complementary repeat regions at the 3' ends of the nucleic acid substrates and is dependent on RNase H cleavage of vRNA sequences.
annealed to (−) SSDNA (45).

**Figure 4.4:** Effect of A3G on minus-strand transfer reactions. (A) Effect of A3G on the time course of RNase H cleavage in the absence or presence of NC. 32P-labeled TAR RNA (0.1 pmol) and TAR DNA (0.2 pmol) were heat annealed and the hybrid was incubated at 37°C in reaction buffer with 0.4 pmol HIV-1 RT with or without NC (7 nt/NC, 0.1 µM), with or without A3G (80 nM). Samples were loaded on a 15% denaturing gel. Positions of the major cleavage products are indicated on the right.

(B) Time course of annealing of 32P-labeled DNA 128 to RNA 148 incubated in the absence or presence of A3G (80 nM) with or without NC (3.5 nt/NC, 0.4 µM). Symbols: no NC/+A3G (filled circles); +NC/no A3G (open squares); and +NC/+A3G (open circles). (C) Schematic diagram illustrating the minus-strand transfer assay system. The R homology is 94 nt; U5 and U3 are 34 and 54 nt, respectively. (D) Graph of percent transfer product plotted versus incubation time. To quantify the percentage of strand transfer, the amount of transfer product was divided by the total amount of DNA, multiplied by 100. Symbols: no NC/no A3G (filled circles); no NC/+A3G (open squares); +NC/no A3G (open circles); and +NC/+A3G (open triangles).
In Figure 4.4a, we show the time course of RNase H cleavage with the heat-annealed 59-bp TAR RNA/DNA hybrid, which was incubated with and without NC, in the absence or presence of A3G. Reactions containing A3G appeared to have slightly faster cleavage rates. However, the overall cleavage pattern was the same under all four conditions. These results indicated that A3G does not interfere with RNase H cleavage and is consistent with A3G's limited ability to bind to an RNA–DNA hybrid (161,167).

To assay the effect of A3G on minus-strand transfer, we first measured NC-mediated annealing using DNA 128 and acceptor RNA 148 (see schematic diagram in Figure 4.4c) in the absence or presence of A3G (Figure 4.4b). NC is required to transiently destabilize the complementary TAR RNA and DNA structures within the repeat region, before hybrid formation can occur (45). Interestingly, A3G had only a minimal effect on annealing in the presence of NC. The rates were fairly similar with and without A3G and there was only a small reduction (11%) in the extent of the reaction when A3G was added. In a control reaction, we found that the extent of annealing minus NC, plus A3G was -10%, in accord with the value obtained in the absence of NC and A3G (56).

We also assayed minus-strand transfer with the complete system, which depends upon both annealing and RT-catalyzed elongation reactions (Figure 4.4c). The percent of total DNA present as the 182-nt strand transfer product (% Transfer Product) was quantified and plotted versus Time (Figure 4.4d). As we showed previously (197,206,207), NC significantly enhanced minus-strand transfer. However, when A3G was added, NC-mediated strand transfer was dramatically reduced (Figure 4.4d). Note
that in the absence of NC, the amount of transfer product made was extremely small and little effect of A3G was observed.

Viewed collectively, the results of Figure 4.4 strongly suggest that A3G inhibited minus-strand transfer by blocking RT-catalyzed DNA elongation. In contrast, A3G did not significantly interfere with NC-facilitated annealing of (−) SSDNA to acceptor RNA. These findings parallel the results obtained for primer placement (Figure 4.1a and b) and tRNA-primed synthesis of (−) SSDNA (Figure 4.1c).

4.3.3. Effect of A3G on (+) SSDNA synthesis and plus-strand transfer

While elongation of minus-strand DNA is being completed, the PPT RNA primer initiates synthesis of a short DNA termed (+) SSDNA (Figure 2.2, steps 4 and 5) (45). To determine whether A3G interferes with this step, we used a simple oligonucleotide assay that was previously developed in our laboratory (196,204). The 15-nt PPT was heat-annealed to a 35-nt minus-strand DNA template and the hybrid was then incubated with RT, which catalyzed the extension and subsequent removal of the PPT. The amount of 20-nt DNA product formed in reactions with and without A3G was quantified and plotted (Figure 4.5a).

The results demonstrated that A3G reduced both the rate and extent of (+) SSDNA synthesis by 2- to 3-fold. The effect of A3G on this reaction was lower than what we observed for (−) SSDNA synthesis (Figure 4.2a and b) or minus-strand transfer (Figure 4.4d). This is expected since long templates were used for those assays. By contrast, in the experiment shown in Figure 4.5a, only one or at most two molecules of A3G can bind to the available 20-nt ss region in the template (167,208). Nevertheless, it is clear
from the data that A3G was also able to inhibit RNA-primed DNA-dependent DNA polymerization by RT.

**Figure 4.5:** Effect of A3G on PPT initiation and plus-strand transfer. (A) Time course of PPT-primed plus-strand DNA synthesis. The 15-nt PPT RNA was heat-annealed to a 35-nt minus-strand DNA template and was then extended by HIV-1 RT. The 20-nt DNA product was internally labeled with [α-32P]dCTP in the absence (filled circles) and presence (open squares) of A3G (80 nM). The amount of 20-nt DNA was plotted as Relative Extension (%) versus Time (min), where 100% represents the end point value for the ‘no A3G’ reaction. (B) Time course of plus-strand transfer. The percentage of 80-nt plus-strand DNA product was calculated as described in the legend to Figure 4.4d. Symbols: no NC/no A3G (filled circles); +NC/no A3G (open circles); and +NC/+A3G (open triangles).
The plus-strand transfer reaction, like minus-strand transfer, consists of two steps: (i) NC-mediated annealing of the complementary PBS sequences in (+) SSDNA and minus-strand acceptor DNA; and (ii) RT-catalyzed elongation of both the plus- and minus-strand DNAs to yield a double-stranded DNA product (Figure 2.2, steps 6–8) (45). In our assay, only the plus-strand 32P-labeled 80-nt DNA is detected. In the absence of A3G, NC stimulated plus-strand transfer by 3-fold (Figure 4.5b), as reported previously (198). When A3G was added, reduction of synthesis was observed beginning at 10–15 min (e.g. at 15 min, by ~4-fold), whereas the extent of the reaction at 120 min was decreased by 2-fold. These results showed that A3G inhibited DNA-dependent DNA polymerization.

Taken together, the in vitro data demonstrate that A3G interfered with all of the elongation reactions catalyzed by RT. In contrast, A3G was shown to have virtually no inhibitory effect on NC-mediated annealing reactions.

4.3.4. Single molecule DNA stretching and FA-binding measurements

In an effort to understand these results, the nucleic acid binding properties of A3G, NC and RT were investigated using single molecule DNA stretching and FA. When optical tweezers are used to stretch single DNA molecules by applying forces approaching 60 pN (Figure 4.6), a force-induced melting transition occurs, in which dsDNA is converted to ssDNA (29). In the absence of protein, this transition occurs over a very narrow force range due to the cooperative melting of DNA. Saturating levels of HIV-1 NC result in a significant increase in the width of the force-induced melting transition (62), which correlates with NC’s relatively efficient chaperone
activity. In addition, the reversibility of DNA stretch/relax curves in the presence of NC (i.e., DNA stretch curves show very little hysteresis), suggests that NC has a fast nucleic acid binding on/off rate and is therefore capable of rapidly switching between dsDNA and ssDNA bound states (55).

