Nucleic acid binding and chaperone properties of HIV-1 Gag and nucleocapsid proteins

Margareta Cruceanu1, Maria A. Urbaneja2, Catherine V. Hixson2, Donald G. Johnson2, Siddhartha A. Datta3, Matthew J. Fivash4, Andrew G. Stephen5, Robert J. Fisher5, Robert J. Gorelick2, Jose R. Casas-Finet6, Alan Rein3, Ioulia Rouzina7 and Mark C. Williams1,8,*

1Department of Physics, Northeastern University, 111 Dana Research Center, 110 Forsyth Street, Boston, MA 02115, USA, 2AIDS Vaccine Program, SAIC-Frederick, Inc., NCI at Frederick, Frederick, MD 21702, USA, 3HIV Drug Resistance Program, NCI-Frederick, Frederick, MD 21702-1201, USA, 4Data Management Services, Inc., NCI-Frederick, Frederick, MD 2170, USA, 5Protein Chemistry Laboratory, SAIC Frederick, Inc., NCI at Frederick, Frederick, MD 2170, USA, 6MedImmune, Inc., 35 W. Watkins Mill Road, Gaithersburg, MD 20878, USA, 7Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, 6-155 Jackson Hall, 321 Church Street SE, Minneapolis, MN 55455, USA and 8Center for Interdisciplinary Research on Complex Systems, Northeastern University, 111 Dana Research Center, 110 Forsyth Street, Boston, MA 02115, USA

Received December 5, 2005; Revised and Accepted January 5, 2006

ABSTRACT

The Gag polyprotein of HIV-1 is essential for retroviral replication and packaging. The nucleocapsid (NC) protein is the primary region for the interaction of Gag with nucleic acids. In this study, we examine the interactions of Gag and its NC cleavage products (NCp15, NCp9 and NCp7) with nucleic acids using solution and single molecule experiments. The NC cleavage products bound DNA with comparable affinity and strongly destabilized the DNA duplex. In contrast, the binding constant of Gag to DNA was found to be ~10-fold higher than that of the NC proteins, and its destabilizing effect on dsDNA was negligible. These findings are consistent with the primary function of Gag as a nucleic acid binding and packaging protein and the primary function of the NC proteins as nucleic acid chaperones. Also, our results suggest that NCp7’s capability for fast sequence-nonspecific nucleic acid duplex destabilization, as well as its ability to facilitate nucleic acid strand annealing by inducing electrostatic attraction between strands, likely optimize the fully processed NC protein to facilitate complex nucleic acid secondary structure rearrangements. In contrast, Gag’s stronger DNA binding and aggregation capabilities likely make it an effective chaperone for processes that do not require significant duplex destabilization.

INTRODUCTION

The human immunodeficiency virus type 1 (HIV-1) contains two copies of single-stranded RNA (ssRNA), which are transcribed into double-stranded DNA (dsDNA) by a complex reverse transcription process. The linear dsDNA that results from this process is integrated into the DNA of the host cell, and from there it will be transcribed and translated using the gene expression machinery of the host cell. The main structural proteins of HIV-1 are expressed as a single polyprotein Gag (Figure 1A). Gag alone is sufficient for the formation of virus-like particles (VLP) in a mammalian cell (1). In fact, recombinant Gag, in the presence of nucleic acid and an additional small molecule cofactor, is sufficient for the assembly of normal VLPs in vitro (2–4) which, except for the absence of a surrounding membrane, are morphologically very similar to the immature viruses formed in vivo (3,5). However, this self-assembly does not occur in the absence of nucleic acid (2,4,6,7). Surprisingly, the self-assembly of structurally normal VLPs has been shown to be supported by the presence of virtually any RNA or DNA molecules (2,4,6–8). However, the HIV-1 Gag protein is also capable of preferentially packaging viral RNA from a large pool of cellular nucleic acid

*To whom correspondence should be addressed. Tel: 1 617 373 7323; Fax: 1 617 373 2943; Email: mark@neu.edu
Correspondence may also be addressed to Ioulia Rouzina. Tel: 1 612 624 7468; Fax: 1 612 624 5121; Email: rouzina@cb.s.umn.edu

© The Author 2006. Published by Oxford University Press. All rights reserved.

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use, please contact journals.permissions@oxfordjournals.org
Even slower cleavage events (400-fold slower than the primary cleavage) reside between the MA and CA domains (site 2a) and between the CA and NC domains (site 2b) of the C-terminal part of Gag. Two secondary cleavage sites, characterized by ~9-fold slower cleavage rates, reside between the MA and CA domains (site 2a) and between the p1 and p6 domains (site 2b) of the C-terminal part of Gag. Even slower cleavage events (~400-fold slower than the primary cleavage) occur at two additional sites, as the p2 and p1 peptides are cleaved from the C-terminal domains of CA and NC, respectively (sites 3a and 3b). In the fully mature, infectious virus, all proteins are present predominantly in their completely processed form (18). The three main products of the Gag processing by PR that contain the NC domain of Gag were termed according to their molecular weight and in the order of their appearance as follows: NCp15 (123 amino acids), NCp9 (71 amino acids) and NCp7 (55 amino acids) proteins (17,19).

All of the Gag cleavage events appear to be related to the changes in the structure and function of the virus during the course of its life cycle. Thus, the cleavage of NCp15, p2 and MA from Gag results in the formation of the cone-shaped capsid shell composed of CA, which is much smaller than the spherical membrane-bound MA shell (5,20,21). Within this cone-shaped mature core, there is still a dense phase of NC bound to the viral RNA dimer, a complex of the two RNA molecules within the virus (20,22,23). The RNA undergoes a process of further maturation, in which its secondary structure is altered. In a later stage of the retroviral life cycle, this RNA is reverse transcripted into proviral cDNA. The latter process involves several strand transfer steps. These events require major rearrangement of the nucleic acid secondary structure. While this restructuring results in a lower free energy of the nucleic acid molecules, and is therefore spontaneous, it appears to be extremely slow. NCp7 has been shown to significantly accelerate these reactions. The capability to facilitate the rearrangements required for these reactions is referred to as nucleic acid chaperone activity. However, it is not known to what extent the complete processing of Gag to its final form, NCp7 (Figure 1B) is required for the enhancement of specific rearrangements of nucleic acid secondary structure. For example, a recent study showed that while both NCp7 and Gag are effective at annealing tRNA\textsubscript{Lys}\textsuperscript{39} to the primer-binding site on the HIV-1 genome, Gag is more effective than NCp7 (24). However, only processed NC has been shown to be capable of facilitating minus-strand transfer (25,26).

