Published online January 24, 2006

472–484 *Nucleic Acids Research, 2006, Vol. 34, No. 2*
doi:10.1093/nar/gkj442

Complex interactions of HIV-1 nucleocapsid protein with oligonucleotides

Robert J. Fisher*, Matthew J. Fivash[1], Andrew G. Stephen, Nathan A. Hagan[2], Shilpa R. Shenoy[3], Maxine V. Medaglia, Lindsey R. Smith, Karen M. Worthy, John T. Simpson, Robert Shoemaker[4], Karen L. McNitt[1], Donald G. Johnson[5], Catherine V. Hixson[5], Robert J. Gorelick[5], Daniele Fabris[2], Louis E. Henderson[5] and Alan Rein[6]

Protein Chemistry Laboratory, SAIC-Frederick, Inc., NCI Frederick, Frederick, MD 21702, USA, [1]Data Management Services, Inc., NCI Frederick, Frederick, MD 21702, USA, [2]Department of Chemistry and Biochemistry, University of Maryland Baltimore County, Baltimore, MD 21250, USA, [3]Molecular Targets Development Program, Basic Research Program, SAIC-Frederick, Inc., NCI Frederick, Frederick, MD 21702, USA, [4]Screening Technologies Branch, Developmental Therapeutics Program, Division of Cancer Treatment and Diagnosis, NCI Frederick, Frederick, MD 21702, USA, [5]AIDS Vaccine Program, SAIC-Frederick, Inc., NCI Frederick, Frederick, MD 21702, USA and [6]HIV Drug Resistance Program, National Cancer Institute at Frederick, Frederick, MD 21702, USA

Received October 7, 2005; Revised and Accepted December 23, 2005

ABSTRACT

The HIV-1 nucleocapsid (NC) protein is a small, basic protein containing two retroviral zinc fingers. It is a highly active nucleic acid chaperone; because of this activity, it plays a crucial role in virus replication as a cofactor during reverse transcription, and is probably important in other steps of the replication cycle as well. We previously reported that NC binds with high-affinity to the repeating sequence $d(TG)_n$. We have now analyzed the interaction between NC and $d(TG)_4$ in considerable detail, using surface plasmon resonance (SPR), tryptophan fluorescence quenching (TFQ), fluorescence anisotropy (FA), isothermal titration calorimetry (ITC) and electrospray ionization Fourier transform mass spectrometry (ESI-FTMS). Our results show that the interactions between these two molecules are surprisngly complex: while the K_d for binding of a single $d(TG)_4$ molecule to NC is only ~5 nM in 150 mM NaCl, a single NC molecule is capable of interacting with more than one $d(TG)_4$ molecule, and conversely, more than one NC molecule can bind to a single $d(TG)_4$ molecule. The strengths of these additional binding reactions are quantitated. The implications of this multivalency for the functions of NC in virus replication are discussed.

INTRODUCTION

The nucleocapsid (NC) protein of human immunodeficiency virus type 1 (HIV-1) is a very small, basic protein containing two retroviral zinc fingers. These zinc fingers exhibit the following conserved retroviral-type arrangement of zinc-coordinating residues: $C\text{-}X_2\text{-}C\text{-}X_4\text{-}H\text{-}X_4\text{-}C$. NC is initially synthesized as part of the Gag polyprotein (1). As a domain of Gag, it participates in selection of the genomic RNA for encapsidation into the nascent virus particle (2) and its interactions with RNA are somehow critical in the structure of the particle (3,4). It also acts as a nucleic acid chaperone, unwinding a cellular tRNA molecule and catalyzing the hybridization of a portion of the tRNA to the viral genome, where it serves as a primer for reverse transcription (5,6). After the virus particle is released from the virus-producing cell, the Gag polyprotein is cleaved by the viral protease, resulting in the release of NC from the polyprotein (1). NC is complexed with genomic RNA in the mature, infectious virion; the nucleic acid chaperone activity of the protein causes a stabilization of the dimeric linkage between the genomic RNA molecules within the particle (7,8). When the particle infects a new host cell, this chaperone activity plays a crucial role in reverse transcription (9,10) and possibly in integration as well (11,12).

Thus, NC interacts with nucleic acids in remarkably diverse ways during the retroviral replication cycle. Because its functions are so critical for retrovirus replication, it is important to understand its interactions with nucleic acids in as much detail as possible. As one approach to this goal, we have studied the

*To whom correspondence should be addressed. Tel: +1 301 846 5154; Fax: +1 301 846 7392; Email: fisher@ncifcrf.gov

binding of this protein to very short, simple oligonucleotides. We and others previously reported that in 0.15 M NaCl, NC binds with significantly higher affinity to the repeating sequence, $d(TG)_n$, than to other tested sequences (13,14). We have now extended this analysis in several ways. The binding of NC to $d(TG)_4$ and related oligonucleotides was characterized by tryptophan fluorescence quenching (TFQ), fluorescence anisotropy (FA), isothermal titration calorimetry (ITC) and electrospray ionization Fourier transform mass spectrometry (ESI-FTMS), as well as in further experiments using surface plasmon resonance (SPR). All the results were subjected to a quantitative analysis. The results of these studies show that both NC and $d(TG)_4$ are multivalent, despite their small sizes. We have quantitated the binding affinities for formation of higher-order, as well as 1:1, complexes between these two molecules. We propose that these properties of the system make important contributions to the nucleic acid chaperone activity of NC, and thus to HIV-1 replication.

MATERIALS AND METHODS

Mutant and wild-type NC proteins

All NC proteins are the mature, 55 amino acid form of NC from pNL4-3 (GenBank accession no. AF324493). Procedures for the preparation of wt NC, NC1:1, NC2:1, NC2:2, NCSSHS and NCF16A mutant recombinant proteins have been published previously (11,15). The NC N-term mutant ('N-term ala'), in which all positive charges are removed from the N-terminus before the first zinc finger, was made by introducing the following nucleotide changes in the NC gene before cloning into the pET32a vector: A23G/A24C, A35G/G36C, A44G/G45C, A47G/A48G and A56G/A57C, where nucleotide 1 is the 5′ end of the NC gene. These changes resulted in the substitution of alanine residues for the lysines at positions 3, 11 and 14, and the arginines at positions 7 and 10. Mutagenesis was performed with the Quick Change Site Directed Mutagenesis Kit (Stratagene). The recombinant N-term ala protein was expressed and purified essentially as described previously (11). In some experiments, the zinc fingers of wild-type NC were disrupted by treating the protein with 5 mM EDTA in the presence of 100 μM TCEP. Experiments with SSHS NC were also performed in the presence of 5 mM EDTA. All protein concentrations were determined by quantitative amino acid analysis on a Beckman System 6300 amino acid analyzer (Beckman Coulter, Inc., Fullerton, CA) and/or spectrophotometrically. The ε^{280} was taken as 5.7×10^3 cm^{-1} M^{-1} for wild-type NC and was adjusted as appropriate for mutant NCs. Virtually all results presented have been reproduced with multiple batches of NC protein.

Oligonucleotides

All oligonucleotides were oligodeoxynucleotides and were synthesized by Leo Lee (SAIC-Frederick). They were purified after synthesis by high-performance liquid chromatography (HPLC) and their concentrations were determined from their absorbance at 260 nm. ε^{260} values were: for $(TG)_4$, 75 400; for $(A)_{30}$, 363 400; for $(TG)_{10}$, 188 200; for $(A)_{10}(TG)_5(A)_{10}$, 363 400; and for $(TG)_5(A)_{10}(TG)_5$, 308 800. For attachment to the SPR surface, the oligonucleotides were synthesized with biotin at the 3′ end using biotin TED CPG (Glen Research, Sterling, VA). Free biotin was removed using a G25 Microspin column (Amersham Biosciences, Piscataway, NJ) according to the manufacturer's instructions. Oligonucleotides used in the FA studies were synthesized with a fluorescein CPG (Glen Research) at the 3′ end. These fluoresceinated oligonucleotides were then purified by HPLC using a Waters Xterra MS C_{18} column (4.6 × 50 mm, particle size 2.5 μm; Milford, MA). The oligonucleotides were eluted over 20 min with a linear gradient of 2–98% Buffer B in Buffer A. (Buffer A consisted of 95% 0.1 M triethylammonium acetate/5% acetonitrile and Buffer B consisted of 70% 0.1 M triethylammonium acetate/30% acetonitrile.) The molecular weights of the eluted oligonucleotides were verified by MALDI-MS.

SPR analysis

Binding experiments were performed on BIAcore 3000 and BIAcore S51 instruments (BIAcore Inc., Piscataway NJ). Biotinylated oligonucleotides were immobilized on streptavidin-coated sensor chips. These chips were prepared in the laboratory by amine coupling using the BIAcore amine coupling kit (#BR1000-50) according to the manufacturer's instructions. Approximately 3500–3800 RU's of streptavidin were attached to BIAcore 3000 chips. Chips for the BIAcore S51 instrument were coated with ∼500 RU's of streptavidin. Oligonucleotides were reconstituted in buffer consisting of 10 mM Tris (pH 7.5), 300 mM NaCl and 1 mM EDTA. The amount of oligonucleotide immobilized on each surface is listed in the relevant figure legends. Proteins were serially diluted in a running buffer, consisting of 1× HBS [10 mM HEPES, 150 mM NaCl (pH 7.5)], 5 mM β-mercaptoethanol, 100 μM TCEP, 0.005% Tween-20 and 1 μM $ZnCl_2$. The concentrations described in each sensorgram were injected at 25°C. After each injection of analyte during a BIAcore experiment, bound analyte is allowed to dissociate freely for a specified period of time. Disruption of any complex that remained bound after dissociation was achieved using an injection of 0.1% SDS. Samples with different concentrations of protein were injected in non-sequential order. The BIAcore 3000 data were X- and Y-transformed using BIAevaluation 3.01 software and non-random noise was removed by subtracting a buffer injection from each curve. Small refractive index mismatches were removed using MATLAB software. The curves obtained from all four flow cells and all NC concentrations were globally fit using the complex model shown in Figure 3, as described in the Supplementary Data.