To probe the nucleic acid binding properties of A3G, we stretched single lambda DNA molecules in the presence of varying amounts of A3G (representative curve shown in Figure 4.6a) and characterized the transition width and hysteresis. The transition width increased from 3.7 ± 0.2 pN in the absence of protein to 18.3 ± 2.4 pN at protein saturation (~90 nM A3G and above). Examination of the stretching and relaxation curves showed that there was significant hysteresis (Figure 4.6).
4.6a), which increased with A3G concentration, reaching a value of ΔG_{hysteresis} = 0.56 ± 0.02 kcal/mol per bp at 150 nM A3G, as compared to ΔG_{hysteresis} = 0.22 ± 0.06 kcal/mol per bp for saturated HIV-1 NC binding (Figure 4.6b) (55). The greater hysteresis in the presence of A3G reflects the inability of DNA to reanneal due to the presence of bound protein, which dissociates more slowly than the relaxation step time of ~1 s. In other experiments, RT was found not to have any measurable effect on lambda DNA stretching (Wang, F. and Williams, M.C., unpublished data).

The apparent binding affinities of HIV-1 NC, RT and human A3G to a 20-mer ssDNA oligonucleotide, 5'-FAM-JL587D, are given in Table 4.1. HIV-1 NC binds to the 20-mer DNA with an approximately 3-fold greater affinity than human A3G. Both A3G and HIV-1 NC bind ssDNA with a significantly higher affinity (8- and 22-fold, respectively) than HIV-1 RT. These studies suggested that A3G should compete very effectively with RT for binding to ssDNA, but would not readily displace NC. Taken together with the stretching data, these results help to explain why A3G inhibits RT-catalyzed elongation reactions, but fails to impact NC-mediated annealing.

<table>
<thead>
<tr>
<th></th>
<th>K_d (nM)^a</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIV-1 NC</td>
<td>84.1 ± 7.8</td>
</tr>
<tr>
<td>HIV-1 RT</td>
<td>1840 ± 390</td>
</tr>
<tr>
<td>A3G</td>
<td>238 ± 95</td>
</tr>
</tbody>
</table>

*Table 4.1: Apparent binding affinities of HIV-1 NC, HIV-1 RT and A3G to 20-mer DNA oligonucleotide (5'-FAM-JL587D)*

4.4. Discussion

Previous studies demonstrated that A3G inhibits reverse transcription during infection with Δvif HIV-1 (183-186), with effects on synthesis of both early and late products (160,165,168,170,190,192). In the present study, for the first time, we show how A3G affects each individual step in reverse transcription (Figure 2.2), using a series...
of well-defined, reconstituted assay systems and highly purified A3G, RT and NC proteins. The results demonstrate that A3G inhibits elongation by HIV-1 RT directly, and not by blocking NC's nucleic acid chaperone activity (45). These findings are consistent with our previous result showing that NC and A3G do not compete for binding to RNA (167) and taken together, represent strong evidence that at a physical and functional level, NC and A3G do not interfere with each other's activities.

More specifically, we report that A3G inhibits –SSDNA and +SSDNA synthesis, minus- and plus-strand DNA transfer, and elongation of minus- and plus-strand DNAs. Using a deaminase-deficient A3G mutant, we also show that (–) SSDNA synthesis was inhibited in the absence of A3G's enzymatic activity (Figure 4.3). Additionally, during –SSDNA synthesis in the presence of A3G, we find that RT pauses at unique sites, which map to bases in or near ss regions in the RNA template structure (Figure 4.2c), consistent with A3G's preference for binding to ss nucleic acids (6,12). Since a direct interaction between A3G and RT could not be detected in pull-down assays, the results further suggest that A3G binding to the template physically blocks RT movement along the template. It should be noted that A3G has a more profound effect on DNA-primed DNA polymerase activity than the ss-binding protein T4 gene 32. Thus, in a primer extension assay with a DNA template, gene 32 protein inhibited minus-strand DNA synthesis at a concentration of 400 nM. In contrast, an A3G effect was already observed at a concentration of 12.5 nM and by 50 nM, the DNA product was no longer detectable.

Systematic study of A3G inhibition of individual reverse transcription reactions is not amenable to investigation with more complex cell-based assays. However, PCR
analysis has indicated that during infection, A3G inhibits synthesis of late DNA products to a greater extent than early products (170,190,192). Conceivably, successive inhibition at each step has a cumulative effect. Recently, Guo et al. (190) showed that reduction of early reverse transcripts in A3G-expressing cells infected with Δvif virions is correlated with decreased tRNA priming in vitro. The authors suggested that an NC–A3G interaction might be inhibiting viral DNA synthesis. However, it was not determined whether A3G affected annealing and/or subsequent tRNA extension. In a new in vitro study, this group reports that in the presence of NC, A3G reduced $tRNA_{lys}^{3Lys}$ annealing by a little less than 2-fold (209). The reason for the difference between these results and our data showing that A3G does not inhibit NC-mediated primer placement (but has a strong effect on primer extension) (Figures 4.1 and 4.3) in all likelihood reflects significant differences in the experimental conditions for annealing with respect to the nt/NC ratio and the solution ionic strength.

Recently, it has been reported that A3G (170,192) and A3F (170) inhibit HIV-1 integration as well as reverse transcription. A3G and A3G C291S could each be specifically coimmunoprecipitated with NC and integrase (IN) present in HIV-1 Δvif virions; coimmunoprecipitation of A3F with virion-associated IN could also be shown (170). Sequence analysis of 2-LTR circle junction clones from unintegrated DNA synthesized in the presence of A3G showed that in some cases, DNA at the U5 end had an additional 6 RNA bases derived from the 3' terminus of $tRNA_{lys}^{3Lys}$. This suggests that A3G causes a defect in the tRNA removal step that limits plus-strand transfer and ultimately integration (192). Since we present definitive data showing that A3G does not
inhibit RNase H cleavage of RNA in a preformed hybrid (Figure 4.4a), we suggest that the aberrant cleavage observed \textit{in vivo} results from A3G binding to site(s) on the tRNA primer and/or the ssDNA template, thereby interfering with synthesis of plus-strand DNA (Figure 4.5a) and formation of a proper substrate for tRNA removal. Based on their results, Mbisa et al. (192) have made a similar proposal.

It was of interest to determine whether the A3G concentrations in the present study were within the physiological range. A recent report showed that only 7 (±4) molecules of A3G per virion are incorporated into vif-deficient HIV-1 produced from human PBMCs, indicating that only a few molecules of A3G are sufficient to inhibit HIV-1 replication (210). Interestingly, even the highest concentrations of A3G used in our experiments (up to 80 nM) are well below the estimated concentration of A3G in the virion (13 ± 8 µM). This value was obtained by assuming that the virus is a sphere of radius 60 nm (211).