In order to understand the effects of the processing of NC on its nucleic acid chaperone and binding activities, we will examine the interaction of DNA with Gag, NCp7, NCp9 and NCp15 using a single DNA molecule stretching technique with optical tweezers as well as two standard solution techniques: tryptophan fluorescence and fluorescence anisotropy (FA) binding assays. In this combined approach, we are able to extract both traditional and novel information on DNA–protein interactions. First, we are able to determine the binding constants to the nucleic acid of the stoichiometrically bound proteins by DNA stretching, obtaining results that agree with both fluorescence studies. In addition, we are able to obtain information on the effect of each protein on force-induced DNA melting, i.e. strand separation due to an external pulling force. In turn, this enables us to determine the effect of each protein on DNA duplex stability and on the kinetics of the strand separation and annealing. While the nucleic acid binding properties of a protein can be compared with previous solution binding data, as is done here, the information on the effect of NC on DNA melting is unique and important for determining the nucleic acid chaperone properties of various forms of NC. Despite our use of long polymeric dsDNA, it is likely that our results are relevant for the binding and duplex destabilizing capabilities of NC to RNA as well. To demonstrate the relevance of our DNA stretching results to RNA binding, we compare binding of both Gag and NCp7 with similar short DNA and RNA oligonucleotides. The major conclusion of this study is that the nucleic acid chaperone activity improves greatly as the Gag precursor is progressively processed to create NCp15, NCp9 and NCp7.

**MATERIALS AND METHODS**

**Protein preparation**

NCp7, NCp9 and NCp15 from HIV-1 were expressed and purified as described (25,27,28). Briefly, HPLC fractions containing the expressed NC sequences (identified by MALDI-mass spectrometry and N-terminal sequencing) were collected, pooled and lyophilized. NC proteins were resuspended in the presence of 1 equivalent of Zn\textsuperscript{2+} per finger,
and their concentration was calculated based on amino acid analysis measurements. Solutions were aliquoted into sterile polypropylene vials, relyophilized and stored at −70°C until used. The molecular masses of the proteins (as determined by MALDI-MS) were consistent with the following sequences: NCp7 = IQKGNFRNQRKTVCFCNCGKGEHIAKNCRPRKKGKWCGKEGHQMDCTERQAN; NCp9 = NCp7 + p1\textsuperscript{Gag} = NCp7 + FLGKIWPShKGGPS-GNP; NCp15 = NCp7 + p1\textsuperscript{Gag} + p6\textsuperscript{Gag} = NCp9 + LQSR-PeptaPeesFrFGeETtPPERIDKELYPLAS-LRSFSGDFPSSQ.

Gag protein (lacking myristic acid at its N-terminus and the p6 domain at its C-terminus, referred to previously as Gag–Δp6) was purified from bacteria as described (2), except that EDTA was eliminated from the purification procedure so that zinc ions would be retained in the zinc fingers of the NC domain. This change required modification of several steps in the procedure, as the protein with zinc bound bacterial nucleic acids considerably more tightly than that lacking zinc. Gag concentrations were measured from their absorbance in 6 M guanidine hydrochloride, pH 6.5.

**Fluorescence binding studies**

Poly(dT) was obtained in lyophilized form from Sigma (St Louis, MO). Poly(eA) was purchased from Pharmacia. LTR DNA oligonucleotides were purchased from Invitrogen with the following sequences: U5 sense, 5’- TAA CTA GAG ATC CCT CAG ACC CTT TTA GTC AGT GTG GAA AAT CTC TAG CAG T-3’; U5 antisense, 5’- ACT GGA AGG GCT AAT -ACT GCT AGA GAT -ACT GGA AGG GCT AAT TCA CTC CCA AAG ACA AGA TAT CCT TGA TCT GTG G-3’; U3 sense, 5’- ACT GGA AGG GCT AAT TCA CTC CCA AAG ACA AGA TAT CCT TGA TCT GTG G-3’; U3 antisense, 5’- CCA ATC AAG GAT GAT ATC TTG TCT TCT TTG GGA GTG AAT TAG CCC TTC CAG T-3’.

The oligonucleotides were aliquoted in five OD vials by the manufacturer and lyophilized. Once dissolved, the concentration of each solution was checked by UV absorbance. To anneal the oligonucleotides in solution, equimolar amounts of the complementary strands were mixed in Eppendorf tubes, heated to near boiling and allowed to cool slowly. A spectrophotometric temperature ramp showed >20% hyperchromism for the annealed duplex. Sodium monophosphate and biphosphate, magnesium and sodium chloride were purchased from Aldrich Chemical Inc. (Milwaukee, WI).

Fluorescence titrations were carried out with a SPEX Fluoromax-2 spectrofluorimeter. NC Trp emission was monitored at 350 nm (5 nm bandwidth) with 280 nm excitation (1 nm bandwidth) in a thermostated capped quartz cell (0.2 × 1.0 cm; Hellma Cells, Inc., Jamaica, NY); stepwise additions of nucleic acid to 0.5–1.0 μM NC solutions, in 10 mM sodium phosphate, pH 7.0 were performed. Readings were corrected for dilution and inner filter effects. NC protein–nucleic acid complex formation was monitored by reduction in the initial fluorescence of the protein [or ethenoA fluorescence enhancement, in the case of poly(eA)].

NC fractional saturation was inferred from the ratio of observed quenching to maximal quenching (at saturation), \( \Delta F_{lim} \), for each of the nucleic acid additions. Limiting quenching was calculated using the method of Kelly et al. (29).

The plateau’s intersection with the initial slope indicates the occluded binding site size (\( n \)) for a titration carried out under near-stoichiometric conditions such as that shown in Figure 2A. Polynucleotide binding cooperativity was analyzed using the model of McGhee and von Hippel (30).

Salt-back titrations were carried out by addition of a 5.0 M NaCl solution to a preformed NC–nucleic acid complex (Figure 2B). Double-logarithmic plots of log \( K \) versus log [Na\textsuperscript{+}] (Figure 2C) were used to derive binding constants under various ionic strength conditions (31). The binding free energy can thus be resolved into two components: an

![Figure 2](image-url)

**Figure 2.** (A) Normalized fluorescence versus concentration ratio of nucleic acid to protein. Poly(dT) was added to NCp15 (0.5 μM in 1 mM sodium phosphate, pH 7.0, and 25°C). The plateau’s intersection indicates the stoichiometry (\( n \), occluded binding site size in the case of a polynucleotide) for a titration carried out under near-stoichiometric conditions such as that shown. (B) Fluorescence recovery (\( F/F_0 \)) after the addition of 5 M NaCl aliquots to a pre-formed NCp15–poly(dT) complex. (C) Double-logarithmic plot for the association of NCp15 to poly(dT). The slope indicates the number of Na\textsuperscript{+} ions displaced; the y-intercept is log \( K_0 \) at 1 M Na\textsuperscript{+}, when only non-ionic interactions contribute to binding free energy.
electrostatic contribution calculated from the slope of the regression line in the double-logarithmic plot, and a hydrophobic (or non-electrostatic) contribution from the y-intercept (32). In the absence of net anion uptake or release during binding, dividing the slope by the number of ions bound (in the thermodynamic sense) per phosphate group in the double-stranded or single-stranded nucleic acid lattice, yields the number of electrostatic interactions incurred by the protein. The y-intercept represents the logarithm of the binding affinity at 1 M Na\(^+\) (log \(K^d\)), the standard state where electrostatic interactions are considered null. Low salt buffer was used in the initial titrations to maximize nucleic acid binding to NC proteins, permitting also a broader range of salt concentrations to be explored in the subsequent salt-back titration.