The data with very low densities of oligonucleotide were gathered on the BIAcore S51 using a 1 min injection of NC protein at 100 μl/min, followed by a 1 min dissociation and 30 s regeneration with 0.1% SDS. The NC injections were done in duplicate, and non-random noise was removed by subtracting a reference surface from each surface during a simultaneous measurement. The corrected data were globally fit to a 1:1 binding model as described in the Supplementary Data.

Tryptophan fluorescence assays

A Perkin Elmer (Wellesley, MA) luminescence spectrophotometer (LS50B) was used to measure fluorescence quenching. The instrument was set to an excitation maximum of 290 nm

and an emission maximum of 350 nm. The excitation and emission slit widths were 2.5 and 10 nm, respectively. The filter cut-off was set at 350 nm and the integration time for sample readings at 10 s. The buffer was comprised of 1× HBS, 1 mM $ZnCl_2$, 100 μM TCEP and 5 mM β-mercaptoethanol. A total of 3 ml buffer was added to two paired Hellma quartz cuvettes. NC was added to one cuvette for a final concentration of 400 nM, while an equivalent volume of buffer was added to the other cuvette as a control. Aliquots of the octanucleotide $d(TG)_4$ (referred to as $(TG)_4$ in the present work) were added to both cuvettes, and fluorescence readings were taken in triplicate. The average fluorescence readings of the control solution were subtracted from that of the NC solution to correct for dilution effects.

FA

Equilibrium binding isotherms for HIV-1 NC binding to $(TG)_4$ were monitored by FA. Serial dilutions of wild-type or mutant NC were mixed in 96-well Costar polypropylene plates (Corning, NY) with 10 nM $(TG)_4$ tagged at the 3′ end with fluorescein in 1× HBS, 5 mM β-mercaptoethanol, 100 μM TCEP, 1 μM $ZnCl_2$ and 0.04% polyethylene glycol 20 000. After 10 min incubation, 2 aliquots of 50 μl were removed and transferred into 384-well Costar polypropylene plates (Corning, NY) and read using a Tecan Ultra plate reader (Durham, NC) with excitation and emission wavelengths of 485 and 535 nm, respectively.

ESI-FTMS

Sample solutions containing the desired concentrations of NC and oligonucleotides were prepared in 0.15 M ammonium acetate (pH 7.0) according to procedures described earlier (16). The NC protein employed in these experiments is identical to NC_{WT} (Figure 1), except for M in place of I at the N-terminus. After mixing the different components, 5 μl aliquots were loaded on to quartz nanospray emitters (New Objective, Woburn, MA) and a platinum wire was inserted from the end to provide the necessary voltage. All determinations were performed on a Bruker (Billerica, MA) Apex III Fourier transform mass spectrometer equipped with a 7T shielded superconductive magnet and a thermally-assisted Apollo atmospheric pressure ion source. All data were acquired in negative ion mode and processed using XMASS 6.0.1 (Bruker, Billerica, MA). Typical analysis provided ∼150 000 resolution and <10 p.p.m. mass accuracy in broadband mode, using a three-points external calibration.

Mathematical treatment of results

All results were analyzed as described in detail in Supplementary Data.

RESULTS

SPR analysis

The sequences of HIV-1 NC protein and of the mutants used in the present work are shown in Figure 1. We had previously observed that HIV-1 NC exhibits sequence-specific binding to oligonucleotides containing the sequence $d(TG)_n$ (13), and that 5 bases of the repeating sequence $d(TG)_n$ are necessary

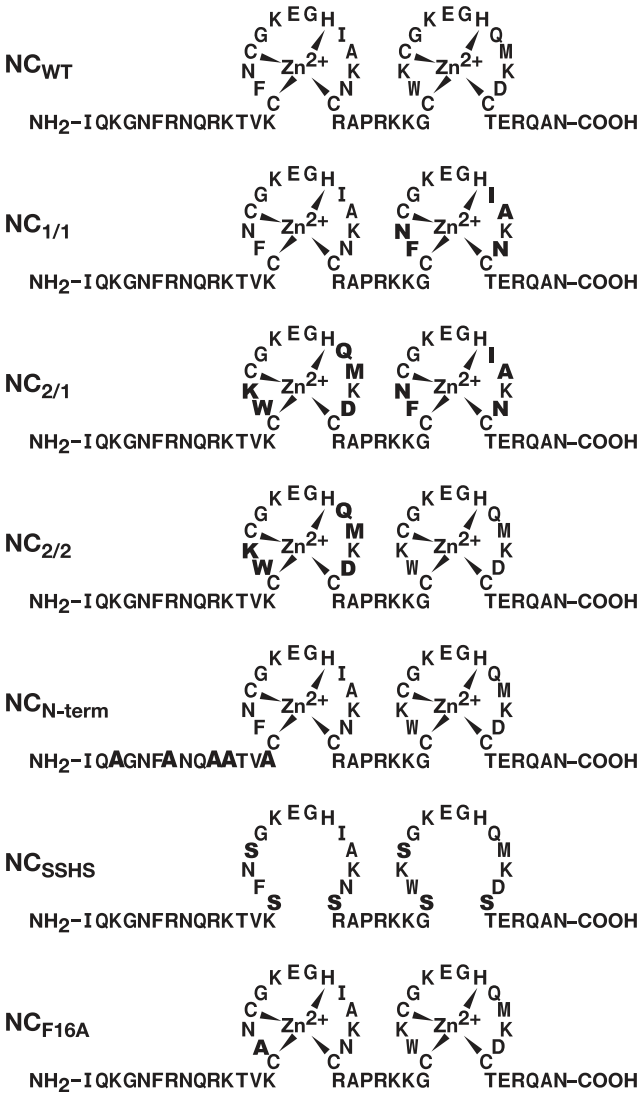

Figure 1. Sequences of NC and NC mutants. Changes from wild-type (WT) are shown in bold.

and sufficient for stable NC binding. We first measured the binding of wild-type NC to the octanucleotide $d(TG)_4$ by SPR, using a buffer containing 0.15 M NaCl. As in our previous studies (13), the oligonucleotide was immobilized on a BIAcore chip. Solutions of NC (at a series of different NC concentrations) were passed across the chip, and bound NC was then eluted with buffer. The instrument records the time-course of binding of NC to, and its subsequent dissociation from, the oligonucleotide on the chip. Our initial experiments were with a chip containing the oligonucleotide at an extremely low density: the average distance between the immobilized 3′ ends of the oligonucleotide molecules on this chip, as calculated from the concentration of oligonucleotide on the chip and the Poisson distribution, was 420 Å. Since NC contains only 55 amino acids, it would be only ∼200 Å long if it were fully extended. Thus, immobilization on the chip served to physically isolate oligonucleotide molecules from each other, so that NC bound to one $(TG)_4$ molecule could not interact with other $(TG)_4$ molecules. The resulting SPR profile is shown in Figure 2A, in which the horizontal line represents

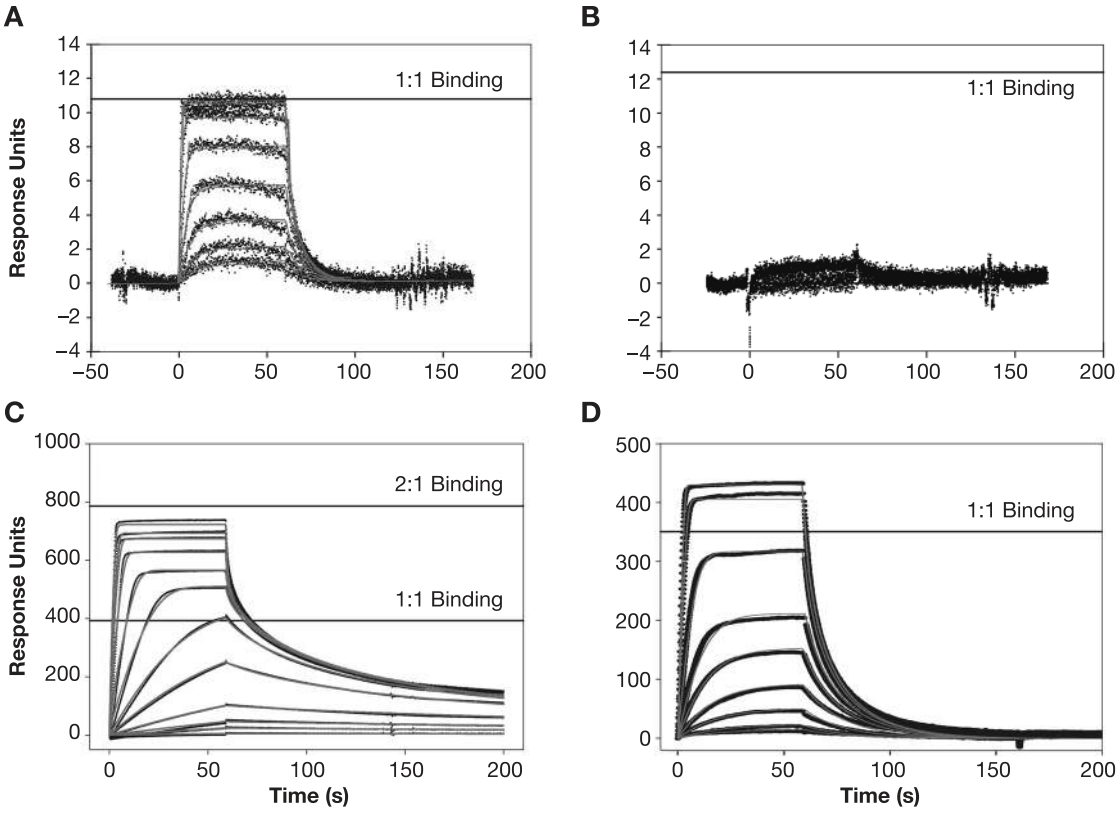

Figure 2. SPR results: (**A**) Binding of NC to a 5.4 RU surface of $(TG)_4$ in 150 mM NaCl. Solutions of 1.7, 3.13, 6.25, 12.5, 25, 50, 100 and 200 nM NC were passed over the surface at 100 µl/min. The horizontal line represents the response expected for a 1:1 complex of NC and $(TG)_4$. The grey lines are the profiles expected from fitting the data to a simple 1:1 binding model as described in the Supplementary Data. (**B**) Binding of NC to $(A)_8$. Solutions of NC as in (A) were passed over a 5.8 RU $(A)_8$ surface. (**C**) Binding of NC to a 156.6 RU surface of $(TG)_4$. Solutions of 1, 2.5, 5, 10, 25, 50, 100, 200, 400, 600, 800 and 1000 nM NC were passed over this surface. The two horizontal lines represent the signals expected if one (lower line) or two (upper line) NC molecules binds to each $(TG)_4$ molecule. The grey lines are the profiles expected from fitting the data to the complex binding model shown in Figure 3, as described in the Supplementary Data. (**D**) Binding of NC to $(TG)_4$ in buffer containing 250 mM rather than 150 mM NaCl. Solutions of 5, 10, 25, 50, 100, 200, 400, 800 and 1200 nM NC were passed over a surface containing 159.4 RU of $(TG)_4$. The horizontal line represents the response expected for a 1:1 complex of NC and $(TG)_4$. Grey lines are as in (C).