Despite the high estimated concentration of A3G in the virus, the expected ratio of A3G:total nt of genomic RNA \textit{in vivo} is much lower than the ratio that is typically used \textit{in vitro}. However, RNA is highly folded and rarely single-stranded (44). Indeed, the vast majority of vRNA is likely to be involved in secondary and tertiary interactions (212-214). Since A3G binds poorly to dsRNA (161,167), the effective ratio of A3G:ss nt \textit{in vivo} is higher than predicted on the basis of total nt of RNA. Thus, the preferential binding of A3G to ssRNA (161,167) and the slow rate of A3G dissociation (Figure 4.6) could still result in A3G-induced inhibition of reverse transcription during infection.
The interpretation of the results presented here follows from our studies on the nucleic acid binding properties of NC, A3G and RT. Thus, the DNA stretching data in the presence of A3G show an increase in the force-induced melting transition width, suggesting that it is capable of binding to both ss and dsDNA. However, relative to NC, considerably higher concentrations of A3G are required to observe changes in the shape of the DNA stretching curve. Moreover, in sharp contrast to the small hysteresis observed with NC (Figure 4.6b), which decreases with increasing protein concentration (55), significant hysteresis is observed with A3G (Figure 4.6a). In this case, the amount of hysteresis increases when the protein concentration is elevated. Similar behavior is observed with the ssDNA-binding protein T7 gene 2.5 (23), whose primary role in DNA replication is to stabilize ssDNA.

The DNA stretching results indicate that A3G binds preferentially to ssDNA and is unable to rapidly switch between binding to ssDNA and dsDNA. This finding can help to explain A3G's inhibition of reverse transcription, which requires rapid access of RT to the ssDNA or RNA template. Similarly, the fact that HIV-1 NC does not interfere with DNA synthesis is consistent with its ability to rapidly adjust to different binding states. The stretching data complement the FA–binding measurements, which show that the ssDNA binding order of the three proteins is NC>A3G>>RT (Table 4.1). Taken together, these data explain why NC is not readily displaced by A3G.

A major issue in research on A3G's antiviral effect is the question of whether this activity is deaminase-dependent or -independent (or both), but there is still no clear consensus. There are data favoring the dependent mechanism, as shown in several
studies of G to A hypermutation (158,159,161) and in some reports on reverse
transcription in Δvif HIV-1-infected cells (170,192). However, there is also accumulating
evidence indicating that both types of mechanisms might be involved in A3G's effect on
reverse transcription in vivo (168,170). Our studies do not exclude a requirement for
editing, but based on biochemical and biophysical analysis, provide support for a novel
molecular mechanism that could account for deaminase-independent inhibition of
reverse transcription and virus replication that is observed in vivo.

In summary, our findings demonstrate for the first time that A3G can inhibit RT-
catalyzed elongation in a deaminase-independent manner without interfering with NC-
mediated chaperone activity. These results suggest that A3G has an intrinsic effect on
viral DNA synthesis, which is independent of the replication steps following reverse
transcription. Interplay of the equilibrium and kinetic differences between HIV-1 NC,
A3G and RT with respect to their nucleic acid binding interactions is likely to be a major
determinant of deaminase-independent A3G inhibition of RT-catalyzed DNA extension.
5. A single amino acid substitution in ORF1 dramatically decreases L1 retrotransposition and provides insight into nucleic acid chaperone activity

This work was collaborated with Sandra Martin’s lab and originally published in Nucleic Acids Research, October 2008; 36: 5845 - 5854. It’s reformatted to fit in this thesis. Bulk experiments were done by Sandra Martin’s lab and single molecule stretching experiments by Fei Wang.

5.1. Background

L1 is an autonomous mammalian retrotransposon that has successfully amplified to comprise 17 and 19% of the human and mouse genomes, respectively. Most of the >600 000 copies of mouse L1 are inactive due to truncations and/or point mutations, but 3000 are estimated to be functional for further transposition (215,216). Full-length, active copies of L1 are 7 kb in length and encode two proteins necessary for retrotransposition. The ORF1 protein (ORF1p) acts as an RNA-binding protein and nucleic acid chaperone protein in vitro (217). The RNA-binding activity of ORF1p is necessary but not sufficient for retrotransposition (1,218), and retrotransposition efficiency depends upon nucleic acid chaperone efficacy (1). The ORF 2 protein (ORF2p) has three essential domains; two of these provide the endonuclease (EN) (219) and reverse transcriptase activities (220) required for the target-primed reverse
transcription reaction (TPRT) (221) that characterizes the replication mechanism of L1 and other non-LTR retrotransposons.

Retrotransposition rates vary widely among different copies of L1. The evolution of L1 is episodic, typically characterized by one or a few distinct subtypes of L1 that dominate the dispersal process within a species and then become extinct (222). In mice, there are three subfamilies represented among the 3000 active copies of L1. These subfamilies, TF, A and GF, are distinguished by their distinct 5'-end sequences. Within each subfamily, individual members vary in their retrotranspositional activity as much as several 100-fold, as measured by an antisense-intron (AI) reporter gene assay in cultured cells (215,216). Individual elements from the currently active subfamily of human L1 similarly exhibit different activities in the cultured cell assay. A total of 40 of 82 full-length human L1 sequences in the human genome database that contain intact ORFs were able to retrotranspose when tested in cultured cells. These active elements varied widely in their retrotransposon rates, however, with most of the total retrotransposition activity of the group (84%) being attributable to just six individual elements. Significantly, of these six elements, the one with the greatest activity had an amino acid sequence most similar to the subfamily consensus (223).

A mouse L1 element on the X chromosome, T_{FC}, has a sequence most like the consensus of the TF subfamily and was found to retrotranspose 15 times more efficiently than another element of the same subfamily, T_{Fspa} (1). T_{Fspa} recently inserted into the beta-glycine receptor gene, hence it is a known active mouse L1 (224). A total of 20 nt substitutions, including three that cause amino acid replacements, distinguish the
two elements. The goal of this study was to define the substitution responsible for this
dramatic effect on L1 retrotransposition and determine its mechanism of action.

The significant substitution mapped to one of the altered amino acids in ORF1 far N-
terminal to the previously described nucleic acid interaction domain of the ORF1 protein
(2,225,226). The substitution affects a late step in retrotransposition and significantly

Figure 5.1: Locating the substitution(s) responsible for elevated retrotransposition of T_Fc (white box) versus T_Fspa (grey box). (A) Structure of T_Fc. From 5’ to 3’, monomers (small rectangles, with promoters), 5’ nc region (line), ORF1 with relative locations of the coiled-coil domain (C-C) and the conserved domain [comprised of the middle or M, domain and the C-terminal domain, CTD (2)], ORF2 with the endonuclease (EN) and reverse transcriptase domains (RT), the 3’ nc region (line), followed by a polyA tail (pA). ORF1 and ORF2 are not aligned because they are separated by a 40 nt intergenic region (4). (B) Autonomous retrotransposition assay with chimeric L1s, grey boxes indicate the locations of the restriction fragments from T_Fspa that were substituted into the T_Fc backbone: NheI to BstWI with two replacement substitutions in ORF1 or BstWI to SspI with one silent and one replacement substitution in ORF2, as depicted in panel A. The fold change in retrotransposition as measured with Ai-eGFP is given relative to T_Fspa. (C) Retrotransposition assay on T_Fc (grey) and T_Fspa (white) elements with single amino acid substitutions. Retrotransposition causes G418R; representative wells show stained G418-resistant cells after 10 d of selection. The uniform dark stain is due to confluent cells, whereas the spots are colonies of G418R cells, indicative of fewer retrotransposition events.
alters the nucleic acid chaperone activity of the ORF1 protein in vitro. The results of this work strengthen the hypothesis that the nucleic acid chaperone activity of ORF1p is required for TPRT during L1 retrotransposition (1,217), as well as increase our understanding of the mechanism of action of nucleic acid chaperone proteins.