**Fluorescence anisotropy binding studies**

FA is based on the observation that when a fluorescent molecule is excited with plane-polarized light, it emits polarized fluorescent light into a fixed plane if the molecules are stationary between excitation and emission (33,34). The FA is determined by measuring the ratio of the emission intensities from the parallel and perpendicular planes. Because the FA of a molecule is proportional to its molecular volume, this ratio is used to study molecular interactions.

\(d(TG)_5\) labeled with fluorescein at the 3′ end was synthesized by Loeo Lee, SAIC-Frederick and purified by HPLC. \((UG)_5\) labeled with fluorescein at the 3′ end was from Dharmaco (Chicago, IL). To determine the dissociation constants of Gag and NCp7 to these oligonucleotides, 10 nM stock solutions of either \(d(TG)_5\) or \((UG)_5\) were prepared in HBS buffer (10 mM HEPES, 150 mM NaCl, pH 7.5) with 100 \(\mu\)M Tris(2-carboxyethyl)phosphine hydrochloride (TCEP), 5 mM \(\beta\)-mercaptoethanol and 1 \(\mu\)M ZnCl\(_2\) and aliquoted into 96-well Costar polypropylene plates (Corning, NY). NC and Gag proteins were serially diluted into both buffer and \(d(TG)_5\). An aliquot of 40 \(\mu\)l was removed and transferred into 384-well Costar polypropylene plates (Corning, NY). Samples were excited at 485 nm and the emission intensities at 535 nm from the parallel and perpendicular planes were measured using a Tecan Ultra plate reader (Durham, NC). Reactions were performed as three independent experiments. Curves were fit assuming a single binding site per oligonucleotide, using the relation (34):

\[
An(C) = \frac{(R \cdot An_{\text{bound}} - An_{\text{free}}) \cdot \left(\left(K_a + C + Lt\right) - \left(K_a + C + Lt\right)^2 - 4LtC^{-1/2}\right) / \left(2Lt\right) + An_{\text{free}}}{(R-1) \cdot \left(K_a + C + Lt\right) - \left(K_a + C + Lt\right)^2 - 4LtC^{-1/2} \left(2Lt\right) + 1}
\]

where \(An(C)\) is the measured anisotropy as a function of concentration, \(Lt\) is the oligonucleotide concentration, \(An_{\text{free}}\) is the anisotropy of unbound oligonucleotide, \(An_{\text{bound}}\) is the anisotropy of the completely bound oligonucleotide solution, \(R\) is the ratio of the fluorescence intensity of saturated bound oligonucleotide relative to free oligonucleotide, \(C\) is the protein concentration and \(K_d\) is the measured dissociation constant resulting from the fit. Because we do not actually know that there is only a single binding site per oligonucleotide, below we refer to the \(K_d\) measurements resulting from this fit as ‘apparent \(K_d\).’

### Force-induced melting

The optical tweezers instrument and data acquisition were described previously (35,36). Biotinylated DNA was obtained as previously described (37). The DNA was labeled at the 3′ ends only such that it would be free to rotate when stretched. All proteins were aliquoted in volumes of 10 \(\mu\)l and 10 or 25 \(\mu\)M protein concentration in specific NC-storage buffer (20 mM HEPES, 5 mM \(\beta\)-mercaptoethanol, 0.1 mM TCEP, pH 7.5), and stored at −80°C before use. NC proteins were reconstituted with two molar equivalents of zinc. The tethering buffer used in this study for all force-induced melting experiments (FIM buffer) contained 10 mM HEPES with 45 mM NaCl and 5 mM NaOH, pH 7.5. Lambda DNA, dNTPs (dATP, dTTP, dGTP), enzymes and buffers were purchased from Promega, the biotin-11-dCTP from Enzo, HEPES, NaCl and NaOH from Sigma, and the streptavidin-coated beads from Bang Laboratories. After capturing a single DNA molecule in the tethering buffer, the molecule was stretched to verify that the correct force-extension curve was obtained in the absence of bound protein. To measure the effect of the proteins on this transition, a buffer solution containing a fixed concentration of protein dissolved in FIM buffer was added to the experimental cell until the buffer surrounding the captured DNA molecule was completely exchanged.

### Results

Below we will present our results in two main sections. In the first, we use a single DNA molecule stretching technique and in the second we use a tryptophan fluorescence assay. These experiments are designed to elucidate the specific features of the interaction with DNA of the HIV-1 NC protein while in the context of Gag as well as in its three differently processed forms. The overall goal is to understand how the sequential proteolysis of HIV-1 NC affects its nucleic acid interaction properties, and to determine whether this can be correlated with the change of NC’s function in various stages of the viral replication cycle. In a third section, we compare the binding of NC and Gag with a very short oligoribonucleotide and a very short oligodeoxynucleotide to determine whether the 2′ OH (or the methyl group in thymidine) plays a significant role in the interactions of these proteins with nucleic acids.

### Single DNA molecule stretching studies of NC–DNA and Gag–DNA interactions

In this work we apply a single DNA molecule stretching technique to obtain information on DNA interactions with each of the HIV-1 NC and Gag proteins. Some information about the NCp7–DNA interaction obtained by this method was already discussed in our previous work (35,36). In this report we perform a much broader analysis of the HIV-1 NC–DNA interaction as a function of protein concentration, which allows us to determine the binding constant of each protein to DNA. In addition, we analyze the equilibrium and non-equilibrium features of our stretching curves, which allow us to determine both equilibrium and kinetic properties of protein–DNA interactions.
An equilibrium DNA stretching curve, \( F(x) \), is extremely sensitive to the presence of DNA binding proteins, such as NC, that bind stoichiometrically along the whole length of the DNA molecule. It, therefore, can be used to monitor this binding. To do this, we obtain the stretching curves of individual DNA molecules undergoing the force-induced melting transition. At large enough extensions the DNA under tension is a combination of melted and double helical regions, with the relative proportion of both forms varying with the molecular extension. Because the HIV-1 NC proteins are known to bind dsDNA and single-stranded DNA (ssDNA) with comparable affinities (Table 2), we will not try to distinguish between these modes of binding, but instead use our \( F(x) \) curves in order to obtain an estimate of the average composite binding constant to both forms of DNA. While this method cannot separately determine dsDNA and ssDNA binding, it allows us to quantify the interaction of NC with polymeric natural DNA. Such DNA more strongly resembles the actual substrate for NC binding in vivo relative to other artificial constructs such as oligonucleotides and homopolymers. This method was used successfully in a previous work to obtain binding constants for LINE-1 ORF1p and several mutants, and the results obtained were in agreement with bulk binding experiments (38).