Table 1. Binding of wild-type NC to $(TG)_4$: summary of K_d determinations by different methods

Method	[NaCl] (M)	K_d1[a] (nM)–NO	K_d2 (nM)–NON	K_d3 (nM)–ONO	ΔAIC[b]
SPR, very low density	0.15	2.95 ± 0.021	–	–	–
SPR	0.15	5.4 ± 0.08	25 ± 0.5	1600 ± 970	+
TFQ	0.15	4.70 ± 0.67	49.9 ± 10.23	239.9 ± 20.05	+
FA	0.15	2.7 ± 0.4	21 ± 6.6	5300 ± 10 700	+
SPR	0.25	200 ± 0.5	1300 ± 350	–	+

[a]K_d values were determined by global analysis of the SPR, TFQ or FA data using the equations in Supplementary Data. The error terms were obtained using a linear approximation.
[b]ΔAIC: Akaike's Information Criterion statistic (see Supplementary Data). ΔAIC > 0 implies that the complex model (see Figure 3) gives a better fit to the data than a 1:1 binding model.

the response expected if each NC molecule bound to one $(TG)_4$ molecule. It can be seen that a clear saturation was obtained at a level corresponding to one NC bound to each oligonucleotide molecule. This level shows that under these conditions, only a single NC molecule can bind to the octanucleotide $(TG)_4$. The data were analyzed as described in Materials and Methods, and were consistent with a single binding system with a K_d of only 3.0 nM (Table 1). The grey lines in the Figure are the curves obtained from the fit to this model. Figure 2B shows that no binding to $(A)_8$ can be detected under these conditions, as expected from our earlier analysis using SPR chips more densely loaded with oligonucleotide (13).

We also repeated the SPR analysis using a chip displaying a higher density of $(TG)_4$. As can be seen in Figure 2C, the results were quite different in this case. First, the binding did not reach saturation; rather, increasing the concentration of NC applied to the chip led to a rise in the amount of NC bound, exceeding that expected for equimolar binding to the $(TG)_4$ (this level is indicated by the lower horizontal line in the Figure). Secondly, the kinetics of elution from the chip

were sharply biphasic, with some NC washing off very rapidly and another portion eluting much more slowly: note that by 75–100 s, the binding curves in Figure 2A have fallen to zero, while in Figure 2C some NC remains bound to the oligonucleotide at 150–200 s, even in the most dilute NC solutions. The results suggest that at high $(TG)_4$ densities, NC is 'trapped' on the chip. These data are in complete accord with our previous findings (13).

Taken together, the SPR results indicate that the interaction between NC and $(TG)_4$ is potentially quite complex. The binding of NC at a level greater than equimolar (Figure 2C) implies that under these conditions, additional NC can bind to an NC:$(TG)_4$ complex. In addition, the fact that the SPR profile is a function of the $(TG)_4$ density on the chip (compare Figure 2C with A) suggests that when NC is bound to oligonucleotide, it can still interact with a second oligonucleotide molecule, provided that the latter is in close enough proximity. [Indeed, we previously presented SPR data showing that the stable binding of NC to the tetranucleotide $(TG)_2$ is dependent upon the spacing of $(TG)_2$ molecules on the chip (13)]. In other words, both NC and $(TG)_4$ appear to be bivalent in NC:$(TG)_4$ interactions.

These properties obviously have significant implications for the quantitative analysis of these interactions. For example, when $(TG)_4$ is titrated into an NC solution, the latter will initially be in vast excess. While the first binding event will lead to a 1:1 complex (which we will refer to as 'NO', where 'N' and 'O' represent NC and oligonucleotide, respectively), the NO complexes are evidently capable of interacting with additional NC, generating 'NON' complexes (as indicated in the upper pathway in Figure 3). Conversely, when NC is added stepwise to a $(TG)_4$ solution, the initial NO complexes will likely interact with a second $(TG)_4$ molecule to form 'ONO' complexes (lower pathway in Figure 3). Moreover, if, as suggested by these considerations, both reactants are bivalent, then higher-order complexes can also be formed: e.g. if NC has two binding sites for oligonucleotides, then the two NC molecules in an NON complex will each have an unfilled site. Thus an NON complex can potentially bind additional $(TG)_4$ molecules, giving rise to ONON or ONONO. In turn, since $(TG)_4$ can evidently bind two NC molecules, ONON has the potential to form NONON, etc. The most probable result would appear to be a heterogeneous mixture of complexes between the protein and the oligonucleotide.

As noted above, the results in Figure 2C are clearly incompatible with a simple model in which interactions are limited to NO formation as in the first pathway in Figure 3. We analyzed the data using the full scheme shown in Figure 3. Assuming the occurrence of the additional binding reactions, leading to NON and ONO complexes in addition to NO, gave an excellent fit to the data, as shown by the red lines in the Figure.

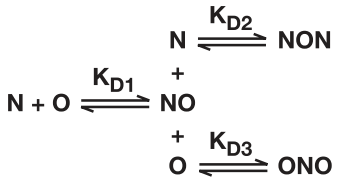

Figure 3. Proposed reaction-scheme for interaction of NC with $(TG)_4$, showing three possible binding reactions, each with its respective K_d. N, NC; O, $(TG)_4$.

The K_d's calculated for the three reactions are 5.4 nM for NO, 1.6 μM for formation of ONO from NO, and 25 nM for formation of NON from NO, as shown in Table 1. The utility of the model shown in Figure 3 for fitting the data is indicated in Table 1 by the positive value for the Akaike's Information Criterion (ΔAIC) statistic. This parameter includes a penalty for introduction of additional variables, and a positive ΔAIC thus implies that each variable makes a valuable contribution to the fit of the data (see Supplementary Data).

It was of interest to probe the contribution of electrostatic forces to the interaction between NC and $(TG)_4$. Therefore, we also analyzed the binding of NC to the oligonucleotide in 250 (rather than 150) mM NaCl, again using an oligonucleotide density similar to that in Figure 2C. As seen in Figure 2D, no saturation was reached, even at 1 μM NC. However, the amount of NC bound to the chip was significantly lower than at 150 mM NaCl (Figure 2C), and the increase in binding above the equimolar level was relatively modest. These data imply that the formation of NON and ONO complexes detected at 150 mM NaCl (Figure 2C) was greatly reduced at 250 mM NaCl. We attempted to fit the results in Figure 2D with a simple 1:1 binding model, and found that the results were satisfactory at NC concentrations up to 100 nM (the 5th curve). However, the level of binding with more concentrated NC solutions was not consistent with this model. Assuming that NON and ONO complexes could also form under these conditions greatly improved the fit of the data at these concentrations, as indicated by red lines in Figure 2D and by the ΔAIC statistic (Table 1). The K_d's for the three reactions, leading to NO, NON and ONO, respectively, are shown in Table 1. The drastic reduction in affinities accompanying an increase in the ionic strength of the solution indicates that there is a major electrostatic component in the interaction of NC with $(TG)_4$.

TFQ

NC contains a single tryptophan residue. The fluorescence of the tryptophan is quenched when NC binds nucleic acids, and many analyses of NC–nucleic acid interaction have used this response as a reporter of the binding. Figure 4 shows the fluorescence at 350 nm as $(TG)_4$ was titrated into a 0.4 μM NC solution in the standard buffer containing 150 mM NaCl. It can be seen that the fluorescence decreased nearly to zero over the course of the titration. The dashed line in Figure 4 shows an attempt to fit these data using the assumption that a single NC molecule interacts with a single $(TG)_4$ molecule; this curve would extrapolate to zero fluorescence at 400 nM $(TG)_4$, when one mole of $(TG)_4$ had been added per mole of NC in the solution. It is evident that the curve (representing the fit) exhibits systematic deviations from the data; thus, the results are not precisely compatible with this model.