5.2. Materials and Methods

5.2.1. Constructs

TFc and TFspa constructs for the autonomous retrotransposition assay were described previously (1). Chimeric constructs that place either the two ORF1 replacement substitutions or the single ORF2 replacement substitution of TFspa into the backbone of TFc were made by moving either the NheI-BstWI or the BstWI-SspI fragments, respectively, from TFspa into the homologous sites of TFc (Figure 5.1b). The single point mutations to reciprocally alter the two ORF1 amino acids that differ between TFc and TFspa were made by site-directed mutagenesis in either a TFc or a TFspa subclone. The NheI-BstWI fragment containing the mutation was then used to replace the homologous fragment of the intact L1 in the retrotransposition assay vector after verification that the desired point mutation was the only change by DNA sequencing.

5.2.2. Cell culture and autonomous retrotransposition

Retrotransposition assays were done in 143B cells as described previously (1), using either G418 resistance or expression of eGFP as the marker of retrotransposition events. Briefly, these reporter cassettes measure retrotransposition because expression of the marker requires excision of an AI in the L1 transcript, then conversion to cDNA
and insertion into the genome by TPRT before the reporter can be transcribed into an mRNA that encodes functional protein, i.e. before either growth on medium containing G418 or detection of green fluorescence, is possible. Cells were transfected the day after seeding using Lipofectamine2000 (Invitrogen, Carlsbad, CA, USA), according to the manufacturer's recommendations. Transformants were selected for 24 h in 10 µg/ml puromycin beginning 24 h after transfection. For experiments involving the eGFP marker, fluorescence was followed daily by microscopy and quantified by flow cytometry 6 days post-transfection as described (1). For experiments involving the neo marker, cells were transfected as for eGFP, except that cells were allowed to recover for 1 day after puromycin treatment, then placed in G418 (400 µg/ml) and allowed to grow for 10 additional days before fixing and staining with crystal violet. Transfections were identical for immunoflorescence assays except they were done using cells plated on polylysine-coated coverslips.

5.2.3. Immunofluorescence microscopy of L1 ORF1p

L1-transfected 143B cells were rinsed with phosphate buffered saline (PBS) and then fixed for 20 min in 4% paraformaldehyde at various times post-transfection. The fixed cells were rinsed in PBS, then blocked and permeabilized for 2 h in PBS containing 3% BSA and 0.1% Triton X-100 (block). The coverslips were incubated overnight at 4°C with rabbit polyclonal anti-ORF1p antibody (3.4 µg/ml) in block and then rinsed three times in block. Cells were then incubated with 1:500 Cy-5 conjugated goat anti-rabbit antibody (Jackson ImmunoResearch, West Grove, PA, USA) for 1 h at 4°C, rinsed three
times in block and mounted on slides with Fluoromount G (Southern Biotech, Birmingham, AL, USA). Fluorescence was imaged and captured using a Zeiss LSM510.

5.2.4. RNA, DNA and protein analyses from transfected cells

Timepoints were taken from L1-transfected 143B cells by harvesting cells every 24 h post-transfection. Cells were recovered with trypsin, washed in PBS and stored frozen as cell pellets at −80°C. Pellets were resuspended in five volumes of PLB (140 mM NaCl, 200 mM Tris–HCl, pH 8.5, 2 mM MgCl₂) to which a 1/20 volume of 5% NP40 was added. After gentle mixing followed by 5 min on ice, the lysate was centrifuged 10 min at 2000g. The supernatant was recovered for RNA and protein analysis and the pellet was used for DNA analysis. About 10 µg/ml protease inhibitors (P8340 Protease Inhibitor Cocktail, Sigma-Aldrich, St. Louis, MO, USA) were added to aliquots collected for western blot analysis.

Retrotransposition events were detected at the DNA level by PCR amplification of the spliced reporter gene. About 200 ng of total DNA from various days post-transfection were amplified for 20 cycles with primers A7 (5’-CGTCCATGCGGAGGTGTATCC) and A8 (5’-GCTACGTCCAGGGAGCGCACCATC), followed by 35 cycles with primers A16 (5’-GCTACGTCCAGGAGCGCACCATC and A8).

ORF1p was detected by western blotting using 20 µg of transfected cell lysate (Bradford Assay, Bio-Rad, Hercules, CA, USA) and affinity purified, rabbit anti-ORF1p antibody (227) with the HRP-conjugated goat anti-rabbit antibody Plus Western Blotting Reagent Pack (GE Healthcare, Pittsburg, PA, USA). Images were captured on a Typhoon 9400 (GE Healthcare). Anti-actin antibody (sc-8432) was purchased from Santa Cruz
Biotechnology (Santa Cruz, CA, USA) and used as recommended by the manufacturer to detect actin on the same blots.

L1 RNA was detected and quantified using RT–PCR. Two microgram of RNA isolated from PLB supernatants using TRIzol LS (Invitrogen) were treated with RQ1-DNase (Promega, Madison, WI, USA) according to the manufacturer’s instructions. One microgram of the DNase-treated RNA was added to 20 µl reverse transcriptase reactions as recommended by the manufacturer (Reverse Transcription System, Promega). Reactions were diluted to 100 µl in nuclease-free water and 10 µl were used in 25 µl PCR reactions with primers that amplify a 265 nt region of ORF2 (172 nt downstream of the AUG) in L1 from both T_FC and T_Fspa. The oligonucleotides were 5'-GACACTACCTCAGAATCAAAGGCTGG (forward primer) and 5'-GTGAGGCGCAATGTGTGCTTTGAGC (reverse primer). It was empirically determined that 23 or 25, cycles remained within the linear range of the assay, on days 1–3 or 4–6, respectively. PCR products were separated by electrophoresis through 2% agarose gels, and then stained with ethidium bromide. Fluorescence images were captured on the Typhoon 9400 and analysed using ImageQuant (GE Healthcare).

5.2.5. Assays using purified proteins

D159 (T_FC) and H159 (T_Fspa) ORF1 proteins were purified to apparent homogeneity after expression in baculovirus-infected insect cells as described (1). Briefly, insect cells were lysed 48 h post-infection and the His-tagged ORF1 proteins were purified using affinity (Ni-NTA agarose, Qiagen, Valencia, CA, USA) chromatography followed by size-exclusion (Sephacryl S300, GE Healthcare).
chromatography. RNA binding, strand exchange and melting temperature (Tm) assays were done as described previously (1,217,228).