In order to estimate the equilibrium binding constant of the protein to DNA, we can follow the variation of the force-extension curve \( F(x) \) with the protein concentration \( c \). The most distinct feature of the changes induced by the presence of the NC proteins in DNA stretching and relaxation behavior is the significant increase in the slope of the force-induced melting transition upon addition of the protein. This increase in transition width is likely due to NC’s sequence-specific binding and subsequent reduction in DNA melting cooperativity, as well as intercalation of NC into dsDNA upon stretching. All of these effects contribute to NC’s nucleic acid chaperone capabilities. Here, we use this growing transition width as a quantitative parameter to monitor the protein binding to ssDNA. We can define the transition width \( \Delta F \) as a force interval over which the DNA molecule switches from an entirely ds state to an entirely ss state. To estimate this transition width, we fit a tangent line to the transition midpoint (determined by linear regression analysis of several stretching curves), as shown in Figure 3. This midpoint corresponds to half the extension change between dsDNA only (solid line, closed circle), and ssDNA only (solid line, closed triangle) in the presence of protein. The slope of this line is the change in force from the beginning of the transition to the end of the transition, divided by the change in DNA length during the transition. Thus, the force transition width \( \Delta F \) is the difference in forces at which this tangent line intersects tangent lines to the dsDNA and ssDNA extension curves (Figure 3). The values obtained in this way are plotted as a function of protein concentration \( c \) for all four proteins studied (Figure 4). In all cases the transition width first increases linearly with protein concentration, and saturates at \( \Delta F_{sat} \). We see that the plots for all three forms of NC protein used in this study are similar within the experimental errors and very different from the plot for Gag. The slope of the linear region of this plot can be used to estimate the binding constant. Indeed, at sub-saturating protein concentrations, the fraction \( \Theta \) of DNA binding sites bound by protein is very small, \( \Theta \ll 1 \), and increases linearly with the concentration of the protein \( c \):

\[
\Theta = K \cdot c,
\]

where \( K \) is the apparent binding constant of protein to DNA. Note that the free protein concentration does not depend on the extent of binding because the protein is in great excess over the DNA in this experiment. In addition, NCp7 is known to exist as a monomer in solution even at high concentrations (39), and while Gag can form dimers in solution, at the concentrations used here the number of dimers should be negligible (S. A. Datta and A. Rein, unpublished data). Assuming that the change in the transition width, \( \Delta F = \Delta F - \Delta F_{0} \), is proportional to the fraction of sites bound by protein, \( \Theta \), and with the saturated value of this parameter, \( \Delta F_{sat} \), we can also write:

\[
\delta F(c) = \Theta \cdot \delta F_{sat}
\]

Here, \( \Delta F_{0} \) is the transition width in the absence of protein. Combining Equations 2 and 3 we can estimate the protein–DNA equilibrium binding constant as:

\[
K = \frac{1}{\Delta F_{sat}} \cdot \frac{\delta F}{c} = \frac{1}{\delta F_{sat}} \cdot \frac{d(\delta F(c))}{dc} \big|_{c \rightarrow 0}
\]

The fact that these stretching curves are not completely reversible does not preclude the measurement of equilibrium binding isotherms. The stretching curves used for these measurements did not depend on pulling rate (at the low pulling rates used here) and the measurements therefore are indistinguishable, within the reported experimental error, from equilibrium measurements (40). Note that irreversible thermal melting curves (which do not depend on heating rate) are routinely used to obtain equilibrium thermodynamic measurements (41).
of protein and data in the low concentration limit, as follows: NCp7:

\[ \Delta F_{sat} \] (pN) \[ d(\Delta F)/dC_0 \] (pN/nM) \[ K^c \] (10^5 M^{-1}) \[ \Delta G_{total}^f \] (kcal/mol bp) 

None 3.66 ± 0.16 — — 2.31 ± 0.05 1.36 ± 0.03
NCp7 48.4 ± 1.9 5.02 ± 0.97 1.22 ± 0.25 0.65 ± 0.02 0.38 ± 0.01
NCp9 40.9 ± 1.1 4.69 ± 0.58 1.40 ± 0.19 0.58 ± 0.02 0.34 ± 0.01
NCp15 39.1 ± 1.6 3.29 ± 0.22 1.04 ± 0.09 0.57 ± 0.02 0.33 ± 0.01
Gag 11.5 ± 2.0 7.54 ± 0.93 9.43 ± 2.6 1.86 ± 0.07 1.08 ± 0.04

*Measurements were performed in FIM buffer (see Materials and Methods).
\[ \Delta F_{sat} \] is the measured transition width at 30 nM HIV-1 NC or 3 nM Gag concentration (saturation) as shown in Figure 4.
\[ d(\Delta F)/dC_0 \] is the slope of the linear fit to experimental data in the low binding fraction limit (solid lines in Figure 4).
\[ K^c \] is calculated using Equation 4.
\[ \Delta G_{total}^f \] is the melting transition free energy reported in units of kcal/mol per base pair (bp), where \( k_B \) is the Boltzmann constant and \( T = 293 \) K is room temperature (1 \( k_B T = 581 \) cal/mol = 4 \( \times 10^{-21} \) J).

Figure 4. Dependence of the transition width \( \Delta F \) on the protein concentration. Data are for: (A) NCp7, (B) NCp9, (C) NCp15 and (D) Gag. Each data point for the NC proteins is given as the mean ± standard error for seven experiments, and with fewer experiments for Gag. Data are taken in FIM buffer (see Materials and Methods), and room temperature. The solid lines are linear fits to the data in the low concentration limit, as follows: NCp7: \( \Delta F = (5.02 ± 0.97)C_0 \); NCp9: \( \Delta F = (4.69 ± 0.58)C_0 \); NCp15: \( \Delta F = (3.29 ± 0.22)C_0 \); Gag: \( \Delta F = (7.54 ± 0.93)C_0 \). Here, \( \Delta F \) is the difference between \( \Delta F \) in the presence of protein and \( \Delta F \) in the absence of protein.

The parameters obtained when fitting the experimental data for all proteins studied (Figure 4) are shown in Table 1. We see that while the binding constants for the three processed forms of NC protein are indistinguishable, \( K \sim 1 \times 10^5 \) M^{-1}, the saturated value of the transition width increases slightly from NCp15 to NCp7. That is, \( \Delta F_{sat} \) = 35, 37 and 45 pN for NCp15, NCp9 and NCp7, respectively. On the other extreme, the Gag protein has higher binding affinity for ssDNA, \( K \sim 9 \times 10^8 \) M^{-1}, but the increase in the transition width is much lower, such that \( \Delta F_{sat} \sim 11 \) pN.