We attempted to fit the results to the scheme shown in Figure 3, in which NO can in turn participate in NON or ONO formation (and assuming that all NC molecules in NO, NON and ONO complexes completely lose fluorescence). As shown in the solid line, this model gave an excellent fit with the data. In particular, it is notable that at the lower $(TG)_4$ concentrations, where NC is in excess, the fluorescence falls more steeply than in the dashed line representing the 1:1 binding model. This is presumably because under these

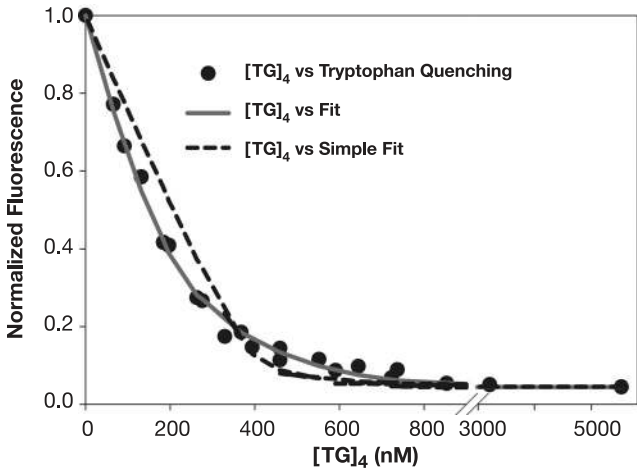

Figure 4. TFQ results. $(TG)_4$ was titrated into a 400 nM solution of NC and tryptophan fluorescence was monitored as described in Materials and Methods. The fluorescence readings were divided by the fluorescence of NC before addition of $(TG)_4$ to obtain the normalized fluorescence shown here. Attempts were made to fit the data assuming a 1:1 binding system (dashed line) (giving a K_d of 3.4 ± 0.07 nM) or the more complex model shown in Figure 3 (solid grey line). The results shown here are normalized by dividing by the fluorescence value of NC alone. As indicated in Supplementary equation 5, the final fluorescence obtained after several equivalents of $(TG)_4$ had been added [i.e. the residual fluorescence of NC bound to $(TG)_4$] was subtracted from these values before attempting to fit the curves.

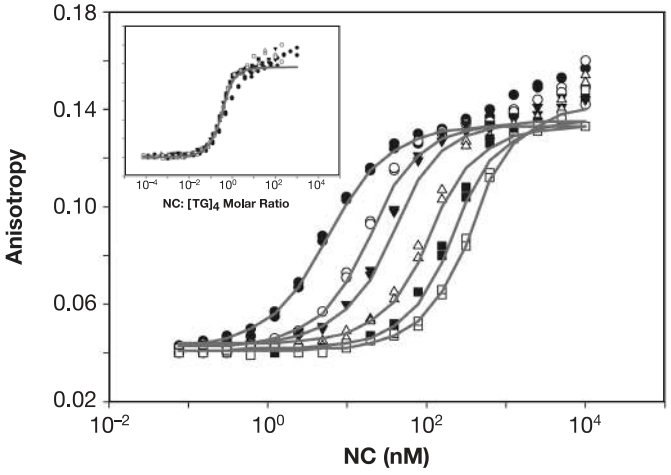

Figure 5. FA results. NC was titrated into solutions containing 10, 50, 100, 300, 600 or 1000 nM fluoresceinated $(TG)_4$ and the anisotropy of the oligonucleotide was measured. Grey lines are fits to the data using the model shown in Figure 3. FA of two aliquots was measured independently for each point in the titrations. Inset: the FA results in the figure were plotted vs. the NC:$(TG)_4$ ratio in the solution.

conditions, a single $(TG)_4$ molecule can quench the fluorescence of both of the NC molecules in a NON complex. In contrast, at higher $(TG)_4$ concentrations (i.e. ~450–650 nM), the fluorescence is quenched slightly more slowly than predicted by the 1:1 binding model; we would suggest that this is due to the binding of some $(TG)_4$ molecules to NO complexes, forming ONO and having no effect on NC fluorescence. The three K_d's calculated from this fit are shown in Table 1, and are very similar to those obtained by SPR. Again, the superiority of the fit using this model is indicated by the positive ΔAIC value. Results of fits using only K_d1 and K_d2 or only K_d1 and K_d3 are shown in Supplementary Figure S2.

FA

We also analyzed interactions between NC and $(TG)_4$ by FA. NC was titrated into solutions of fluoresceinated $(TG)_4$ and the anisotropy of the fluorescein reporter was monitored. Results of this experiment, for a series of different $(TG)_4$ concentrations between 10 nM and 1 μM, are shown in Figure 5. It can be seen that addition of NC increased the anisotropy of the fluoresceinated oligonucleotide in a dose-dependent manner. As expected, each $(TG)_4$ concentration gave a different titration curve; i.e. the concentration of NC required to shift the fluorescein anisotropy is dependent on the $(TG)_4$ concentration. It is striking to note that there was no strict plateau in the anisotropy values; rather, the anisotropy continued to rise slowly, even at the highest NC concentrations. This observation supports the idea that complexes between NC and $(TG)_4$ are still capable of interacting with additional NC to form NON-type complexes.

The data shown in Figure 5 was analyzed in several ways. The inset shows the same titration data, but here each FA value is plotted against the molar ratio of NC to $(TG)_4$ present in the solution. It is evident that when the results are graphed in this way, all of the titrations except that with 10 nM $(TG)_4$ fall on the same line. (Presumably the curve obtained with 10 nM $(TG)_4$ diverges from the others because this concentration of oligonucleotide is near the K_d1 of the reaction, so that the two reactants do not associate with each other completely until excess NC has been added.) The curve representing the higher NC concentrations was analyzed to determine the 'valency' of NC, i.e. the maximum number of $(TG)_4$ that can be bound by an NC molecule, as described in Supplementary Data. In three separate experiments, this value was found to be 2.17 ± 0.25. The fact that the titration does not saturate until ~2 equivalents of NC have been added shows clearly that the two reactants can interact further after formation of the expected 1:1 complex.

We also fit the data in Figure 5 using the complex model shown in Figure 3. As indicated by the lines in Figure 5, this treatment gave good fits to the data, yielding the three K_d values shown in Table 1. The values for the 3 K_d's are similar to those obtained by SPR and TFQ, and the superiority of the fit obtained from the model in Figure 3 is indicated by its positive ΔAIC value. (It will be evident that the fit is imperfect at very high molar ratios; this is presumably due to the formation of higher-order complexes such as NONON, which are not used in fitting the data.)

Dissection of the interaction of NC with $(TG)_4$

NC is a highly basic protein, with a particularly basic N-terminal region; it also contains two zinc fingers (Figure 1). The zinc fingers are known to participate in binding to nucleic acids *in vitro* and are critical in some tests of nucleic acid chaperone activity *in vitro* (15,17–20). They are also essential for selective packaging of genomic RNA during virus assembly by the Gag polyprotein *in vivo* (2).

In order to explore the role of the basic N-terminal residues in binding of NC to $(TG)_4$, we analyzed binding by the

Table 2. Summary of K_d values for mutant NC proteins

Protein	Method	K_d1[a] (nM)–NO	K_d2 (nM)–NON	K_d3 (nM)–ONO	ΔAIC[b]
N-term	FA	62 ± 8.8	–	–	–
SSHS	FA	30 500 ± 14 600	–	–	–
F16A	SPR	119 ± 0.5	14 000 ± 3000	987 ± 11	+
1:1 mutant	SPR	10 ± 0.8	–	–	–
2:2 mutant	SPR	48 ± 0.5	–	718 ± 40	+
2:1 mutant	SPR	34 ± 0.6	–	2350 ± 72	+

[a]K_d values in 0.15 M NaCl were determined by global analysis of the SPR or FA data using the equations noted in Supplementary Data.
[b]ΔAIC: Akaike's Information Criterion statistic (see Supplementary Data).

'N-term' mutant, in which the five lysine and arginine residues within the first 14 residues had been replaced with alanines (Figure 1), by FA. We found (data not shown) that the anisotropy induced by the mutant NC rises to a clear plateau, so that there is no increase as the protein concentration is raised from 2 to 10 μM (a 1000-fold excess of protein over $(TG)_4$). In contrast, the curve obtained with wt NC continues to rise slowly over this range, indicating the formation of NON-type complexes. The results with the mutant are consistent with a simple 1:1 binding system with a K_d of 62 nM (see Table 2). Thus, the basic residues in the N-terminus are critical both for the high-affinity of wt NC for $(TG)_4$ and for its ability to interact with NO complexes.

We also tested the contribution of the zinc fingers to the interaction of NC with $(TG)_4$. Again using FA, we analyzed binding of the 'SSHS' NC protein, in which the six zinc-coordinating cysteines were all replaced with serine. We found that the affinity of NC for the oligonucleotide was drastically reduced under these conditions. Only a single binding system with a K_d of 30.5 μM was detectable in this experiment (Table 2). Similar results were obtained with wt NC when the titration was performed in the presence of 5 mM EDTA (data not shown). These data imply that the zinc fingers, as well as the N-terminal basic residues, are crucial for the high-affinity interaction of NC with $d(TG)_4$ and for its ability to form ONO- or NON-type complexes.

As an extension of our exploration of the role of the zinc fingers in the interaction with $(TG)_4$, we also analyzed 'finger-swap' NC proteins, in which the sequence of one or both of the zinc fingers was replaced with that of the other (see Figure 1). Using SPR, we found that the '1:1' mutant protein formed NO complexes with nearly the same K_d1 as wt NC, while the '2:2' and '2:1' proteins had ~10-fold reduced affinity (Tables 1 and 2). Interestingly, the data indicated that only the latter two proteins retained the ability to form ONO complexes, and all three mutants were defective with respect to NON formation (see Table 2). We also investigated the binding of the F16A mutant of NC, in which an alanine residue is substituted for the phenylalanine within the N-terminal finger [known to engage in hydrophobic interactions with nucleic acid bases (21)]. In this case, the affinity of the protein for $(TG)_4$ was reduced to ~20–40-fold relative to wt NC, but it could still form both NON and ONO complexes at detectable levels.

ITC analysis

We also characterized the interactions between NC and $(TG)_4$ by ITC. The results of this analysis are shown in Figure S3 and are discussed in Supplementary Data. Briefly, ~18 kcal/mol was released when the oligonucleotide was titrated into a solution of 18 μM NC. The titration curve centered around a molar ratio of 0.5 and reached saturation at a molar ratio of ~1.0, in agreement with the idea that a single oligonucleotide molecule can bind two NC molecules in a NON-type complex (22). We also performed a 'reverse' titration in which NC was titrated into a solution of 10 μM $(TG)_4$. Approximately 12 kcal/mol was released over the course of the titration, but no clear saturation was observed. Thus the ITC data did not point to a precise stoichiometry for the interactions between NC and $(TG)_4$, as might be expected from their apparent ability to form NO, ONO, NON and other, higher-order complexes (see above). The dependence of the total heat released upon the direction of the titration also underscores the complexity of the interactions between NC and $(TG)_4$. In addition, we probed the role of electrostatic interactions in the binding, both by repeating the titration of NC into $(TG)_4$ in 250 mM, rather than 150 mM, NaCl and by using the 'N-term' mutant protein, in which the five lysine and arginine residues between positions 3 and 14 had been replaced with alanine. The amount of heat released was drastically reduced in both of these experiments, suggesting that that coulombic attractions play a major role in the interactions between NC and $(TG)_4$.