DNA stretching experiments were performed using a dual beam optical tweezers instrument, described in previous chapters (29). To quantify the effects of these proteins on the DNA helix–coil transition, we calculated the transition width as a function of protein concentration, as described previously (1). The results were fit to the McGhee–von Hippel binding isotherm:

\[
\Theta = K \cdot n \cdot C \cdot \frac{(1-\Theta)^n}{\left(1-\Theta + \frac{\Theta}{n}\right)^{n-1}}
\]

(5.1)

where \( \Theta \) is the fractional change in transition width, \( K \) is the equilibrium association constant for protein binding to DNA and \( n \) is the binding site size. For simplicity we set \( n = 1 \) so that the equilibrium constant is the only fitting parameter.

Kinetic experiments on oligonucleotides were performed on an upgraded Biacore 3000 instrument at 25°C. The 5'-biotinylated AAAAAAGTACACAGTCTAACTCAACTCGC was annealed to either 5'-GCGAGTTGATGGTAGACTGTGTACTTTTT to make a perfectly matched, dsDNA duplex or to 5'-GCGAGTTGACGTCAGACCGTGCACTTTTT to make the mismatched dsDNA duplex. Biotinylated oligonucleotide was captured on a CM4 chip first derivitized with NeutrAvidin biotin-binding protein (Pierce, Rockford, IL, USA) via amine coupling. dsDNA constructs (perfectly matched and mismatched) were hybridized on the chip in running buffer (50 mM phosphate buffer, 250 mM NaCl, 0.1 mM EDTA, pH 7.6). The instrument
was programmed for iterative cycles in which each kinetic cycle consisted of: (i) 300 s protein injection phase, (ii) 300 s or greater dissociation phase depending on affinity and (iii) a 120 s regeneration phase. A flow rate of 20 µl/min was maintained throughout the cycle. The concentration of proteins analysed ranged from 10 to 300 nM. The surface plasmon resonance (SPR) signal was recorded in real time every 0.5 s. Each sensorgram obtained was corrected for bulk refractive index changes by subtracting the corresponding protein injection cycle on a blank NeutrAvidin surface. The association and dissociation rate constants (\(k_{\text{on}}\) and \(k_{\text{off}}\), respectively) for the interaction were calculated by globally fitting the data using different kinetic models available in the BIA evaluation software package with a simple 1:1 bimolecular Langmuir interaction model.

5.3. Results

T\(_{\text{FC}}\), a mouse L1 element with the sequence of the TF subfamily consensus, retrotransposes 15-fold more frequently than T\(_{\text{Fspa}}\) in an autonomous retrotransposition assay (1). The 20 nt substitutions that distinguish these two elements are represented schematically in Figure 5.1a. Ten of these are in the monomers, which potentially could impact transcription and hence retrotransposition \textit{in vivo}. However, the monomers are not present in the constructs used for the autonomous assay, rather the CMV promoter is used to drive transcription of both TFspa and T\(_{\text{FC}}\); thus, those 10 substitutions cannot account for the difference observed between the two elements using this assay. Just three of the remaining 10 nt substitutions cause amino acid replacements: two in ORF1 and one in ORF2. In addition, there are three single nucleotide substitutions in ORF2.
that are silent at the amino acid level, another in the 5' non-coding region and three in the 3' non-coding region.

The ORF2 replacement altered a conserved EN domain in ORF2p (229), whereas the two ORF1 replacements both lie in a relatively non-conserved coiled-coil forming region of ORF1p (2,230). All three of the replacement substitutions were rare among the 539 aligned sequences (1) that were used to derive the consensus. In ORF1, D53G was present in just five of the sequences including T\textsubscript{Fspa}, D53N occurred once and D159H was present only in T\textsubscript{Fspa}. In ORF2, F224L also appeared only once in T\textsubscript{Fspa}, but two of the sequences had F224Y and two others had F224C. Chimeric L1 constructs were made to test the importance of the ORF2 versus the two ORF1 substitutions in the retrotransposition assay by replacing either of two restriction fragments of T\textsubscript{FC} with the homologous fragments from T\textsubscript{Fspa}; these two chimeric elements are both comprised of largely T\textsubscript{FC}, but introduce either the single amino acid replacement (together with one nearby silent substitution) of ORF2, or the two amino acids of ORF1 from T\textsubscript{Fspa} into the T\textsubscript{FC} backbone. The elevated retrotransposition activity of T\textsubscript{FC} unambiguously mapped to the fragment containing the two replacements in ORF1, rather than to the fragment which altered ORF2 (Figure 5.1b). To determine whether one or both amino acid replacements in ORF1 were important for the enhanced activity of T\textsubscript{FC} and the diminished activity of T\textsubscript{Fspa}, as well as to eliminate the possibility that one of the ‘silent’ substitutions elsewhere was critical, the two ORF1 amino acid replacements of T\textsubscript{FC} were individually introduced into T\textsubscript{Fspa} and the two replacements of T\textsubscript{Fspa} were likewise introduced into T\textsubscript{FC}. Subsequent retrotransposition assays with these altered elements
identified the aspartic acid at position 159 (D159) in ORF1p as the crucial amino acid responsible for most, if not all, of the 15-fold increase in retrotransposition activity exhibited by Trfc (Figure 5.1c).

![Image](image_url)

**Figure 5.2:** Timecourse of L1 retrotransposition. (A) Representative micrographs of 143B cells captured daily after transfection with Trfc or Trfspa. Cells with new insertions by retrotransposition are green, cells expressing ORF1p are red, each field is 1.3 mm². Arrows on images for days 3 and 4 point to clusters containing at least two green cells. (B) Quantification of cells expressing eGFP following transfection. (C) Schematic of the L1 retrotransposition reporter, showing location of primers used for RT–PCR detection of unspliced and spliced eGFP DNA. The antisense eGFP gene is depicted as a white rectangle interrupted by a grey intron (in the same orientation as the L1 ORFs). Amplification across the intron results in a 1334 nt PCR product from transfected DNA, whereas after splicing and cDNA synthesis due to retrotransposition the product is 263 nt. (D) Ethidium bromide-stained agarose gels of PCR products obtained following a two-step PCR amplification of genomic DNA isolated from transfected cells (days 1–6 post-transfection with Trfc or Trfspa) or the control transfected plasmid DNA (lane C). PCR was 20 cycles of amplification using the A7 and A8 primers followed by a second amplification using the primers A16 and A8. A16 spans the splice junction. The 1334 nt band amplified from transfected DNA containing the intron is readily distinguished from the 263 nt band amplified from retrotransposed DNA and thus lacking the intron. The intronless band is observed earlier and is more intense throughout the timecourse in DNA from Trfc-transfected cells compared with Trfspa-transfected cells. The marker lane (M) shows the bands from 2000 to 200 nt from the 1 Kb Plus DNA Ladder (Invitrogen).
The kinetics of retrotransposition, as well as the expression of L1 RNA and ORF1p were examined in cells transfected with L1 T<sub>FC</sub> or T<sub>Fspa</sub> marked with the eGFP Ai reporter. Transfected cells were examined daily for the presence of green cells, which indicate retrotransposition of L1 from the transfected plasmid into genomic DNA (231). Cells transfected with T<sub>FC</sub> expressed eGFP at higher frequency throughout the timecourse compared with those transfected with T<sub>Fspa</sub>; green cells were also initially apparent on day 3 post-transfection with T<sub>FC</sub>, versus day 4 with T<sub>Fspa</sub> (Figure 5.2a and b). Interestingly, simultaneous detection of ORF1p by immunofluorescence suggests that ORF1p accumulates similarly when expressed from the two elements (Figure 5.2a).