In addition to measuring the binding constant of each protein to DNA, our DNA stretching curves can be used to determine the dsDNA melting free energy, as discussed previously (42,43). In this report we will outline this approach and then use it for all four proteins at saturated binding in order to characterize the maximum effect of these proteins on DNA duplex stability. Briefly, the area of the mechanical cycle in the \( (F - x) \) plane delimited by the experimental stretching curve and the extrapolated stretching curve for ssDNA, with or without protein bound, gives the melting free energy in each case. Note that if the extension is per base pair, then the free energy is also per base pair, averaged over the whole length of polymeric DNA. The mechanical energy required to melt an average base pair, in turn, equals the free energy of the duplex relative to the single-stranded state of a base pair in the absence of applied force. The latter value is a function of solution conditions, including the concentration of DNA-binding protein. This method allows us to measure directly the effect of the protein on the stability of the duplex. Importantly, this approach does not rely on any model of the protein–DNA interaction or on the details of the DNA melting transition. We note that when the stretching curves are not reversible, the work done by stretching may not represent the equilibrium melting free energy. If that is the case, our results will slightly overestimate the melting free energy. However, this effect should be very small due to the relatively high cooperativity of DNA melting under the conditions studied here as well as the observed rate-independence (at
Representative DNA stretching and relaxation curves for all four proteins are shown in Figure 5. The stretching portion of the curve is used to obtain the DNA melting free energy per base pair in the presence of each protein, calculated at saturated protein binding. The results of this calculation are presented in Table 1. The addition of saturating amounts of any of the three NC proteins (20 nM) results in a low DNA helix stability of 0.35 ± 0.03 kcal/mol bp. This value is about 1 kcal/mol bp smaller than in the absence of protein at the 50 mM ionic strength used. The significant duplex destabilization induced by NC is the result of the preferential interaction between its zinc fingers and the single-stranded nucleic acids. In addition, NC produces some duplex stabilization due to increased electrostatic screening of DNA phosphate charges. The latter effect depends on the ionic strength of the buffer used and results in a salt dependence of the total NC-induced duplex destabilization. Thus, at higher salt, the net NC-induced DNA destabilization can become slightly larger than 1 kcal/mol bp. The effect of Gag on duplex stability is less significant, i.e. the free energy of base pair melting is only 0.3 kcal/mol bp lower than in the absence of NC, due to the overall increase in the melting force and small change in the transition width (Figure 5). One of the possible reasons why Gag does not significantly destabilize the DNA duplex could be that its NC domain is not capable of interacting with the ssDNA bases via its zinc fingers as efficiently as in the processed NC. However, this would be contradicted by the results of our direct measurement of NC and Gag binding to short ss oligonucleotides, discussed below, showing identical NC and Gag binding. The alternative explanation is that the binding of Gag molecules to dsDNA is always accompanied by strong interaction between the neighboring Gag molecules. The latter may produce a torsional constraint on the local unwinding of dsDNA, thus stabilizing it relative to the unwound ssDNA. This hypothesis is supported by our observation of the increasing force at the midpoint of the DNA force-induced melting transition upon Gag binding. Such a strong increase in the DNA melting force was previously observed when the stretched dsDNA was torsionally constrained (44). Thus, our results argue that the Gag polyprotein is inefficient as a duplex destabilizer, in contrast to all processed forms of NC. This finding suggests that Gag should be much less efficient at complex nucleic acid secondary structure rearrangements than the processed NC proteins.

Figure 5 also shows DNA relaxation curves for all four proteins, all of which show different amounts of hysteresis (i.e. disparity between stretch and relax curves). We have also studied the complete stretch–relax cycles of λ-DNA in the presence of many other retroviral NC proteins (M. Cruceanu, K. Stewart, R.J. Gorelick, K. Musier-Forsyth, I. Rouzina, M.C. Williams, manuscript in preparation), as well as several mutants of HIV-1 NCp7 (M. Cruceanu, M.-N. Vo, R.J. Gorelick, K. Musier-Forsyth, I. Rouzina, M.C. Williams, manuscript in preparation), and the ORF1 protein from mouse LINE-1 retrotransposon (38). Among all DNA binding proteins studied using force-induced melting of DNA, including the proteins presented herein, HIV-1 NCp7 protein produces the smallest hysteresis of the DNA stretch–relax cycle. As shown in Figure 5, the hysteresis induced by the proteins studied here, at saturation, increased in the following order: HIV-1 NCp7, NCp15, NCp9 and Gag. However, while the hysteresis contains rich information on protein–DNA interactions, its interpretation is complex, and has never been thoroughly understood.

Figure 5. Stretching (solid lines) and relaxing (dashed lines) curves for λ-DNA in the presence of the NC proteins or Gag: (A) NCp7; (B) NCp9; (C) NCp15; (D) Gag. Protein concentrations are indicated in the figure. Data were collected in FIM buffer (see Materials and Methods) at room temperature. The stretching and relaxing curves in the absence of protein are also shown (black solid line and filled diamonds, respectively).
been previously discussed. In Discussion, we present comments on the general information that can be obtained from studying the hysteresis of single DNA molecule stretching, followed by several conclusions regarding the proteins studied in this work. We will show that, although the information from studying the non-equilibrium features of the DNA stretch–relax cycle is not quantitative, it can serve as an important basis for understanding the kinetics of protein–nucleic acid interactions.

Fluorimetric titrations with HIV-1 NC proteins

The nucleic acid binding properties of HIV-1 NC in its three differently processed forms were investigated with the long single-stranded homopolymeric DNA and RNA molecules, poly(dT) and poly(εA), respectively (45). Both nucleic acid sequences were bound by the proteins nearly stoichiometrically under the conditions of the reverse titration experiment (see Materials and Methods). Binding to poly(dT) resulted in a large reduction (86%) in the fluorescence emission of the sole Trp residue (Trp37) in NCp7, as reported previously (45,46). The presence of the additional Trp residue (Trp61) in NCp9 and NCp15 resulted in a decrease in the limiting quenching at saturation for these NC sequences (Table 2), in agreement with earlier reports that Trp61 is not directly involved in the interaction of NCp9 with nucleic acids (46).