Competition experiments using SPR

Although SPR technology is capable of measuring the binding of NC to high-affinity nucleic acid sequences such as the repeating oligonucleotide, $(TG)_n$ [see reference (13)], no binding to $(A)_n$ is detected under the same conditions, i.e. 150 mM NaCl (pH 7.5) (Figure 2B). Therefore, we devised an indirect SPR approach to detect the interaction of low-affinity ligands such as $(A)_n$ with NC. These experiments used an SPR chip coated with $(TG)_4$ at a density of 344.2 response units (RU), similar to the conditions used in Figure 2C. In preliminary tests, a series of solutions containing different concentrations of NC was passed over the chip, and the initial binding rate of NC to the chip was measured. As expected, the initial binding rate under these conditions was directly proportional to the NC concentration (Figure 6A, B). This result shows that the initial binding rate can be used to measure the concentration of free NC in solution.

To exploit the 'competition' technique to detect binding of NC to $(A)_{30}$, we mixed NC at 100 nM with a series of concentrations of $(A)_{30}$. In all cases, the oligonucleotide was in excess over NC in the solution. The mixtures were incubated for 1 h and then passed over the $(TG)_4$ chip surface. As expected, the initial binding rate of NC to immobilized $(TG)_4$ decreased as the concentration of $(A)_{30}$ increased (data not

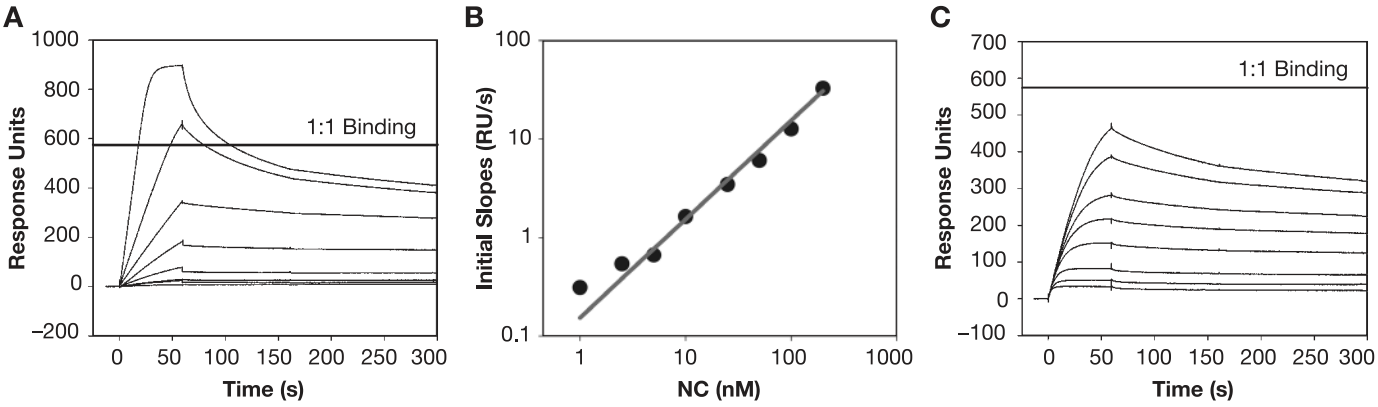

Figure 6. Analysis of competition between oligonucleotides in solution and immobilized $(TG)_4$ for NC. (**A**) Binding of NC to $(TG)_4$. One min injections of (from bottom to top) 0, 1, 2.5, 5, 10, 25, 50, 100 and 200 nM NC were passed over a surface of 344.2 RU's $(TG)_4$ at 64 µl/min. (**B**) Rate of binding of NC to $(TG)_4$ as a function of NC concentration. Initial slopes in (A) were measured using BIAevaluation 3.01, and the plot of initial slopes versus NC was fit using linear regression. (**C**) Effect of $(TG)_4$ in solution on binding of NC to immobilized $(TG)_4$. A total of 100 nM NC was incubated at 4°C with (from top to bottom) 0, 0.1, 0.25, 0.5, 1.0, 2.5, 5.0, 10.0, 25.0, 50.0 and 100.0 µM $(TG)_4$ for at least 1 h before injecting a sample for 1 min at 64 µl/min.

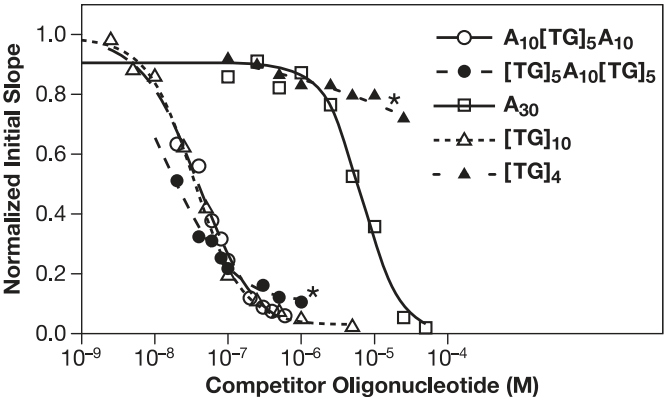

Figure 7. Competition by oligonucleotides in solution with the binding of NC to immobilized $(TG)_4$. Solutions of 100 nM NC were incubated with varying amounts of different oligonucleotides for at least 60 min. These solutions were then passed over an SPR surface containing 327 RU's of $(TG)_4$ and the initial slope of the SPR profile, representing the initial rate of binding to the immobilized $(TG)_4$, was measured. The open squares represent preincubation with 0.1, 0.25, 0.5, 1.0, 2.5, 5, 10, 25 and 50 µM $(A)_{30}$; open triangles: preincubation with 2.5, 5.0, 10, 25, 50, 100, 250, 500, 1000 and 5000 nM $(TG)_{10}$; open circles: preincubation with 20, 40, 60, 80, 100, 300, 500 and 1000 nM $(A)_{10}(TG)_5(A)_{10}$; closed circles: preincubation with 20, 40, 60, 80, 100, 300, 500 and 1000 nM $(TG)_5(A)_{10}(TG)_5$; closed triangles: preincubation with 0.1, 0.25, 0.5, 1.0, 2.5, 5, 10 and 25 µM $(TG)_4$. The lines are fits to a 4-parameter logistic equation, as described in Supplementary Data, yielding the values shown in Table 3. Asterisks highlight the two curves in which IS_{min} is greater than zero.

shown). This decrease reflects the removal of free NC from solution as it is bound by $(A)_{30}$. The relationship between $(A)_{30}$ concentration and the amount of NC removed provides a way of measuring the binding of NC to $(A)_{30}$. As shown in Figure 7 (open squares), addition of sufficient $(A)_{30}$ completely eliminated binding of NC to $(TG)_4$, as expected. The results of these experiments were analyzed using the 4-parameter logistic equation as described in Supplementary Data, yielding an IC_{50} value (the concentration of oligonucleotide in solution required to depress the initial rate of binding to $(TG)_4$ halfway to its final value) of 6.7 µM (see Table 3). The fact that the binding can be completely suppressed by addition of sufficient $(A)_{30}$ is indicated in Table 3 by the value of zero for the 'minimum initial slope' (IS_{min}) (the normalized initial rate of binding to the immobilized $(TG)_4$ that would still occur in an infinite concentration of oligonucleotide in solution).

Table 3. IS_{min} and IC_{50} values for competition by oligonucleotides

Protein	Oligonucleotide	IS_{min}[a]	IC_{50} (nM)[a]
wt NC	$A_{10}(TG)_5A_{10}$	0.00	39.7 ± 3.67
wt NC	$(TG)_5A_{10}(TG)_5$	0.11 ± 0.023	15.96 ± 1.68
wt NC	$(A)_{30}$	0.00	6676 ± 785
wt NC	$(TG)_{10}$	0.00	36.96 ± 2.73
N-term	$(TG)_4$	0.00	602.6 ± 71.7
EDTA–NC	$(TG)_4$	–	–
wt NC	$(TG)_4$	0.77 ± 0.023	263.6 ± 100.8

[a]IC_{50} and IS_{min} values were determined from the SPR competition data using a 4-parameter logistic equation as described in Supplementary Data.

This competition protocol can be used to analyze the binding to high-affinity, as well as low-affinity substrates. We performed the same experiment using $(TG)_{10}$ in place of $(A)_{30}$. As shown in Figure 7 (open triangles), the IC_{50} for $(TG)_{10}$ was ~37 nM, and the IS_{min} in this interaction was zero (see Table 3).

Non-ideal behavior in competition experiments

We also tested the ability of $(TG)_4$ in solution to compete with immobilized $(TG)_4$ on the SPR chip surface. Since these oligonucleotides are identical, we expected the $(TG)_4$ in solution to compete for NC with an IC_{50} in the low nanomolar range. However, the results were quite different from our predictions and from the results obtained with $(A)_{30}$ or $(TG)_{10}$. As shown in Figure 6C, the final height of the curves, i.e. the final amount of NC bound to immobilized $(TG)_4$ at steady-state, was, as expected, quite sensitive to the concentration of $(TG)_4$ in the solution. Thus, in all cases the $(TG)_4$ in solution was able to depress the binding of NC to the immobilized $(TG)_4$, and at high $(TG)_4$ concentrations the SPR signal was reduced nearly to zero. However, the initial slopes (i.e. the initial rates of binding of NC to the chip) obtained with solutions of different $(TG)_4$ concentrations were all very similar to each other, even when the $(TG)_4$ in solution was in 1000-fold excess over NC

(see Figure 6C). This observation implies that the mechanism of the competitive inhibition of NC binding is not that expected from a classical 1:1 binding event.