![Figure 5.3: Timecourse of ORF1p expression. Representative western blot of proteins recovered from untransfected cells (-t) or after transfection with T<sub>FC</sub> (C) or T<sub>Fspa</sub> (S) L1 constructs at days 1–4 and 6 following transfection. Lanes marked 3 ng contain baculovirus-expressed ORF1p fusion protein (47 kDa). Sizes on the left are in kDa, the position of native ORF1p is indicated by the arrow. The actin signal from the same lanes on these blots is shown below to verify approximately equal loading.](image)

because there is no evidence for a lag in expression in cells transfected with T<sub>Fspa</sub> compared with T<sub>FC</sub>, nor does there appear to be any difference in either the intensity or the subcellular localization of ORF1p over the timecourse examined (see Figure 5.3).

With both T<sub>FC</sub> and T<sub>Fspa</sub>, retrotransposition leading to eGFP expression occurred in a
minor fraction of the ORF1p expressing cells.

Figure 5.4: Timecourse of L1 RNA expression. 143B cells were transfected with either TFC or Tspa containing the eGFP Ai reporter gene in the 3′-UTR, and then harvested on the indicated days following transfection for isolation of RNA. (A) Schematic of L1RNA and location of primers to detect L1RNA; these primers amplify a 265 nt region near the 5′-end of ORF2. (B) Fluorescence image of ethidium bromide-stained gels after RT–PCR; each set of four lanes contains the PCR product from the same DNAse-treated RNA RT–PCR in triplicate plus a control reaction without reverse transcriptase. Twenty-three cycles of PCR were performed on cDNA from days 1 and 2, and 25 for days 4 and 5, to remain in the linear range of the assay. (C) Steady-state levels of L1 RNA as determined by semi-quantitative RT–PCR using primers within ORF2. PCR bands were quantified from fluorograms and are plotted as mean ± SD from triplicates; the actual above background intensity values for days 4 and 6 were divided by four to correct for the cycle number difference between these 2 days and days 1–3. This experiment is representative of the results obtained in three separate transfection experiments. The steady-state levels of L1 RNA in transfected cells were not significantly different between TFC (grey bars) and Tspa (white bars) on any day across the time course.

The results of the imaging studies were corroborated and extended using biochemical techniques. The finding that transposition occurred earlier in cells expressing TFC than Tspa was confirmed by PCR amplification of genomic DNA using a strategy that allowed simultaneous detection of the transfected eGFP DNA (intron present) and the retrotransposed copy (intron absent, Figure 5.2c). The spliced eGFP gene, indicative of successful retrotransposition, appeared earlier in genomic DNA after transfection and accumulated to higher levels in TFC-transfected cells compared with those transfected with Tspa, as expected based upon the results obtained by following the appearance of green cells. Consistent with the intensity and distribution of ORF1p
observed by indirect immunofluorescence, the steady-state levels of ORF1p detected by western blotting did not differ between the two elements over the timecourse examined (Figure 5.3). Likewise, there were no significant differences in the steady-state level of L1 RNA detected by semi-quantitative RT–PCR between $T_{FC}$ and $T_{Fspa}$ (Figure 5.4).

ORF1p is an RNA-binding protein that forms large complexes with L1 RNA (5,12,16,17,21). It is also a nucleic acid chaperone; mutations that compromise chaperone activity block or diminish retrotransposition (1). In order to examine these activities of ORF1p, we isolated the D159 ($T_{FC}$) and H159 ($T_{Fspa}$) variant forms of ORF1p from baculovirus-infected insect cells. The two proteins behaved identically throughout protein purification, eluting from size-exclusion chromatography in the identical fraction characteristic of the elongated trimer form of the protein as described previously for the $T_{Fspa}$, H159 variant (230). The ellipticity and $T_m$ of these two purified proteins were also equivalent as determined by circular dichroism (–26 000 and –26 800 and 51 and 49.5°C, for H159 and D159, respectively).

The affinity of both proteins for RNA was measured using a nitrocellulose filter-binding assay. In a side-by-side comparison containing increasing concentrations of D159 or H159 ORF1p and 25 pM of an antisense 111 nt L1 RNA in 250 mM NaCl, the apparent $K_d$ of D159 for RNA was 8.9 ± 1.0 nM and H159 was 7.0 ± 0.6 nM, respectively. This 20% change is not significant based upon multiple experiments and is unlikely to explain the 15-fold decrease in L1 retrotransposition associated with D159H because a 38% drop in apparent affinity for RNA by the ORF1p mutant, R298K, decreases retrotransposition by just 56% (1).
The nucleic acid chaperone activity of ORF1p can be assessed by determining the 
$T_m$ of a mismatched dsDNA oligonucleotide in the presence of protein (1). About 30 nM 
H159 ORF1p shifts the $T_m$ of a 29 nt dsDNA oligonucleotide with four non-contiguous 
mismatches from 42°C to 22°C. At the same concentration of protein, D159 ORF1p has a 
significantly different effect, with only a small fraction of the mismatched double-
stranded oligonucleotide being converted to single-stranded form (Figure 5.5a).

**Figure 5.5**: $T_m$ assay for nucleic acid chaperone activity of ORF1p. Addition of ORF1p alters the $T_m$ of a 
mismatched 29 nt dsDNA oligonucleotide as measured by conversion of the double-stranded form to 
single-stranded form. (A) ORF1p from TFC (D159) and TFSpa (H159) at the same concentration (trimer) 
compared with naked DNA. (B) $T_m$ of the mismatched oligonucleotide in the presence of increasing 
concentrations of D159 ORF1p or (C) H159 ORF1p. ORF1p (trimer) concentrations are identical to 
those in (B). The temperature range needed to fully denature the DNA in the absence of protein is 
shown; both proteins begin to denature at 35°C, which enhances their charge neutralization

This effect occurs over a broad concentration range of protein (compare b and c in 
Figure 5.5), and suggests that H159 ORF1p interacts more strongly with single-stranded 
DNA than D159 ORF1p as the duplex transiently or fully melts.

This difference in the interaction of ORF1p with a mismatched DNA 
oligonucleotide was explored in more detail using SPR. The interactions of both ORF1p 
proteins with ssDNA and with dsDNA containing either perfect or imperfect 
heteroduplex were examined for comparison. The interactions of the two proteins with
the ssDNA oligonucleotide, as well as with the perfectly matched dsDNA oligonucleotide were similar. In contrast, a relatively large difference was observed between these two proteins in their interaction with the imperfect double-strand duplex of the same length; D159 ORF1p displays rapid kinetics of association and dissociation with the mismatched duplex whereas H159 ORF1p dissociates 10 times more slowly (Table 5.1).