Table 2. Nucleic acid binding properties of HIV-1 NC proteins

<table>
<thead>
<tr>
<th>System</th>
<th>$n_{av}$</th>
<th>$\Delta F_{lim}$ (%)</th>
<th>$K$ (M$^{-1}$)</th>
<th>$\Delta \log K$</th>
<th>$\log K^0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCp7 + poly(dT)</td>
<td>6.4</td>
<td>-86</td>
<td>9.3 x 10$^7$</td>
<td>2.54</td>
<td>2.89</td>
</tr>
<tr>
<td>NCp9 + poly(dT)</td>
<td>9.5</td>
<td>-83</td>
<td>1.1 x 10$^8$</td>
<td>-2.61</td>
<td>2.83</td>
</tr>
<tr>
<td>NCp15 + poly(dT)</td>
<td>10.3</td>
<td>-70</td>
<td>1.4 x 10$^9$</td>
<td>-3.35</td>
<td>2.45</td>
</tr>
<tr>
<td>NCp7 + poly(εA)</td>
<td>6.6</td>
<td>+135</td>
<td>1.4 x 10$^9$</td>
<td>2.02</td>
<td>5.54</td>
</tr>
<tr>
<td>NCp9 + poly(εA)</td>
<td>12.2</td>
<td>+210</td>
<td>3.8 x 10$^9$</td>
<td>-2.06</td>
<td>11.1</td>
</tr>
<tr>
<td>NCp15 + poly(εA)</td>
<td>11.8</td>
<td>+195</td>
<td>1.7 x 10$^9$</td>
<td>-2.04</td>
<td>5.40</td>
</tr>
<tr>
<td>NCp7 + 3’LTR-dsDNA</td>
<td>2.7 bp</td>
<td>-85</td>
<td>8.3 x 10$^4$</td>
<td>-3.26</td>
<td>2.40</td>
</tr>
<tr>
<td>NCp9 + 3’LTR-dsDNA</td>
<td>4.3 bp</td>
<td>-61</td>
<td>3.8 x 10$^3$</td>
<td>-3.65</td>
<td>2.28</td>
</tr>
<tr>
<td>NCp15 + 5’LTR-dsDNA</td>
<td>2.9 bp</td>
<td>-86</td>
<td>3.4 x 10$^8$</td>
<td>-3.21</td>
<td>2.49</td>
</tr>
<tr>
<td>NCp7 + 5’LTR-dsDNA</td>
<td>4.4 bp</td>
<td>-59</td>
<td>8.9 x 10$^9$</td>
<td>-3.96</td>
<td>0.23</td>
</tr>
<tr>
<td>NCp9 + 5’LTR-dsDNA</td>
<td>4.0 bp</td>
<td>-92</td>
<td>4.3 x 10$^6$</td>
<td>-2.31</td>
<td>3.03</td>
</tr>
<tr>
<td>NCp15 + 3’LTR/U5</td>
<td>8.3</td>
<td>-81</td>
<td>6.2 x 10$^5$</td>
<td>-2.50</td>
<td>3.46</td>
</tr>
<tr>
<td>NCp9 + 3’LTR/U5</td>
<td>4.8</td>
<td>-93</td>
<td>2.2 x 10$^6$</td>
<td>-2.96</td>
<td>2.77</td>
</tr>
<tr>
<td>NCp15 + 5’LTR/U3</td>
<td>9.5</td>
<td>-86</td>
<td>3.9 x 10$^8$</td>
<td>-2.85</td>
<td>3.23</td>
</tr>
</tbody>
</table>

$^{a}$Thermodynamic results from fluorimetric binding isotherms of NCp7, NCp9 and NCp15 with various nucleic acid oligomers were obtained by two methods: (i) addition of nucleic acid aliquots to protein (Trp quenching) and (ii) addition of protein aliquots to poly(εA), both in a solution of 10 mM sodium phosphate buffer, pH 7.0, at 25°C. Average errors were as follows: ±1% for $n_{av}$, ±1.6% for $\Delta F_{lim}$, ±28% for $K$; ±0.07 for $\log K$/log[Na$^+$]; and ±0.19 for $\log K^0$. $K^0$ is the association constant extrapolated to 1 M salt.

$^{b}$These results were measured only at higher salt than that reported for the other measurements in the table. To allow for comparison of results, they have been extrapolated to 10 mM sodium phosphate buffer conditions.

Fluorescence anisotropy binding studies with Gag and NCp7

Because the DNA stretching studies were done with DNA only, we wanted to test whether or not our conclusions concerning Gag and Gag cleavage products containing NC were applicable to Gag and NC binding to RNA. The interactions of NC with either RNA or DNA are both potentially biologically important in the context of NC-induced facilitation of reverse transcription. To this end, we performed FA binding studies of Gag and NCp7 binding to 10 nt DNA and RNA oligonucleotides, d(TG)$_5$ and r(UG)$_5$. The results of these measurements are presented in Table 3. The measured values of the binding constant for NC are in agreement with its value obtained in our stretching and tryptophan fluorescence quenching studies (see Tables 1 and 2). It seems likely that this agreement is in fact a result of mutual compensation of two effects. The FA binding studies were performed at higher (150 mM NaCl) salt, as compared with 50 mM NaCl in DNA stretching and 10 mM NaCl in tryptophan quenching experiments. Higher salt decreases the binding affinity of NC to nucleic acids. However, this weakening is compensated by the fact that TG and UG oligonucleotide sequences bind NC more strongly than random nucleic acid sequences (48,49). Perhaps the most important observation is that both NC and Gag bind similarly to RNA and DNA

Table 3. Binding constants of HIV-1 NCp7 and Gag proteins to 10 nt ssRNA and ssDNA oligomers measured by FA

<table>
<thead>
<tr>
<th>Oligonucleotide</th>
<th>$K$ (M$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(TG)$_5$</td>
<td>(7.2 ± 1.5) x 10$^7$</td>
</tr>
<tr>
<td>(UG)$_5$</td>
<td>(3.4 ± 0.9) x 10$^7$</td>
</tr>
</tbody>
</table>

Binding measurements were performed as described in Materials and Methods.
oligomers. This result is consistent with previous observations of comparable NC binding to DNA and RNA molecules (49,50).

Another very important result is that the binding constants of NC and Gag to these short oligonucleotides are very similar, see Table 3. This result is in contrast to 10-fold stronger binding of Gag to long dsDNA, observed in our stretching study (Table 1). One possible explanation for this discrepancy is that Gag binds the longer dsDNA by some additional region besides NC, most likely by the matrix domain of Gag (MA). The alternative explanation is that the longer nucleic acid molecule is capable of binding several Gag molecules simultaneously, while the short oligonucleotides used (10 nt) can likely accommodate only one Gag molecule. In this case, even if the individual Gag–DNA and NC–DNA binding constants are equally strong, Gag–Gag interactions can contribute to the stronger apparent Gag binding observed in DNA stretching experiments. In Discussion below we explore this hypothesis and estimate the DNA binding cooperativity parameter of Gag, and the corresponding free energy of Gag–Gag interaction.