The results of a similar experiment are summarized by the closed triangles in Figure 7. It can be seen that the initial slope was reduced only ~20% by preincubation with 10^{-5} M $(TG)_4$ (a 100-fold excess of $(TG)_4$ over NC), and responded only slightly to an additional 2.5-fold increase in the concentration of $(TG)_4$ in the solution. These data are readily explained by the hypothesis that complexes formed in solution between NC and $(TG)_4$ retain the ability to interact with $(TG)_4$ on the SPR chip, forming ONO-type complexes, although the steady-state level of these complexes is low in the presence of excess $(TG)_4$ in solution. The results are summarized in Table 3, showing that the IS_{min} is 77%. In other words, the initial rate of binding of NC to immobilized $(TG)_4$ would only be reduced 23% by an infinite concentration of $(TG)_4$ in solution. These results underscore the remarkable ability of NC to enter into higher-order complexes with nucleic acids containing the sequence $(TG)_n$.

It was of interest to assess the contributions of the NC zinc fingers and basic residues to the interactions with $(TG)_4$ observed here. We therefore tested the N-term ala mutant in an experiment analogous to that shown by the closed triangles in Figure 7. We found (Table 3) that the IS_{min} was zero for the mutant, with an IC_{50} of ~600 nM. Thus, the basic residues in the N-terminal region of NC are required for the ONO-type complex formation detected in these experiments. NC from which the Zn^{+2} had been removed with 5 mM EDTA showed no detectable binding to $(TG)_4$ in an SPR experiment.

Binding of NC to chimeric oligonucleotides containing both low-affinity and high-affinity sequences

As another approach to understanding the structure and mode of formation of ONO complexes, we also analyzed the binding of NC to a pair of oligonucleotides which contained both low-affinity [i.e. $(A)_n$] and high-affinity [$(TG)_n$] sequences. In one of these oligonucleotides, the $(TG)_n$ sequences were at the ends and the $(A)_n$ sequence was internal; in the other, terminal $(A)_n$ sequences surrounded a $(TG)_n$ stretch.

NC was mixed with different concentrations of these oligonucleotides, and the mixtures were then analyzed in a competition experiment by passing them over an SPR chip containing $(TG)_4$, as in Figure 6. We found (Figure 7, open circles) that the 30-base oligonucleotide with terminal low-affinity sequences, $(A)_{10}(TG)_5(A)_{10}$, like $(A)_{30}$ (open squares) or $(TG)_{10}$ (open triangles), was able to completely prevent the binding of NC to $(TG)_4$ on the SPR surface (i.e. the IS_{min} was 0, as indicated in Table 3). The IC_{50} with which this oligonucleotide inhibited binding was 40 nM.

The addition of the other chimeric oligonucleotide to the NC solution produced a somewhat different result. As shown in Figure 7 (closed circles) and Table 3, $(TG)_5(A)_{10}(TG)_5$ inhibited the binding of NC to immobilized $(TG)_4$ with an IC_{50} of 16 nM. However, this oligonucleotide could not completely prevent the binding of NC to immobilized $(TG)_4$: the initial rate of binding to the chip was only reduced to ~87% even when this oligonucleotide was at concentrations as high as 1 μM. The non-zero IS_{min}, 0.11, obtained for this oligonucleotide (Table 3) implies that when NC is bound to it in solution, the protein is still capable of interacting with $(TG)_4$ on the SPR surface to a significant degree. In other words, when $(TG)_n$ is at the ends of an oligonucleotide, the oligonucleotide can enter into an ONO complex under our experimental conditions, but if the oligonucleotide has internal $(TG)_n$ and terminal $(A)_n$ sequences, it cannot.

Direct detection of NON complexes

Finally, we characterized the composition of NC:oligonucleotide mixtures by electrospray ionization Fourier transform mass spectrometry (23). This ionization technique has proven capable of preserving relatively weak non-covalent interactions, including those involved in protein–nucleic acid complexes (24). More specifically, we recently employed ESI-FTMS to determine the stoichiometry and binding affinity of the complexes of NC with RNA stem–loops of the HIV-1 packaging signal (16). Under similar conditions, a 5 μM solution of $(TG)_4$ provided intense signals with mass over charge ratio (m/z) of 822.48 and 1234.22, corresponding to the −3 and −2 charge states, respectively (see Figure 8A). Small amounts of $(TG)_4$ in which a guanine-base has undergone hydrolysis [$(TG)_4$-G] were also detected together with dimeric forms of intact and hydrolyzed $(TG)_4$, which could be explained by possible 'wobble' base-pairing between G and T. No other nucleic acid species were detected in this analysis.

The addition of an equimolar amount of NC produced a mixture in which a 1:1 complex is the predominant species (Figure 8B). Consistent with an equilibrium situation, very small amounts of free NC and free $(TG)_4$ were still detectable in the solution. Most interestingly, when the ratio of NC to

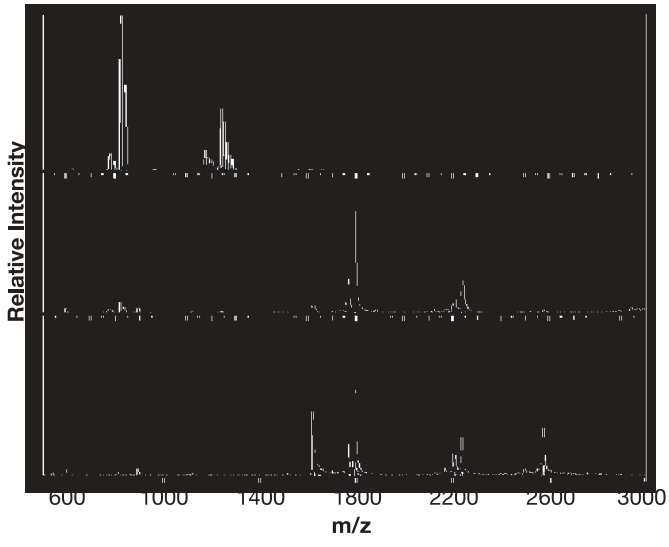

Figure 8. ESI-FTMS spectra of $(TG)_4$ and $(TG)_4$-NC complexes. (**A**) Control 5 μM $(TG)_4$, providing a 2470.46 Da experimental mass (2470.44 Da monoisotopic mass calculated from sequence, for a ~8 p.p.m. mass accuracy). The signal labeled −G corresponds to a product in which a deoxyguanine is formally replaced by deoxyribose (observed decremental mass of ~133 Da). D indicates $(TG)_4$/$(TG)_4$-G dimeric species. (**B**) Mixture containing 5 μM each of $(TG)_4$ and NC. NC provided a mass of 6488.89 Da (6488.91 Da monoisotopic from sequence), while the 1:1 NO complex was 8959.33 Da (8959.35 Da monoisotopic from sequence). (**C**) Sample containing 5 μM $(TG)_4$ and 15 μM NC. The 1:2 NON complex provided a mass of 15 447.93 Da (15 448.25 Da monoisotopic from sequence).

$(TG)_4$ was increased to a final 3:1, an additional species was observed in the corresponding ESI-FTMS spectrum (Figure 8C) consistent with a complex of one $(TG)_4$ molecule with two NC molecules (referred to as NON above). The fact that free NC is still present in the solution suggests that the affinity of NC for an NC:$(TG)_4$ complex (i.e. K_d2) is considerably weaker than its affinity for naked $(TG)_4$. In other words, while NON could be detected here, there is no evidence for cooperativity in the binding of NC to the oligonucleotide. These results are in excellent agreement with those reported in Table 1. It is interesting to note that no signals could be found for dimeric $(TG)_4$ species upon addition of NC (Figure 8B and C), suggesting that the affinity of $(TG)_4$ for NC is higher than its affinity for a second $(TG)_4$ molecule. Further addition of NC to achieve a 5-fold excess over $(TG)_4$ did not result in the formation of detectable levels of higher-order stoichiometries (data not shown), thus indicating that NON is the largest complex observed under these experimental conditions.

When 5 μM $(TG)_4$ was mixed with ~2 μM NC, we only observed the 1:1 NO complex in addition to free $(TG)_4$. This would suggest that the tendency to form ONO complexes under conditions of $(TG)_4$ excess is relatively weak, again in good agreement with the K_d3 results shown in Table 1. We also detected virtually no binding of NC to $(A)_8$ in these experiments, as expected (13). In analogy with the SPR competition experiments presented in Figures 6 and 7 and Table 3, ~6 μM NC was added to a solution containing 5 μM each of $(TG)_4$ and $(A)_8$. The resulting solution was found to contain both NO and NON complexes with $(TG)_4$, but no evidence for binding to $(A)_8$ was obtained (data not shown). Thus, the affinity of NC for a NC:$(TG)_4$ complex (again, K_d2), while weaker than K_d1, is still stronger than its affinity for $(A)_8$.

DISCUSSION

In the present work, we have used a variety of techniques, including SPR, FA, ITC, TFQ and ESI-FTMS to detect and characterize the binding of NC to the octanucleotide $(TG)_4$. These results led to the model presented in Figure 3, in which a 1:1 complex between NC and $(TG)_4$ can be augmented either by an additional NC molecule (leading to an 'NON' complex) or an additional $(TG)_4$ molecule (forming 'ONO'). Direct evidence was obtained for the formation of these higher-order complexes, since (i) the results with SPR were a function of the density of $(TG)_4$ on the SPR chip (Figure 2); (ii) in both SPR and FA, we did not observe a complete saturation when excess NC was added to $(TG)_4$ (Figures 2C and 5); (iii) addition of some oligonucleotides to NC in solution caused only a partial reduction in the rate of binding of NC to $(TG)_4$ on an SPR chip (Figures 6C and 7, Table 3); and (iv) NON complexes were directly observed by ESI-FTMS (Figure 8C). In fact, the only experimental situation in which the results were compatible with simple 1:1 binding of NC to $(TG)_4$ was on SPR chips containing a very low density of $(TG)_4$; this configuration evidently precluded interaction of the NC: $(TG)_4$ complex with additional $(TG)_4$ molecules. The model in Figure 3 was applied in the quantitative analysis of the results obtained by the other techniques used in the present work.