<table>
<thead>
<tr>
<th></th>
<th>ssDNA binding</th>
<th>dsDNA perfect binding</th>
<th>dsDNA mismatch binding</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$k_a$</td>
<td>$k_d$</td>
<td>$K_D$</td>
</tr>
<tr>
<td>H159</td>
<td>4.7E+04</td>
<td>6.0E–04</td>
<td>1.3E–08</td>
</tr>
<tr>
<td>D159</td>
<td>3.7E+04</td>
<td>4.3E–04</td>
<td>1.2E–08</td>
</tr>
</tbody>
</table>

Table 5.1: Kinetics of interaction of ORF1p with short DNA oligonucleotides

Single-molecule analysis of DNA stretching is a sensitive assay for nucleic acid chaperone activity (29). Typical DNA stretching experiments in the absence of protein and in the presence of 15 nM D159 or H159 ORF1p are shown in Figure 5.6a. The stretching curves (solid lines) for DNA in the presence of both proteins show significant changes in the shape of the helix–coil transition. In the absence of protein, at a force of about 60 pN, the DNA stretching force increases very little as the DNA extension is increased by a factor of about 1.7. This plateau represents a cooperative DNA helix–coil transition, and the length at which the force begins to increase dramatically at the end of the transition (indicated by an arrow) demonstrates almost complete conversion of the DNA from double- to single-stranded form. The width of the transition, which is smaller for a more cooperative transition, is only about 4 pN in the absence of protein.
In contrast, the transition width (or the force change from the beginning to the end of the plateau) is much higher in the presence of both proteins, but D159 ORF1p shows a much greater increase in the transition width relative to that observed for H159 ORF1p at this protein concentration. Finally, both proteins induce a change in the extension at which the ssDNA stretching force increases at the end of the transition, shown by the arrows. This change in ssDNA extension represents ssDNA aggregation, in which protein-induced effects make ssDNA attracted to itself, thus effectively decreasing the ssDNA length at a given force. The magnitude of the ssDNA aggregation is clearly much greater in the case of H159 ORF1p. To quantify the effects of these proteins on the DNA helix–coil transition, we calculated the transition width as a function of protein concentration,

![Figure 5.6: Effect of ORF1p on DNA as measured by DNA stretching. (A) Effect of T_{FC} (D159) and T_{Fspa} (H159) ORF1 proteins on the helix–coil transition of λDNA. Stretching (continuous lines) and relaxing (dotted lines) curves for λDNA, either without protein (black) with 15 nM D159 (blue), or H159 (red) ORF1p from T_{FC} and T_{Fspa}, respectively. Arrows indicate ssDNA contour length. The average maximum stretching force was the same for both forms of ORF1p. (B) Binding titrations for ORF1p T_{FC} and T_{Fspa} variant binding to DNA as measured by DNA stretching. Scattered data points represent measurements and continuous curves are fits to the data using the McGhee–von Hippel isotherm (as shown in Equation 1 in Methods section). For fits to D159 ORF1p titration data (blue), we obtain a saturated transition width of 27.5 ± 1.2 pN and an equilibrium association constant of 0.78 ± 0.08 x 10^8 M^{-1}. For H159 (red), we obtain a saturated transition width of 19.2 ± 0.8 pN and an equilibrium association constant of 2.1 ± 0.2 x 10^8 M^{-1}.](image)
as described previously (1). The results are shown as data points in Figure 5.6b, along with lines that represent fits to the McGhee–von Hippel binding isotherm (Methods section). A saturated transition width of 27.5 ± 1.2 pN and an equilibrium association constant of 0.78 ± 0.08 x 10^8 M\(^{-1}\) (\(K_D = 1.28 ± 0.13 x 10^{-8} M\)) was calculated from the fit to the data obtained using D159 ORF1p, compared with a saturated transition width of 19.2 ± 0.8 pN and an equilibrium association constant of 2.1 ± 0.2 x 10^8 M\(^{-1}\) (\(K_D = 0.48 ± 0.05 x 10^{-8} M\)) from the data obtained in the presence of H159 ORF1p. These equilibrium binding constants for ssDNA determined from DNA stretching experiments agree well with those presented in Table 5.1 from bulk measurements.

5.4. Discussion

The human and mouse genomes contain hundreds or thousands of intact and therefore potentially retrotransposition-competent L1s, respectively. The activities of individual L1 elements vary widely when tested in a cultured cell assay (215,216,223). Analogous variations in retrotransposition activity are also detected on an evolutionary timescale through phylogenetic analysis of genomic sequences (232-234); thus this property of individual variation in the ability to produce progeny is likely intrinsic to L1 and not an artifact of the cultured cell assay. \(T_{\text{Fspa}}\), the first active mouse L1 described, was isolated because it had retrotransposed into the glycine receptor beta subunit and disrupted the expression of this critical neurotransmitter receptor gene (224,235). \(T_{\text{FC}}\), on the other hand, was predicted to be an active element based upon its near identity to the consensus sequence of all of the TF elements in the mouse genome that are closely related to \(T_{\text{Fspa}}\) (1). Although both elements are active, \(T_{\text{FC}}\) retrotransposes 15-
fold more effectively than T_{Fspa} in the assay used here. The goal of this investigation was to determine the molecular basis for this difference in order to gain insight into the mechanism and control of L1 retrotransposition.

The increased activity of T_{FC} was mapped to aspartate 159 in ORF1p (Figure 5.1), which affects a step that follows the accumulation of L1 intermediates (RNA and ORF1p) but precedes the successful insertion of a new cDNA copy of L1 into genomic DNA (Figures 5.2–5.4). Thus, the L1 expression products, RNA and ORF1p, are more effectively converted to new insertion events in T_{FC} compared with T_{Fspa}. RNA binding and nucleic acid chaperone activities are the only two essential retrotransposition functions presently attributable to ORF1p (1,218). This study revealed that H159 ORF1p has a diminished nucleic acid chaperone activity compared with D159 ORF1p (Figures 5.5 and 5.6), although its affinities for RNA, a short ssDNA oligonucleotide, a perfectly base paired short dsDNA oligonucleotide and a long dsDNA were not affected (Table 5.1, Figure 5.6). R297K and RR297:298KK substitutions in ORF1p disrupt nucleic acid chaperone activity without affecting RNA binding, but RR297:298AA causes a significant drop in RNA affinity and also destroys nucleic acid chaperone activity (1). Hence, no mutations in ORF1p are known to significantly reduce RNA binding without disrupting nucleic acid chaperone activity; such mutations would be useful for understanding the relationship between the RNA binding and nucleic acid chaperone activities of ORF1p.