DISCUSSION

In this work we used a single DNA molecule stretching technique, as well as a tryptophan fluorescence assay, in order to understand the specific features of the interaction with DNA of the HIV-1 NC protein while in the context of Gag as well as in its three differently processed forms. Our major goal is to see how the sequential proteolysis of HIV-1 NC affects its nucleic acid interaction properties, and whether or not this can be correlated with the change of NC’s function in various stages of the viral life cycle. While the NC protein within the context of Gag appears to have mostly a nucleic acid binding and packaging function, supporting Gag self-assembly into VLP, the processed forms of NC appear to act more as nucleic acid chaperones. The latter finding means that NC can be capable of greatly facilitating nucleic acid restructuring processes that lead to the more stable annealed state of these molecules. While Gag facilitates tRNA_Lys annealing to the HIV-1 primer binding site (51–53), it has not been shown to facilitate more complex nucleic acid rearrangements, such as those involved in strand transfer. Until recently, it was not clear which properties of NC are responsible for its chaperone activity. However, recent studies of the kinetics of annealing (54), strand transfer (25,55–58) and recombination (59–62) have shown that the chaperone activity of NC has two major components: helix destabilization and strand aggregation (63).

HIV-1 NC is able to weakly destabilize all nucleic acid base pairs, allowing it to melt unstable nucleic acid intermediates, but preventing melting of the final, stably annealed structures. Thus, melting by NC of the short helical regions bordered by mismatches or bulges provides the single-stranded nucleation sites for the new helices. In addition, NC-induced duplex destabilization accelerates various processes of strand exchange (63). On the molecular level, the capability of HIV-1 NC to destabilize nucleic acid duplexes relies on its preferential binding to single-stranded nucleic acids via the stacking of the aromatic residues of its zinc fingers with unpaired bases (49,64–66). In our force-induced melting studies, this duplex destabilizing activity of NC is observed directly by measuring the changes in DNA base pair stability upon addition of NC. In the present work this duplex destabilizing activity of all three NC proteins studied appeared to be quite similar, with each base pair destabilized by processed NC proteins by ~1 kcal/mol bp. Since the average stability of a base pair in the absence of NC was ~1.4 kcal/mol in 50 mM NaCl, this is a significant but incomplete destabilization induced by NC, which is consistent with the chaperone function of NC. Similar destabilization produced by all three proteins seems reasonable, taking into account the fact that they all share the same DNA binding domains.

In contrast, according to our measurements, the DNA binding constant of Gag is about 10-fold higher than that of the other three NC proteins. If the NC domain constitutes the main binding domain of Gag to DNA, the enhancement of the protein–DNA association constant comes most likely from an additional Gag–Gag interaction upon binding of the protein to DNA (i.e. cooperative binding) (30), although we cannot rule out additional interactions from other parts of Gag. In the limit of low protein binding, \( \Theta < 1 \), the fraction of protein bound \( \Theta \), is proportional to the protein concentration \( c \), as in Equation 2:

\[
\Theta = K^{\text{eff}} \cdot c,
\]

and the effective binding constant includes the cooperativity parameter, \( \omega \), defined below:

\[
K^{\text{eff}} = K \cdot \omega
\]

\[
\omega = \exp \left[ \frac{(m/2) \cdot \delta G_{\text{Gag-Gag}}}{k_B T} \right]
\]

Here \( \delta G_{\text{Gag-Gag}} \) is the free energy of Gag–Gag interaction per protein molecule, \( m \) is the number of the nearest-neighbor Gag molecules when bound to DNA, the factor 1/2 takes into account that each interaction is shared by the two Gag molecules, and \( k_B T \) is the thermal energy per statistical degree of freedom, which at room temperature is \( \sim 0.6 \text{ kcal/mol} \). The fact that the effective binding constant of Gag is \( \sim 10 \)-fold higher than for either NC protein implies that \( \omega \sim 10 \). This estimate, of course, assumes that the intrinsic binding constant of Gag to DNA, \( K \), is the same as the binding constant of NC to DNA. By making this assumption and using Equation 7, we can estimate that

\[
m/2 \cdot \delta G_{\text{Gag-Gag}} = k_B T \ln \omega \approx 2.3 k_B T = 1.4 \text{ kcal/mol}.
\]

The pair-wise attraction free energy \( \delta G_{\text{Gag-Gag}} \) can be estimated from the net free energy of cooperative interactions per Gag molecule given by Equation 8, for the particular number of nearest neighbors, \( m \). The \( m \) range for Gag is likely from 3 to 4, with individual interacting subunits of capsid forming dimers of trivalent capsid molecules (67–69), each of which has 2–3 neighbors in addition to the dimer partner. In our experiments the number of interacting Gag neighbors bound to the single stretched DNA molecules could be as low as 2, given the fact that the DNA is stretched. We can therefore estimate the Gag–Gag interaction free energy to range between \( \delta G_{\text{Gag-Gag}} = 0.7 \) and 1.4 kcal/mol.
These results suggest that the individual CA–CA interactions of HIV-1 Gag are much weaker than any typical functionally important protein–protein interaction. Unfortunately, other direct measurements of such weak interactions are not yet available for HIV-1 Gag. However, the cooperative Gag–DNA binding constant that we infer from our results, $K_{Gag} = 10^9 \text{M}^{-1}$, is in agreement with the typical values of the reciprocal critical concentration of capsid molecules required for the capsid self-assembly $K_{sp} = 10^8–10^9 \text{M}^{-1}$ for several other viruses studied (70). Thus, cooperative binding of Gag to DNA is consistent with its enhanced capability to bind to and package nucleic acids. However, multiple Gag–Gag interactions are expected to lead to a slowing of this protein’s motion on nucleic acid. This, in turn, should result in the interference of Gag with the fast nucleic acid duplex melting/annealing observed in the presence of NCp7, as reflected in the significant hysteresis observed in DNA stretching curves in the presence of Gag (Figure 5). Therefore, because of this slow unbinding (relative to NCp7) as well as the inability of Gag to destabilize the DNA duplex, the interactions of Gag with nucleic acids are not likely to be optimized for complex rearrangements of nucleic acid secondary structure such as those required for strand transfer.

The second HIV-1 NC activity that contributes to its chaperone function is its ability to promote annealing of nucleic acid strands via their aggregation. This activity is especially important for facilitating the bimolecular nucleation step of strand annealing. The NC-induced rate enhancement of such reactions is especially significant in solutions of low ionic strength, where it can reach up to $10^{-5}$-fold (54,71–73). This is because the NC protein facilitates the new duplex nucleation by screening the negative phosphate charges of the nucleic acid strands, just as is observed in high salt. However, the facilitating effect of HIV-1 NC on duplex nucleation goes far beyond this high salt effect (72). The additional rate enhancement comes from NC-induced nucleic acid–nucleic acid attraction (74,75).

This electrostatic attraction is a phenomenon similar to the nucleic acid aggregation induced by multivalent cations (76–78). The attraction induces nonspecific aggregation of all nucleic acids, increasing the local concentration of strands, thus facilitating the diffusional search of the complementary regions for each other. The most important condition for this mechanism of nucleic acid annealing rate enhancement is that the nucleic acid molecules should remain highly mobile relative to each other within such an aggregate. These conditions seem to be fulfilled in the case of HIV-1 NCp7. Being a highly charged cationic molecule, NCp7 most likely produces the nucleic acid aggregation in a manner similar to that of other simple multivalent cations (76–82) via its electrostatic interaction with nucleic acids. This highly mobile and nonspecific nucleic acid aggregation induced by NCp7 can be contrasted with the much stronger and much less mobile ‘sticking’ of nucleic acid molecules produced by the cooperative association of the hydrophobic parts of the nucleic acid binding proteins, such as Gag. This concept is reflected in the stretching and relaxation curves shown in Figure 5, which in the cases of NCp7, NCp15, NCp9 and Gag proteins shows progressively larger hysteresis.

There are several factors that may cause the hysteresis to increase in the presence of typical DNA binding proteins. First of all, a protein that binds preferentially to ssDNA, as NC does, may be slow to unbind, and therefore might interfere with strand annealing (40,83). Moreover, in order for the relaxation curve to coincide with the stretching curve, the fraction of the base pairs annealed should be in true equilibrium. This means that on the time scale of a single step, i.e. ~1 s, in which the force is being averaged in our experiment, multiple events of melting and reannealing of the base pairs should occur, such that the fraction of the annealing and melting base pairs averages to the thermodynamic equilibrium quantity. This, in turn, implies that many protein molecules bound to the ~1000 bp long piece of DNA that is melted in each step should be able to unbind ssDNA and rebind dsDNA many times on the time scale of 1 s. This is a rather stringent requirement on the ability of the protein to adjust to the state of DNA. Based on the experimental data showing very small hysteresis in the presence of NCp7, we argue here that NCp7 is indeed such a very ‘fast’ binding and unbinding protein.

This hypothesis is in good agreement with the very high mobility of NCp7 when bound to various nucleic acid molecules, as observed in NMR experiments. Indeed, NMR studies have shown that the majority of NCp7 molecules, when bound to nucleic acids, are able to rapidly switch between numerous possible conformations, which are averaged out on the millisecond time scale of an NMR experiment (84–86). Also, recent fluorescent energy transfer studies of the TAR DNA molecule have shown that the NCp7 (12–55) fragment is able to facilitate the rate of the terminal DNA duplex opening and closing on the $10^{-4}$–$10^{-3}$ s time scale by several-fold (87–89). This implies that NCp7, lacking the basic N-terminal tail, is capable of changing its binding mode between the ssDNA- and dsDNA-bound modes even faster than that. This is not impossible, since in order to adjust to a new state of DNA, the protein does not have to completely unbind and rebind the nucleic acid. Instead, it may simply change the conformation of its aromatic residues between DNA bases, i.e. stacked and unstacked, in addition to undergoing other small conformational changes. Such movements can easily occur on the microsecond time scale.

It is likely that protein–protein interactions resulting in cooperative protein binding to DNA are responsible for the observed hysteresis in the presence of Gag. The Gag–Gag interactions estimated from the high value of the Gag–DNA binding constant above are consistent with the large hysteresis in the presence of high concentrations of Gag (1–10 nM). Similar features are present for NCp9 and NCp15 proteins, which induce hysteresis that is much stronger than in the case of NCp7. This finding correlates with the results of Khan and Giedroc (72), who reported that, in contrast to the completely processed NCp7, NCp9 shows moderate DNA binding cooperativity. In addition, these authors found that while NCp7 binds and unbinds DNA very fast on the time scale of a Trp fluorescence quenching measurement, the NCp9 protein was shown to undergo some slow changes in its binding to DNA (72). Taken together, these findings suggest that the distal domains in the partially processed NC protein and, particularly, the unstructured, flexible and highly hydrophobic p1 sequence present in NCp9, may participate in protein–protein contacts.

Thus, the hysteresis and possible DNA–protein binding cooperativity that are observed most strongly for Gag
NCp9 are signatures of a ‘slow’ DNA binding protein. Despite this slow binding characteristic, the capability of Gag to aggregate nucleic acids, most likely makes it a reasonably good chaperone for processes that do not require significant duplex destabilization, such as rRNA annealing (51–53). However, processes that require significant duplex destabilization, such as strand transfer, will likely be optimally facilitated by the final processed NCp7, which exhibits strong duplex destabilization, and which facilitates rapid reannealing of DNA that has been melted by force.

The direct results of the studies presented here can be summarized as follows. Both single DNA molecule stretching experiments and bulk Trp fluorescence assays showed that the three Gag cleavage products NCp15, NCp9 and NCp7 bound DNA with comparable affinity. The FA assays showed that NC and Gag bind with similar affinity to short DNA and RNA oligonucleotides. In contrast, the DNA stretching experiments showed that Gag bound single polymeric DNA molecules with an apparent affinity 10 times higher than that of any of the Gag cleavage products. Based on the change in single DNA molecule melting force observed in the presence of these proteins, we also found that Gag was unable to destabilize the DNA helix, while the Gag cleavage products destabilized DNA significantly. Finally, upon relaxation of the DNA after force-induced melting, we found that the amount of hysteresis was greatest for Gag, followed by NCp9 and NCp15. The final NC Gag cleavage product, NCp7, exhibited the least amount of hysteresis, indicating that this protein is most efficient at reannealing the regions of a single DNA molecule that have been melted by force. Thus, the final Gag cleavage product is expected to be most efficient nucleic acid chaperone for processes that require both nucleic acid destabilization and nucleic acid annealing. Both of these characteristics are required for minus-strand transfer, in which the very stable nucleic acid secondary structures, TAR RNA and cTAR DNA, must be annealed to each other. In contrast, Gag is more efficient at binding and packaging viral RNA. It can also perform some nucleic acid annealing functions, such as rRNA annealing, a process that does not require significant duplex destabilization, but involves nucleic acid aggregation by protein.

ACKNOWLEDGEMENTS
This work was funded in part by the National Cancer Institute under contract No. NO1-CO-12400 with SAIC-Frederick, Inc., by the National Science Foundation (MCB-0238190), the National Institutes of Health (GM-072462), by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research, and the Research Corporation. M.A.U. was partially funded by a fellowship for Perfeccionamiento y Movilidad del Personal Investigador from the Departamento de Educación, Universidades e Investigación del Gobierno Vasco (Spain). We thank Prof. Karin Musier-Forsyth for helpful discussions and Prof. Richard L. Karpel for a critical reading of the manuscript. Funding to pay the Open Access publication charges for this article was provided by NIH GM-072462.

Conflict of interest statement. None declared.

REFERENCES