The binding results using all of the techniques are summarized in Table 1. It can be seen that the conclusions derived from SPR, FA and TFQ measurements were generally in very good agreement with each other. We conclude that under our standard conditions [0.15 M NaCl (pH 7.5)], NC binds with remarkably high-affinity (K_d of ~5 nM) to the simple repeating sequence, $(TG)_n$. It can then proceed to form either NON or ONO; the ability to form both types of ternary complex also raises the possibility of further permutations, e.g. NONO or ONON, etc. The data also indicate that the high affinity of NC for $(TG)_4$ depends upon both the positively charged groups in the protein, which presumably engage in electrostatic interactions with the nucleic acid, and the zinc fingers, which are known to be important in binding of NC to other nucleic acids (21,25). Evidence for the electrostatic contribution included the reduction in affinity observed when some of the positive charges were removed from the protein (Table 2) or when additional counterions were added (Figure 2D), and the reduced enthalpy of binding in the presence of added counterions (Supplementary Figure S3).

One important question which is left unanswered by our results is why proximity of $(TG)_4$ molecules, as in Figure 2C, is required for formation of NON as well as ONO complexes. It might be proposed that a high $(TG)_4$ concentration leads to the formation of $(TG)_4$ dimers or oligomers, and that these preformed oligomers then participate in NON formation. Indeed, $(TG)_4$ dimers were detected by ESI-MS (Figure 8A). However, addition of one or more equivalents of NC reduced their concentration to below detectable levels (Figure 8B and C), suggesting that they are dissociated by interaction with NC. This result appears to argue against a role for these dimers in the formation of ternary NC–$(TG)_4$ complexes. Furthermore, if dimer formation were a major contributor to K_d3, then the affinity would be expected to increase with ionic strength. However, raising the salt concentration was found to decrease the affinities for all three binding reactions (Table 1, Figure 2C), thus casting further doubt on the possible contribution of preformed dimers.

We showed previously that a 5-base oligonucleotide is long enough to support binding of NC (13). Thus, the octanucleotide $(TG)_4$ potentially contains as many as four overlapping 5-base binding sites. However, SPR results with the two pentanucleotides, (TGTGT) and (GTGTG), displayed on the chip at densities similar to that used in Figure 2C, were very similar to those shown in Figure 2C (13). Thus the 'redundancy' of $(TG)_4$ is not responsible for the high-affinity or complexity of its interactions with NC. It should be noted that an earlier study (26) measured the affinity of NC for $(TG)_4$ at low ionic strength and the dependence of the affinity upon the Na^+ concentration. Extrapolation of these values to the ionic strength used here gives a significantly lower affinity for the interaction than we observed for K_d1. We are not certain why this extrapolation method gave a different K_d from the mutually consistent results obtained by three independent methods, i.e. SPR, FA and direct TFQ measurements, in the present work.

The ability of NC to form ONO-type complexes indicates that it contains several domains or sites that bind nucleic acids. We cannot say precisely how many such domains are present in NC, since we do not know whether the two zinc fingers can function independently, nor whether basic regions

other than the N-terminus can bind nucleic acids. As an approach to this question, we analyzed 'finger-swap' mutant NC proteins, in which one or both of the zinc fingers was placed in the position of the other. We found that the 2:1 and 2:2 proteins had significantly lower affinity for $(TG)_4$ than the wild-type protein, although this affinity was far higher than when the ability to coordinate zinc was completely eliminated (Table 2). Results with F16A NC, lacking an aromatic residue in the first finger that is known to participate in tight binding to RNA stem–loops (21), were somewhat similar to those with 2:1 and 2:2. The reduced binding of the finger-swap mutants might indicate that the fingers normally function in concert with each other. An alternative explanation, however, is that a single zinc finger can function as an autonomous nucleic acid-binding domain, but its functional boundaries extend beyond the N- and C-terminal cysteines of the finger. The ratio 1:1 NC, in which only the C-terminal finger has been changed, formed NO complexes with nearly the same affinity as wt NC. It is interesting to note that HIV-1 containing 1:1 NC is more replication-competent than those with the other finger-swaps (27).

We had previously reported that the tetranucleotide, $(TG)_2$, cannot support the binding of NC in SPR experiments unless the $(TG)_2$ molecules are in close proximity to each other (13). These experiments suggested that NC could form a relatively stable complex with two tetranucleotide molecules, resembling the ONO-type complexes detected in the present work. We do not know how these complexes are formed, but it is significant that NC bound to a chimeric oligonucleotide, consisting of internal $(TG)_n$ sequences flanked by low-affinity $(A)_n$ sequences, did not engage in detectable ONO-type complex formation (Figure 7, Table 3). This observation raises the possibility that NC must be at the end of one nucleic acid molecule in order to interact with a second. [This may also be the reason that competition by $(TG)_{10}$ gave results in accord with a simple binding model (Figure 7, Table 3): as in the $(A)_{10}(TG)_5(A)_{10}$ oligonucleotide, the internal (TG) sequences provide binding sites for NC which are distant from the ends of the oligonucleotide.]

The possible requirement that NC be near the end of one oligonucleotide in order to interact with a second suggests that it may slide between them in an ONO-type complex. In many ways, this behavior of NC on single-stranded oligonucleotides is analogous to that of other proteins on double-stranded DNA. For example, the ability of lac repressor to rapidly 'find' a high-affinity base sequence in a large double-stranded DNA molecule appears to result from its ability to slide along DNA (28–33). In fact, Winter *et al*. (30) point out that the ability of lac repressor to bind DNA both specifically through sequence-dependent interactions and nonspecifically, i.e. electrostatically, is essential for its facility in finding the operator by a sliding mechanism. It seems possible that the zinc fingers and basic regions of NC play a similar respective roles in the interactions of NC with single-stranded nucleic acid.

NC has been reported to induce aggregation of single-stranded nucleic acids (34–36). These aggregates presumably represent higher-order complexes, which are expected to form if both NC and the nucleic acid are at high concentration and each is able to bind more than one molecule of the other.

What is the biological significance of these remarkable properties of NC protein? The functional importance of the high-affinity of NC for the repeating sequence, $(TG)_n$, or its RNA counterpart, $(UG)_n$, (13) is not clear, but it is noteworthy that alternating $(UG)_n$ sequences are found at several sites in HIV-1 genomic RNA. One (UG)-rich region is nucleotides 556–577 in the U5 region of the NL4-3 isolate of HIV-1. The function of this highly conserved sequence is not known, but it must be crucial for the virus, since subtle changes lead to a profound diminution in replicative capacity (37). These sequences appear to be significant in dimerization of genomic RNA (38), and may also contribute to efficient polyadenylation at the 3′ end of viral RNA (39). Similar sequences are also found in stem–loop 3, a region in the leader that appears to be important for encapsidation of the genome.

As noted above, NC and the NC domain of Gag polyprotein participate in an amazing variety of protein–nucleic acid interactions during the viral life cycle. Many, although not all, of the functions of NC involve its nucleic acid chaperone activity. The molecular mechanisms underlying this activity are not fully understood, but recent work lends strong support to the hypothesis that a major component of chaperone activity is the ability of NC to induce short-range attraction between nucleic acid molecules (10,20,40). Hargittai *et al*. (20) point out that this property is to be expected from the polycationic nature of NC. The results presented here provide the first quantitation of this activity, which we termed 'K_d3', the affinity of a complex between NC and one nucleic acid molecule for a second nucleic acid molecule. It is also notable that this affinity was significantly weaker than K_d1; thus NC does not bind to the short oligonucleotide studied here with positive cooperativity, in contrast to the suggestion that NC:NC interactions contribute to its biological activities (41). However, we cannot exclude the possibility that NC binds cooperatively to longer nucleic acid molecules.

In conclusion, we have used SPR, TFQ, FA, ITC and ESI-FTMS to characterize the binding of HIV-1 NC protein to the simple oligodeoxynucleotide, $(TG)_4$. Our data show that this high-affinity interaction results from the combined action of multiple nucleic acid-binding domains in NC. The protein is capable of binding simultaneously to more than one nucleic acid molecule, and can also bind to complexes between an oligonucleotide and another NC molecule. It seems likely that these attributes contribute to the complex functional properties of NC and of the NC domain of the Gag protein.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We are very grateful to Irina Orlov and Evan Behre at BIAcore Inc. for their work on the BIAcore S51. We also thank Dr Linda Kenney for helpful suggestions on fluorescence anisotropy and Drs. Ioulia Rouzina and Yun-Xing Wang for careful readings of the manuscript. In addition, we thank Leslie Hickman for her helpful analyses of SPR signals and Dominic Scudiero for permission to use the Tecan Ultra fluorescence anisotropy plate reader. This research was supported in part by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research and was funded in part

with Federal funds from the National Cancer Institute, National Institutes of Health, under Contracts NO1-CO12400 and NO1-CO12401. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products or organization imply endorsement by the U.S. Government. The participation of D.F. and N.A.H. was supported by the National Institutes of Health (R01-GM643208 and 1T32-GM066706). Funding to pay the Open Access publication charges for this article was provided by NO1-CO12401.

Conflict of interest statement. None declared.

REFERENCES

1. Swanstrom,R. and Wills,J.W. (1997) Synthesis, assembly, and processing of viral proteins. In Coffin,J.M., Hughes,S.H. and Varmus,H.E. (eds), *Retroviruses*. Cold Spring Harbor Laboratory Press, Plainview, NY, pp. 263–334.
2. Berkowitz,R., Fisher,J. and Goff,S.P. (1996) RNA packaging. *Curr. Top Microbiol. Immunol.*, **214**, 177–218.
3. Muriaux,D., Mirro,J., Harvin,D. and Rein,A. (2001) RNA is a structural element in retrovirus particles. *Proc. Natl Acad. Sci. USA*, **98**, 5246–5251.
4. Khorchid,A., Halwani,R., Wainberg,M.A. and Kleiman,L. (2002) Role of RNA in facilitating Gag/Gag-Pol interaction. *J. Virol.*, **76**, 4131–4137.
5. Feng,Y.X., Campbell,S., Harvin,D., Ehresmann,B., Ehresmann,C. and Rein,A. (1999) The human immunodeficiency virus type 1 Gag polyprotein has nucleic acid chaperone activity: possible role in dimerization of genomic RNA and placement of tRNA on the primer binding site. *J. Virol.*, **73**, 4251–4256.
6. Cen,S., Huang,Y., Khorchid,A., Darlix,J.L., Wainberg,M.A. and Kleiman,L. (1999) The role of Pr55(gag) in the annealing of tRNA3Lys to human immunodeficiency virus type 1 genomic RNA. *J. Virol.*, **73**, 4485–4488.
7. Fu,W., Gorelick,R.J. and Rein,A. (1994) Characterization of human immunodeficiency virus type 1 dimeric RNA from wild-type and protease-defective virions. *J. Virol.*, **68**, 5013–5018.
8. Feng,Y.X., Copeland,T.D., Henderson,L.E., Gorelick,R.J., Bosche,W.J., Levin,J.G. and Rein,A. (1996) HIV-1 nucleocapsid protein induces ''maturation'' of dimeric retroviral RNA *in vitro*. *Proc. Natl Acad. Sci. USA*, **93**, 7577–7581.
9. Rein,A., Henderson,L.E. and Levin,J.G. (1998) Nucleic acid chaperone activity of retroviral nucleocapsid proteins: significance for viral replication. *Trends Biochem. Sci.*, **23**, 297–301.
10. Levin,J.G., Guo,J., Rouzina,I. and Musier-Forsyth,K. (2005) Nucleic acid chaperone activity of HIV-1 nucleocapsid protein: critical role in reverse transcription and molecular mechanism. *Prog. Nucleic Acid Res. Mol. Biol.*, **80**, 217–286.
11. Carteau,S., Gorelick,R.J. and Bushman,F.D. (1999) Coupled integration of human immunodeficiency virus type 1 cDNA ends by purified integrase *in vitro*: stimulation by the viral nucleocapsid protein. *J. Virol.*, **73**, 6670–6679.
12. Buckman,J.S., Bosche,W.J. and Gorelick,R.J. (2003) Human immunodeficiency virus type 1 nucleocapsid zn(2+) fingers are required for efficient reverse transcription, initial integration processes, and protection of newly synthesized viral DNA. *J. Virol.*, **77**, 1469–1480.
13. Fisher,R.J., Rein,A., Fivash,M., Urbaneja,M.A., Casas-Finet,J.R., Medaglia,M. and Henderson,L.E. (1998) Sequence-specific binding of human immunodeficiency virus type 1 nucleocapsid protein to short oligonucleotides. *J. Virol.*, **72**, 1902–1909.
14. Vuilleumier,C., Bombarda,E., Morellet,N., Gerard,D., Roques,B.P. and Mely,Y. (1999) Nucleic acid sequence discrimination by the HIV-1 nucleocapsid protein NCp7: a fluorescence study. *Biochemistry*, **38**, 16816–16825.
15. Guo,J., Wu,T., Anderson,J., Kane,B.F., Johnson,D.G., Gorelick,R.J., Henderson,L.E. and Levin,J.G. (2000) Zinc finger structures in the human immunodeficiency virus type 1 nucleocapsid protein facilitate efficient minus- and plus-strand transfer. *J. Virol.*, **74**, 8980–8988.
16. Hagan,N. and Fabris,D. (2003) Direct mass spectrometric determination of the stoichiometry and binding affinity of the complexes between nucleocapsid protein and RNA stem-loop hairpins of the HIV-1 Psi-recognition element. *Biochemistry*, **42**, 10736–10745.
17. Guo,J., Wu,T., Kane,B.F., Johnson,D.G., Henderson,L.E., Gorelick,R.J. and Levin,J.G. (2002) Subtle alterations of the native zinc finger structures have dramatic effects on the nucleic acid chaperone activity of human immunodeficiency virus type 1 nucleocapsid protein. *J. Virol.*, **76**, 4370–4378.
18. Williams,M.C., Rouzina,I., Wenner,J.R., Gorelick,R.J., Musier-Forsyth,K. and Bloomfield,V.A. (2001) Mechanism for nucleic acid chaperone activity of HIV-1 nucleocapsid protein revealed by single molecule stretching. *Proc. Natl Acad. Sci. USA*, **98**, 6121–6126.
19. Williams,M.C., Gorelick,R.J. and Musier-Forsyth,K. (2002) Specific zinc-finger architecture required for HIV-1 nucleocapsid protein's nucleic acid chaperone function. *Proc. Natl Acad. Sci. USA*, **99**, 8614–8619.
20. Hargittai,M.R., Gorelick,R.J., Rouzina,I. and Musier-Forsyth,K. (2004) Mechanistic insights into the kinetics of HIV-1 nucleocapsid protein-facilitated tRNA annealing to the primer binding site. *J. Mol. Biol.*, **337**, 951–968.
21. De Guzman,R.N., Wu,Z.R., Stalling,C.C., Pappalardo,L., Borer,P.N. and Summers,M.F. (1998) Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element. *Science*, **279**, 384–388.
22. Wiseman,T., Williston,S., Brandts,J.F. and Lin,L.N. (1989) Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem.*, **179**, 131–137.
23. Comisarow,M.B. and Marshall,A.G. (1974) Fourier transform ion cyclotron resonance. *Chem. Phys. Lett.*, **25**, 282–283.
24. Hofstadler,S.A., Sannes-Lowery,K.A. and Hannis,J.C. (2005) Analysis of nucleic acids by FTICR MS. *Mass Spectrom Rev.*, **24**, 265–285.
25. Amarasinghe,G.K., De Guzman,R.N., Turner,R.B., Chancellor,K.J., Wu,Z.R. and Summers,M.F. (2000) NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the psi-RNA packaging signal. Implications for genome recognition. *J. Mol. Biol.*, **301**, 491–511.
26. Wu,J.Q., Ozarowski,A., Maki,A.H., Urbaneja,M.A., Henderson,L.E. and Casas-Finet,J.R. (1997) Binding of the nucleocapsid protein of type 1 human immunodeficiency virus to nucleic acids studied using phosphorescence and optically detected magnetic resonance. *Biochemistry*, **36**, 12506–12518.
27. Gorelick,R.J., Chabot,D.J., Rein,A., Henderson,L.E. and Arthur,L.O. (1993) The two zinc fingers in the human immunodeficiency virus type 1 nucleocapsid protein are not functionally equivalent. *J. Virol.*, **64**, 3207–3211.
28. Berg,O.G., Winter,R.B. and von Hippel,P.H. (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry*, **20**, 6929–6948.
29. Winter,R.B. and von Hippel,P.H. (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 2. The *Escherichia coli* repressor–operator interaction: equilibrium measurements. *Biochemistry*, **20**, 6948–6960.
30. Winter,R.B., Berg,O.G. and von Hippel,P.H. (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The *Escherichia coli* lac repressor–operator interaction: kinetic measurements and conclusions. *Biochemistry*, **20**, 6961–6977.
31. Barkley,M.D., Lewis,P.A. and Sullivan,G.E. (1981) Ion effects on the lac repressor–operator equilibrium. *Biochemistry*, **20**, 3842–3851.
32. Barkley,M.D. (1981) Salt dependence of the kinetics of the lac repressor-operator interaction: role of nonoperator deoxyribonucleic acid in the association reaction. *Biochemistry*, **20**, 3833–3842.
33. Berg,O.G. and von Hippel,P.H. (1985) Diffusion-controlled macromolecular interactions. *Annu. Rev. Biophys. Biophys. Chem.*, **14**, 131–160.
34. Stoylov,S.P., Vuilleumier,C., Stoylova,E., De Rocquigny,H., Roques,B.P., Gerard,D. and Mely,Y. (1997) Ordered aggregation of ribonucleic acids by the human immunodeficiency virus type 1 nucleocapsid protein. *Biopolymers*, **41**, 301–312.
35. Le Cam,E., Coulaud,D., Delain,E., Petitjean,P., Roques,B.P., Gerard,D., Stoylova,E., Vuilleumier,C., Stoylov,S.P. and Mely,Y. (1998) Properties and growth mechanism of the ordered aggregation of a model RNA by the HIV-1 nucleocapsid protein: an electron microscopy investigation. *Biopolymers*, **45**, 217–229.
36. Remy,E., de Rocquigny,H., Petitjean,P., Muriaux,D., Theilleux,V., Paoletti,J. and Roques,B.P. (1998) The annealing of tRNA3Lys to human immunodeficiency virus type 1 primer binding site is critically dependent on the NCp7 zinc fingers structure. *J. Biol. Chem.*, **273**, 4819–4822.

37. Yamada,O., Kraus,G., Sargueil,B., Yu,Q., Burke,J.M. and Wong-Staal,F. (1996) Conservation of a hairpin ribozyme sequence in HIV-1 is required for efficient viral replication. *Virology*, **220**, 361–366.
38. Russell,R.S., Hu,J., Laughrea,M., Wainberg,M.A. and Liang,C. (2002) Deficient dimerization of human immunodeficiency virus type 1 RNA caused by mutations of the u5 RNA sequences. *Virology*, **303**, 152–163.
39. Bohnlein,S., Hauber,J. and Cullen,B.R. (1989) Identification of a U5-specific sequence required for efficient polyadenylation within the human immunodeficiency virus long terminal repeat. *J. Virol.*, **63**, 421–424.
40. Urbaneja,M.A., Wu,M., Casas-Finet,J.R. and Karpel,R.L. (2002) HIV-1 nucleocapsid protein as a nucleic acid chaperone: spectroscopic study of its helix-destabilizing properties, structural binding specificity, and annealing activity. *J. Mol. Biol.*, **318**, 749–764.
41. Tanchou,V., Gabus,C., Rogemond,V. and Darlix,J.L. (1995) Formation of stable and functional HIV-1 nucleoprotein complexes *in vitro*. *J. Mol. Biol.*, **252**, 563–571.