The location of the amino acid substitution responsible for the improved nucleic acid chaperone function of ORF1p and the elevated retrotransposition activity of T_{FC} was surprising based upon our understanding of the structure and function of L1 ORF1p.
prior to this work. Primary sequence analysis had revealed two domains in ORF1p: (i) an N-terminal, coiled-coil domain which is highly divergent and perhaps not homologous among mammalian L1s and (ii) a C-terminal conserved domain that is shared among all mammalian L1s and some non-LTR retrotransposons in fish (236). Results of several studies indicate that the coiled-coil region is necessary and sufficient for the formation of highly stable ORF1p homotrimeres, and the basic region of the conserved domain is likewise both necessary and sufficient for high-affinity interactions between ORF1p and RNA (225,226,230). The crucial residue for high retrotransposition activity identified here, D159, lies within the coiled-coil domain, but near its C-terminus (2); this is the first report of residues that are critical for functional protein–nucleic acid interactions in this region of ORF1p. Based upon the transfer of 32P from RNA to protein, a polypeptide with residues 244–371 of ORF1p binds RNA, but one with residues 1–251 does not (226). In addition, four consecutive alanine substitutions for REGK beginning at residue 235 in ORF1p from human L1 (homologous to 271 in mouse ORF1p) alter interaction of ORF1p with RNA (218). Thus, D159 in mouse L1 ORF1 is at least 85 amino acids N-terminal to the closest residue previously shown to be involved in nucleic acid interactions.

Although it is possible that these regions are adjacent to one another in the presently unknown 3D structure of ORF1p, our data strongly suggest the presence of a heretofore unrecognized site on the protein for nucleic acid interactions.

The binding constant calculated from DNA stretching was similar for D159 and H159 ORF1 proteins, in agreement with the SPR analysis results shown in Table 5.1. In addition, the amount of hysteresis observed between the stretching and relaxation
curves is also similar for these two variants. There are two primary differences between the DNA stretching results for D159 ORF1p and those obtained for H159 ORF1p. First, the amount of ssDNA aggregation is much greater for H159 relative to that observed in the presence of D159. Second, the maximum helix–coil transition width as predicted from fits to the binding titrations is greater for D159 ORF1p by about 40%. An increase in helix–coil transition width is positively correlated with nucleic acid chaperone activity (24,25,31), however, this increase in transition width with protein binding involves the combination of several effects related to chaperone activity that are not easily separated (62).

Recent studies demonstrated that the primary determinants of efficient nucleic acid chaperone activity are the capability to induce nucleic acid attraction, the ability to partially but not completely destabilize the DNA helix, and the ability to rapidly switch between ssDNA- and dsDNA-binding modes (45,55). Therefore, the chaperone activity of a specific protein is the result of a balance between competing effects. For example, the ability to induce DNA attraction, or aggregation of DNA, facilitates nucleic acid rearrangements by bringing complementary strands together. Conversely, this property also tends to stabilize the DNA helix and inhibit the mobility of the DNA–protein complex, thereby inhibiting rearrangements of nucleic acid secondary structure. HIV-1 nucleocapsid protein (NC) is a well-studied example of a protein that has optimized these various competing effects (45). Subtle changes in the architecture of the zinc fingers in NC destroy this delicate balance, resulting in an inefficient nucleic acid chaperone that is defective in retroviral replication (55).
The two ORF1p proteins studied here exhibit all of the characteristics of a nucleic acid chaperone, but to subtly different extents. They both aggregate DNA, bind preferentially to ssDNA and therefore stabilize the DNA helix, and do not strongly inhibit annealing of long DNA strands. These general nucleic acid chaperone characteristics were previously demonstrated using DNA stretching measurements for the T\textsubscript{Fspa} ORF1p (1), and are also apparent in the results of DNA stretching experiments with T\textsubscript{FC} ORF1p (Figure 5.6a). More subtle features of the stretching experiments distinguish the two ORF1 proteins, however; the results show significantly lower aggregation for the T\textsubscript{FC} ORF1p, suggesting that DNA complexes with this protein will have increased mobility and therefore increased nucleic acid chaperone activity. The hypothesis that lower aggregation of ssDNA results in increased mobility of protein–DNA complexes is supported by the SPR results, which provide evidence of more rapid binding kinetics for the T\textsubscript{FC} ORF1p with a mismatched dsDNA oligonucleotide compared with the ORF1p from T\textsubscript{Fspa}. It is also consistent with much stronger effect in melting the mismatched dsDNA oligonucleotide in the gel-based Tm experiments. Thus, the biophysical data presented here are fully consistent with the observed stronger nucleic acid chaperone capabilities of D159 compared with H159 ORF1p.

The results of earlier studies established that mutations in the C-terminus of T\textsubscript{Fspa} ORF1p greatly inhibited nucleic acid chaperone activity, which in turn abolished L1 retrotransposition (1). There, the primary effect of the mutations on DNA stretching was to induce such strong ssDNA and dsDNA aggregation that the DNA could not be melted by force. Those earlier results illustrated that aggregation that is too strong inhibits
chaperone activity. This hypothesis is supported by other studies in which similar ssDNA aggregation effects were observed for DNA stretching in the presence of HIV-1 Gag, a nucleic acid packaging protein (62), as well as in the presence of HIV-1 NC variants that were identified as poor nucleic acid chaperones (55). Increased DNA aggregation that results in decreased chaperone activity could be due directly to changes in a DNA-binding region of the protein, as is likely the case with mutations in the zinc finger regions of HIV-1 NC (55). It is possible that a tighter interaction with nucleic acids by H159 is a simple reflection of the more basic nature of histidine than glutamate. If D159H does directly alter a DNA binding site in ORF1p, it is not likely to be the previously known site because interactions of ORF1p with nucleic acids have consistently mapped to the distant C-terminal third of the protein (2,218,225,226). The known, C-terminal binding site is apparently unaltered by the D159H substitution, based upon the indistinguishable affinities of D159 and H159 ORF1p for RNA and ssDNA.

Alternatively, increased DNA aggregation that inhibits chaperone activity could also occur with amino acid substitutions that increase or alter homotypic protein–protein interactions, as is likely the explanation for the increased aggregation that was observed for HIV-1 Gag and the Gag cleavage product NCp9. For these two proteins, the DNA binding site remained intact, yet aggregation was increased and nucleic acid chaperone activity decreased compared with NCp7 (62). Given the location of D159H in the coiled-coil domain (2) where it is likely exposed to solvent, it is plausible that this residue could be involved in further protein–protein interactions between trimers. The differences observed between D159 and H159 in the assays involving short
oligonucleotides are not easily explained by such protein–protein interactions, however, because each trimer occupies 50 nt (225). Hence, the 29 nt oligos used for Tm and SPR studies would only bind a single ORF1p trimer and would not be expected to interact with more than one trimer at a time.

A full understanding of the interactions of ORF1p with single and double-stranded nucleic acids, and the relationship between its high-affinity RNA-binding function and the nucleic acid chaperone function necessarily awaits a high resolution structure of the protein with and without its various nucleic acid ligands. Nevertheless, the results presented here illustrate the importance of maintaining a delicate balance between strong DNA binding and DNA–protein complex mobility for efficient nucleic acid chaperone activity. This work also lends further support to the conclusion that the nucleic acid chaperone activity of ORF1p plays an essential role during L1 retrotransposition; this role likely occurs late in the process, consistent with a function in facilitating the strand exchanges that are required to initiate TPRT or melting secondary structure in the RNA template for reverse transcription (217).
Reference:


