



ELSEVIER

Biophysical Chemistry 87 (2000) 149–165

Biophysical
Chemistry

www.elsevier.nl/locate/bpc

Binding of human immunodeficiency virus type 1 nucleocapsid protein to Ψ -RNA-SL3

Michael F. Shubsda, Christopher A. Kirk, Jerry Goodisman,
James C. Dabrowiak*

Department of Chemistry, Center for Science and Technology, R 1-014, Syracuse University, Syracuse, NY 13224-4100, USA

Received 31 May 2000; accepted 17 July 2000

Abstract

The interaction of the nucleocapsid protein NCp7, from the pNL4-3 isolate of HIV-1, with Ψ -RNA-SL3, with the sequence 5'-GGACUAGCGGAGGCUAGUCC, was studied using non-denaturing gel electrophoresis. Two kinds of experiments were performed, using buffered solutions of radiolabeled RNA and unlabeled protein. In the 'dilution' experiments, the total RNA concentration, R_T , was varied for a series of solutions, but kept equal to the total protein concentration, P_T . In the 'titration' experiments, solutions having R_T constant but with varying P_T were analyzed. The solutions were electrophoresed and the autoradiographic spot intensities, proportional to the amounts of the different species present, were measured. The intensities were fit to a number of equilibrium models, differing in species stoichiometries, by finding the best values of the binding constants. It was shown that NCp7 protein and SL3 RNA combine to form at least two complexes. When P_T is below approximately 10 μ M, a complex that contains two RNAs and one protein forms. Increasing P_T to approximately 100 μ M causes the 2:1 complex to oligomerize, forming a species having eight RNAs and four proteins. For the dilution experiments, run at 5°C at an ionic strength of 31 mM, we found K_1 for the 2:1 complex is $\sim 10^{11} \text{ M}^{-2}$ and K_2 for the 8:4 complex is $\sim 10^{16} \text{ M}^{-3}$. The titration experiments returned $K_1 \sim 10^7 \text{ M}^{-2}$ (poorly determined) and $K_2 \sim 10^{19} \text{ M}^{-3}$. The analysis was complicated by the loss of RNA at higher protein concentrations, due to formation of an insoluble species containing both RNA and protein, which does not enter the gel. Correcting for this changes the calculated values of equilibrium constants, but not the molecularities determined by our analysis. The observation that a small complex can oligomerize to form a larger species is consistent with the fact that NCp7 organizes and condenses the genome in the virus particle. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Nucleocapsid; Electrophoresis; Protein-RNA complex

* Corresponding author. Tel.: +1-315-443-4601; fax: +1-315-443-4070.

E-mail address: jcdabrow@syr.edu (J.C. Dabrowiak).

1. Introduction

The nucleocapsid protein, NCp7, has a number of important functions in the life cycle of HIV [1–4]. The protein facilitates the hybridization of tRNA to HIV RNA so that reverse transcription of the viral genome can take place [5]. It also promotes dimerization of HIV RNA [6–9] by converting an unstable intermediate formed between two genomic length RNAs into a stable duplex. When the virus is budding the correct RNA must be selected from a pool of competing cellular RNAs for insertion into the virus particle [10–13]. Selection appears to involve specific binding of the gag-polyprotein, and its attached nucleocapsid portion, to an RNA sequence, known as the Ψ element or packaging region of the genome.

While certain of the interactions between NCp7 and HIV RNA appear to be structure and/or sequence specific, others are clearly non-specific. For example, there are approximately 2000 NCp7 molecules bound to two copies of genomic length viral RNA [8,12]. This produces a very high binding density strongly suggesting that the interaction between the protein and RNA cannot be sequence specific. Lack of binding specificity is also demonstrated by the observations that NCp7 can interact with non-HIV RNA, DNA and even other proteins [14–19].

The binding of the nucleocapsid protein to HIV RNA has been studied using a variety of physical techniques. Mély et al. [20] studied the binding of a form of the nucleocapsid protein containing 72 amino acids to tRNA^{Lys}. By analyzing the fluorescence quenching data from two tryptophan residues of the protein using Scatchard methods, they found that the nucleocapsid protein bound non-cooperatively to tRNA^{Lys} with $K_{\text{obs}} = 3.2 \times 10^6 \text{ M}^{-1}$ and a binding site size at saturation of six nucleotides. Since similar binding parameters were found for another transfer RNA as well as for 5S RNA from *E. coli*, the authors concluded that binding is not sensitive to RNA structure and sequence. The extended form of the nucleocapsid protein and its RNA complexes form insoluble aggregates that can complicate a binding study [20]. Aggregates involving

tRNA^{Lys} [21] and poly A [22] and the protein have been characterized using electron microscopy.

Sakaguchi et al. [7] used a gel shift assay (electrophoresis in non-denaturing polyacrylamide gels) to study the binding of NCp7 to a 44-nucleotide RNA located in the Ψ -element of the HIV strain, HXB2R. The RNA, which possessed two stem-loops, SL-2 and SL-3 [23] (and references therein), formed two complexes with the protein. Cross-linking experiments showed that both complexes contained one protein cross-linked to one RNA molecule. However, the authors suggested that the slower-moving species in the gel actually contains two RNAs but one of them is blocked from forming a covalent attachment to the protein in the photoinduced cross-linking reaction. Based on their observations the authors proposed that NCp7 binds to the first RNA via direct interaction of one of the zinc fingers of the protein with a stem-loop of RNA. The second complex forms by reaction of the second RNA with the free zinc finger of the 1:1 RNA-protein complex.

Using the SELEX technique, Burglund et al. [24] used NCp7 to select an RNA that has a high binding affinity to the nucleocapsid protein. The selected RNA, SelPsi, possessed sequence/structural features similar to those of stem-loop 3 (SL3) of the Ψ element of HIV. Footprinting and filter binding studies showed that NCp7 binds to the loop region of SelPsi with $K = 4.3 \times 10^8 \text{ M}^{-1}$ but the binding stoichiometry (assumed to be 1:1) was not established in the study. Using electrospray ionization mass spectrometry, Loo et al. [25] found that a 14-mer having the SL3 sequence in HIV-1 (LAI) binds to NCp7 to give a molecular ion corresponding to a 1:1 complex. As expected, removing the zinc ions from NCp7 blocked protein binding to RNA.

Clever et al. [11] used a filter binding assay to measure binding of GST fusion proteins of the full-length gag polyprotein and the nucleocapsid protein to a variety of RNA segments located in the Ψ -region of HIV-1. By assuming a 1:1 binding stoichiometry they found that an RNA segment containing SL3 binds protein with $K \sim 5 \times 10^6 \text{ M}^{-1}$.

A detailed atomic structure of the nucleocapsid

protein bound to HIV RNA determined by NMR was recently published by De Guzman et al. [26]. The structure of a 55-amino acid form of NCp7 bound to SL3 (Fig. 1) shows three of the four bases in the tetra-loop portion of the RNA bound to the zinc fingers of the protein. The stoichiometry of the complex was determined to be 1:1 by gel filtration chromatography and the binding constant, $K \sim 10^7 \text{ M}^{-1}$, was estimated using non-denaturing gel electrophoresis.

Recently, Mély et al. [27] studied the binding of the DNA analogue of SL3 to the 72-amino acid form of NCp7 using fluorescence. They found that one protein binds to one DNA molecule with a binding constant, $K \sim 10^7 \text{ M}^{-1}$. By studying a number of single stranded DNAs and DNA hairpins they found that NCp7 prefers to bind to the following sequences: $X_i\text{TGX}_j > X_i\text{GXGX}_j \sim X_i\text{TXGX}_j > X_i\text{GX}_j \gg X_iX_j$, where X corresponds to either A or C.

In an effort to characterize the Ψ -element of HIV-1 as a possible new target for interdiction by small molecules we have been studying the monomer-dimer equilibrium of sequences from HIV RNA and their interaction with proteins and drugs, using footprinting and quantitative gel

techniques. Previously, we reported monomer-dimer equilibrium constants for RNA sequences located in the dimer initiation site (DIS) of HIV-1 (MAL), measured by non-denaturing gel electrophoresis [28]. We found that equilibrium constants for three RNAs, 19, 27 and 41 bases in length, range from 10^8 to 10^5 M^{-1} , with the longest sequence having the smallest value of K . The decrease in K by three orders of magnitude between the shortest and longest sequence was attributed to changes in entropy of rod-like molecules and increased conformational entropy associated with long RNA sequences.

In this work we use non-denaturing polyacrylamide gel electrophoresis to study the equilibrium and kinetics of binding between SL3 and the 55-amino-acid residue protein, NCp7; this is the same system studied by DeGuzman et al. [26], Fig. 1. By changing the concentration of protein and RNA in a buffered solution and analyzing the species present using non-denaturing gel electrophoresis we show that NCp7 forms at least two complexes with SL3. At protein and RNA concentrations below approximately $10 \mu\text{M}$, a complex forms that contains *two* RNAs and *one* protein. However, increasing the concentration to

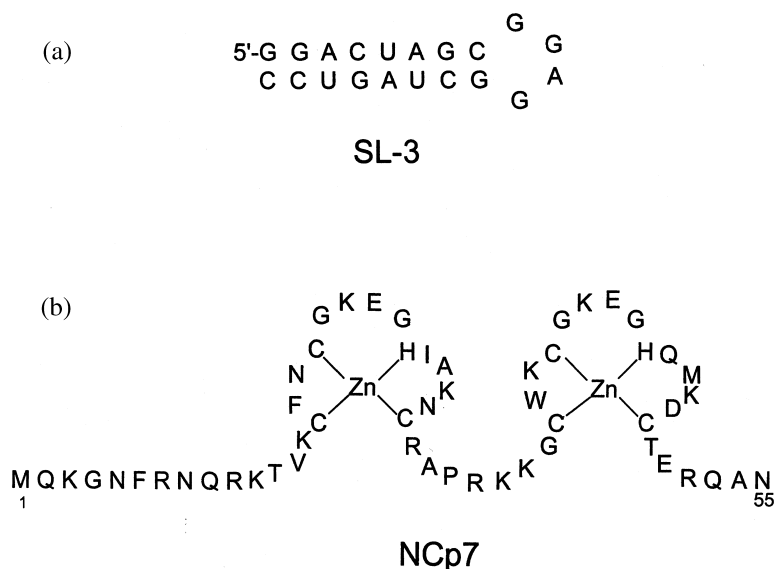


Fig. 1. (a) The sequence of SL3 from the LAI isolate of HIV-1 used in this study. (b) The 55-amino acid residue sequence of NCp7, with the CCHC zinc finger domains indicated, from the pNL4-3 isolate of HIV used in this study.

approximately 100 μM , causes the 2:1 complex to oligomerize to form a species containing eight RNAs and four proteins. Our analysis makes no assumptions about the molecular weights of species in the gel but rather seeks to explain the autoradiographic spot intensities in terms of an equilibrium model involving free RNA and complexes. In this report we present the theory and results for the gel analysis and discuss our findings in light of other binding models and the earlier NMR study on the same system [26].

2. Materials and methods

2.1. Materials

The SL3 RNA (5'-GGACUAGCGGAGG-CUAGUCC) used in this study was synthesized using T7 polymerase [28–30]. To maximize RNA yield and prevent fraying of the stem the coding sequence was modified at the ends from the normal HIV-1 LAI isolate. The molar extinction coefficient at 260 nm used for SL3 is $(1.60 \pm 0.03) \times 10^5 \text{ M}^{-1} \text{ cm}^{-1}$ as measured (five determinations) by previously described methods [28].

The extinction coefficient (ϵ) of the hairpin form of SL3 was determined by complete hydrolysis of the 20-mer [28]. This method, which is not subject to sequence/structural effects that affect other ways of determining ϵ [31], has been used to measure the extinction coefficients of RNA and DNA [32–34]. Since the extinction coefficient was used to measure concentration of SL3, a significant error in this quantity would seriously affect the model used and the constants found in the gel analysis. However, comparison of the value of ϵ for SL3 with those of other RNAs suggests that the value found is consistent with the sequence and structure of SL3 RNA.

The NCp7 with the 55-amino acid residue sequence: M Q K G N F R N Q R K T V K C F N C G K E G H I A K N C R A P R K K G C W K C G K E G H Q M K D C T E R Q A N, from the pNL4-3 isolate of HIV, was supplied by Dr M. Summers, University of Maryland-Baltimore County. The concentration of the protein was

determined optically using an extinction coefficient of $6050 \text{ M}^{-1} \text{ cm}^{-1}$ at 280 nm [35].

2.2. Labeling RNA

The SL3 RNA was labeled with ^{32}P by first removing the 5' phosphate with calf intestinal phosphatase, CIP (Boehringer Mannheim), in 60 μl of CIP buffer containing $\sim 5 \mu\text{M}$ RNA and one unit of CIP [36]. The reaction was run for 1 h at 50°C followed by removal of the enzyme by phenol extraction and purification of the sample by ethanol precipitation. RNA was end-labeled with 20 units of T4 polynucleotide kinase (Promega) in kinase buffer, and 50 μCi of $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ (Amersham). The labeling solution was incubated at 37°C for 1 h and the reaction was stopped by ethanol precipitation. Labeled RNA was purified using denaturing PAGE in $1 \times \text{TBE}$ running buffer (90 mM Tris-borate, 2 mM EDTA, pH 8). RNA was visualized by autoradiography, the band containing the sample excised and RNA recovered by suspension of the gel segment in 1 ml of $1 \times \text{TBE}$ buffer. A 5- μl sample of the RNA recovered by this method yielded $\sim 60\,000$ cpm upon scintillation counting.

2.3. Protein binding studies

Two different types of gel shift assays were done. One type referred to as a 'dilution experiment' was done by taking a concentrated sample of unlabeled RNA, 510 μM , in buffer (25 mM acetate, 0.1 mM ZnCl_2 , 0.1 mM β -mercaptoethanol, pH 6.5) and mixing with an equal molar amount of NCp7. Prior to adding the protein, RNA secondary structure was removed by heating to 100°C for 10 min followed by rapid cooling to 0°C . The equimolar solution was diluted with buffer to prepare samples spanning a wide range of concentration (0.6–510 μM). A 12000 cpm-labeled RNA aliquot that had also been treated to remove secondary structure described above was added to each sample and the samples were equilibrated at 5°C for at least 36 h. Immediately prior to electrophoresis, each sample was diluted by 0.25 by addition of a 50%

water/glycerol mixture at 5°C, briefly vortexed, and loaded on a 20% non-denaturing gel at 5°C. Electrophoresis was for ~16 h at 1000 V in 1× TBE buffer. After electrophoresis, the gel was dried and autoradiography carried out at –20°C using Kodak X-omat film. If the maximum optical density of images exceeded ~1.1 O.D., i.e. the optical density above which the response of the film is non-linear, a second autoradiogram with a shorter exposure time would be made.

In the experiments referred to as ‘titration experiments’ a number of solutions having the same amount of labeled and unlabeled RNA (total RNA concentration of 90 μM) were prepared in buffer and subjected to heating/cooling to remove RNA secondary structure as described above. Protein was then added to give each sample a different concentration of protein (0–90 μM). Samples were equilibrated, usually at 5°C for ~36 h, after which time they were diluted by addition of water/glycerol, briefly vortexed and subjected to electrophoresis.

2.4. Quantitation of gel data

The autoradiograms were digitized using a Hewlett Packard Scan Jet 5p scanner and Sigma Scan software, yielding an intensity profile for each lane of the scanned image. The intensity profiles were fitted to a sum of Gaussian peaks and a linear background using Peakfit software. In the few cases where the optical density of the image exceeded 1.1, a peak was fitted using only data points below 1.1. Peak areas were assigned to the unbound RNA and the two complexes (the faster-running complex appeared as a streak). When bands overlapped, appropriate weighting factors were used to find band areas. The areas corresponding to unbound RNA and the complexes were divided by the total intensity in each lane to give the fractional amounts of each species. The fractional amounts were used in the calculations.

3. Theory

The electrophoretic field causes negatively

charged species in a sample placed in the gel well to migrate out of the well and into the gel proper. In the gel, these species are separated according to their mobilities, and the locations and amounts of radiolabeled species are determined from an autoradiogram of the gel.

When the species enter the gel from the loading well and are separated according to their mobilities, they are no longer in equilibrium since, in a region of the gel where one species is present, the concentrations of other species are close to zero. The absence of equilibrium may lead to decomposition or combination of species during migration in the gel. Only if the rate constants are small enough so that interconversion of species in the gel is negligible, will the separated species appear as spots in different locations on the gel, with spot intensities proportional to the amounts of these species which were present in the gel well. If interconversion is rapid, so that one detectable species converts rapidly into another of different mobility, the interconverting mixture appears as a streak.

Cann [37–39] has shown that the sizes and shapes of spots and streaks in an electrophoretic gel may be understood by solving, numerically, the differential equations for motion of species in the gel (including mobility, diffusion, and reaction). From the concentrations of different species as a function of time and position, one can generate intensity plots (intensity as a function of position in a gel lane) which may be compared to the experiment. We and others [28,40,41] have carried out this program for DNA and RNA, obtaining spots and streaks which resemble those found experimentally. In particular, if species A reacts in the gel to form species B, the streak will extend from the position of A to the position of B, increasing in intensity from the position of A to that of B. If both $A \rightarrow B$ and $B \rightarrow A$ occur, the two streaks may combine to give a broad spot with maximum intensity between A and B.

Observation of a narrow spot indicates a species which is not reacting in the gel, but a broad spot or streak does not necessarily indicate interconversion between species. If a species can exist in a number of different conformations with slightly different mobilities, it may give rise to a number

of close-together spots on the gel. These may appear to form one broad spot or streak, whether or not interconversion between conformations is taking place in the gel. It should be noted that rate constants for interconversion in the gel are often found to be several orders of magnitude smaller than those in bulk solution [28,41].

In the present system, the negatively charged SL3 RNA migrates from the top of the gel to the bottom when the electrophoretic field is on. The positively charged free NCp7, also present in the solution loaded into the well, does not enter the gel. Whether the charge of a protein-RNA complex is positive or negative depends on the relative numbers of RNAs and proteins in the complex. However, since the anionic charge on SL3 is much larger in magnitude than the cationic charge on NCp7, a complex formed between the two will most likely be negative. This is confirmed by our ability to detect NCp7-SL3 complexes in the gel. Since the mobility of a species depends on its charge, the RNA-protein complexes will certainly move more slowly in the gel than unbound RNA. We expect larger complexes to have lower mobilities because of their size, but it is also possible that the combination of negatively-charged and positively-charged species leads to more compact structures with enhanced mobilities in the gel.

It is possible for an NCp7-SL3 complex to dissociate to free protein and free RNA in the gel, leading to formation of a streak increasing in intensity from the position of the complex to the position of the free RNA. This is because more RNA is dissociated from complex earlier, and the RNA dissociated later lags further behind the free RNA entering the well. On the other hand, an NCp7-SL3 complex cannot form in the gel, because there is no free protein available — any protein formed by decomposition of a complex migrates in the direction opposite to that of RNA. Small complexes can combine into larger ones, according to $nA \rightarrow A_n$, since there is an excess of A in one region of the gel; this would lead to a streak increasing in intensity going backward (opposite to the direction of migration) from the position of A, since A_n has a lower mobility than A. The decomposition, $A_n \rightarrow nA$, can also occur, leading to a streak increasing in intensity going

forward (in the direction of migration) from the position of A_n .

In our analysis, it is assumed that the intensity of a spot (or the integrated intensity of a streak) is proportional to the number of radiolabeled SL3 RNAs at the spot's location. This means that, for a complex containing m RNAs, the ratio of the spot intensity to the number of such complexes present is m times the corresponding ratio for a complex containing a single RNA. Since a broad spot or streak is a feature formed from reaction in the gel, its intensity can be assigned to a species from which it originated and contributes to the species concentration in the well.

The actual concentrations of radiolabeled species are never known, so we cannot calculate concentrations from spot intensities. Only fractional intensities, spot intensities divided by the sum of all spot intensities in each lane, are significant. Our procedure is to assume a model, corresponding to an assumed stoichiometry of the complexes involved, and calculate, from the model, fractional intensities for different bands appearing on the autoradiogram. These are then compared with the measured fractional intensities. For models which show promise, calculations are performed for various values of the equilibrium constants, to find the best fit. Goodness of fit is defined by the average squared deviation between calculated and measured fractional intensities.

As shown in Figs. 2–4, our gels show the existence of more than one NCp7-SL3 complex, in addition to free SL3. Assuming two such complexes, with empirical formulas $P_i R_j$ and $P_m R_n$, we have two equilibrium expressions:

$$\frac{[P_i R_j]}{[P]^i [R]^j} = K_1 \quad \text{and} \quad \frac{[P_m R_n]}{[P]^m [R]^n} = K_2 \quad (1)$$

Here, brackets indicate concentrations, P represents free NCp7, R represents free SL3 RNA, and K_1 and K_2 are the equilibrium constants. There are also two conservation relations,

$$[P] + i[P_i R_j] + m[P_m R_n] = P_T \quad (2a)$$

$$[R] + j[P_i R_j] + n[P_m R_n] = R_T \quad (2b)$$

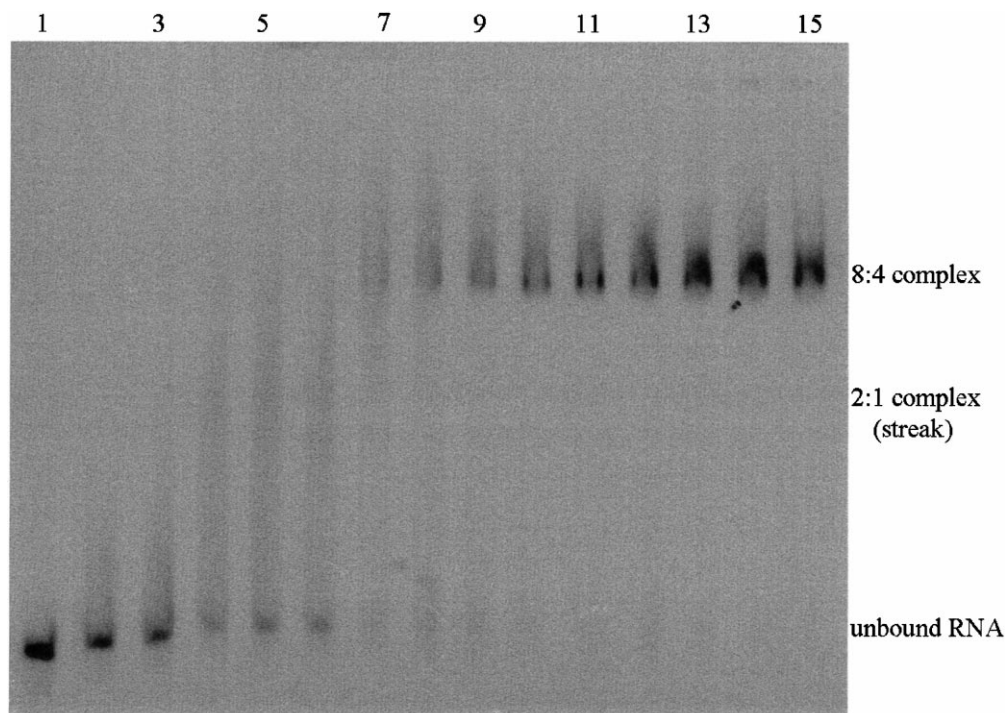


Fig. 2. The autoradiogram of a typical dilution experiment. The direction of migration is from top to bottom. The fastest running species is the unbound RNA, the slowest running species is the 8:4 complex and the streak between the two is the 2:1 complex. The concentration of RNA is the same as the concentration of protein for each lane except for lane 1, in which only radiolabeled RNA is present. This is a control to show the position of unbound RNA. The concentrations in μM and the lane numbers (in parentheses) are: 0.41 (2), 0.71 (3), 1.2 (4), 2.0 (5), 3.5 (6), 5.7 (7), 9.8 (8), 17 (9), 29 (10), 50 (11), 83 (12), 135 (13), 233 (14), 383 (15).

in which P_T and R_T are the known total concentrations of NCp7 and SL3. Using the equilibrium expressions to substitute for $[P_i R_j]$ and $[P_m R_n]$, one obtains two equations for $[P]$ and $[R]$. For given values of K_1 , K_2 , P_T and R_T , these equations are solved by the Newton–Raphson method [42] in two dimensions.

Thus, for each set of values of P_T and R_T one obtains $[P]$ and $[R]$ and, using the equilibrium constant relations, $[P_i R_j]$ and $[P_m R_n]$. Then fractional intensities for free RNA and the two complexes are calculated according to:

$$f_R = \frac{[R]}{[R] + j[P_i R_j] + n[P_m R_n]} \quad (3a)$$

$$f_1 = \frac{j[P_i R_j]}{[R] + j[P_i R_j] + n[P_m R_n]} \quad (3b)$$

and

$$f_2 = \frac{n[P_m R_n]}{[R] + j[P_i R_j] + n[P_m R_n]} \quad (3c)$$

One then has fractional intensities of all three radiolabeled species as a function of concentration.

To study the binding of NCp7 to SL3 two types of experiments were carried out. In dilution experiments, different lanes of the gel correspond to different dilutions of a solution containing equal concentrations of NCp7 and SL3 RNA, so that each lane has different values of P_T and R_T but $P_T = R_T$ in all lanes. In titration experiments, R_T is the same for all lanes of the gel, and P_T varies from 0 to approximately R_T . Titration experiments were also attempted with P_T fixed and

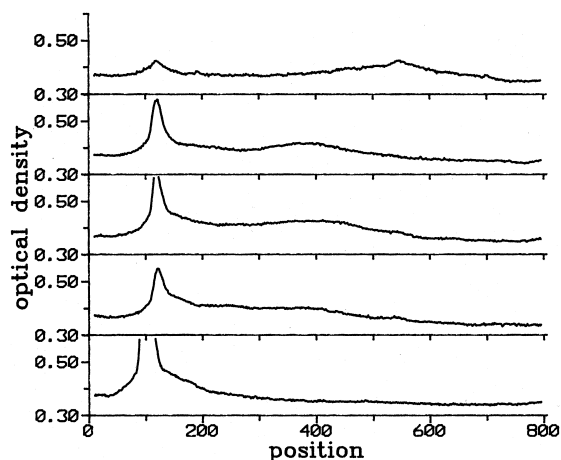


Fig. 3. Optical density as a function of position (arbitrary units) for lanes (from bottom to top): 2, 4, 5, 6 and 7 from the autoradiogram shown in Fig. 2. The direction of migration is from high to low position numbers.

R_T varying from lane to lane, but were unsuccessful. When P_T was much greater than R_T , the radioactivity remained either in the gel well or the reaction tube, suggesting that the protein

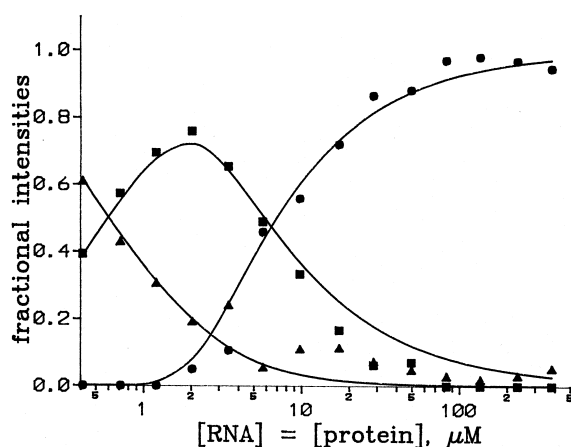


Fig. 4. Plot of the fractional intensities of the species present in each lane of the autoradiogram in Fig. 2 vs. the concentration of RNA or NCp7. Triangles represent the unbound RNA, squares represent the 2:1 complex, and circles represent the 8:4 complex. The solid lines are the calculated fractional intensities with $K_1 = 3.7 \times 10^{12} \text{ M}^{-2}$ and $K_2 = 7.1 \times 10^{16} \text{ M}^{-3}$.

bound the RNA into species which were not soluble, had no charge or were positively charged.

4. Results and analysis

4.1. Temperature of electrophoresis. Stability of species in the gel

In order to determine the equilibration and gel running temperatures to be used in the study, a series of trial electrophoresis experiments was carried out. Two solutions, one containing 103 μM and one containing 91.3 μM NCp7, and both containing 125 μM SL3, were equilibrated at 5°C for 9 days after which a ‘load and run’ experiment [41] was carried out. Aliquots of solution were loaded into a non-denaturing gel with a running temperature of $\sim 35^\circ\text{C}$ at various times after the current was turned on. In this way, the behavior of species as a function of time spent in the gel could be studied. Plots of band intensity for the slowest migrating complex vs. migration distance (time) clearly showed that the concentration of this complex decreases as electrophoresis takes place at $\sim 35^\circ\text{C}$. This shows that the complex decomposes in the gel as electrophoresis takes place. However, when electrophoresis was carried out in a gel with a running temperature of 5°C, no decomposition took place. On the basis of these observations it was decided to carry out all electrophoresis separations at 5°C to minimize decomposition of the complex in the gel.

4.2. Equilibrium / aggregation

To determine if the system reached equilibrium prior to electrophoresis, selected samples were allowed to stand at 5°C for 2–9 days followed by electrophoresis in a 5°C gel. The band patterns were identical (data not shown), indicating that samples reached equilibrium at 5°C in less than two days. This raises the possibility that the system reaches equilibrium in the loading well of the gel during the time (10–15 min) that it takes for the species to enter the gel proper. If this occurs, the sample is diluted in the well, changing its

concentration by a factor $d = V_i/V_w$ where V_i and V_w are the volumes of the sample and well respectively [28]. This disturbs the equilibrium. Then all initial concentrations must be divided by d before calculating binding constants from the spot intensities [28].

Samples of identical composition were equilibrated at two different temperatures, 5°C and 35°C, before being loaded into a 5°C gel for electrophoresis. The resulting band patterns were not the same (data not shown), reflecting the fact that the equilibrium distribution of species at 5°C is different from that at 35°C. Thus, the equilibration time for the system appears greater than the residence time in the well, making a dilution correction unnecessary. Also, loading different volumes of the same NCp7/SL3 solution, equilibrated at 5°C, into a gel at 5°C gave the same relative band intensities and patterns, again indicating that dilution in the well is not important. Therefore, no correction for dilution in the well was used in calculating the binding constants reported below. It should be noted that our calculations show that, while the magnitude of K changes if dilution is considered, the molecularities of the species involved does not.

Previous work by Mély et al. [20] and Al-Ghusein et al. [21] with a 72-amino acid variant of NCp7 indicated that the protein and its RNA complexes form insoluble aggregates. The 55-residue nucleocapsid protein used in these studies is missing the cationic ‘tail’ from the carboxylic acid terminus of the longer variant. How this structural change relates to the nucleation properties of the protein is difficult to predict but Summers and co-workers [26,43,44], who have extensively characterized the 55-residue variant using high resolution NMR, have not reported that the protein itself aggregates or associates in solution. In our analysis we have assumed that interactions between free protein molecules in solution do not take place. However, evidence that the NCp7-SL3 complexes may aggregate to produce insoluble materials may lie in the fact that, for our experiments, the total intensity in a lane of a gel usually decreased as the concentration of NCp7/SL3 increased (data not shown). This indicates that some of the sample may be in a form

which does not allow it to enter the gel. However, although the change in total intensity can be large, we found that the *fractional spot intensities* are unaffected; all of the experiments support the proposed model and return comparable values of K .

4.3. Results of experiments

An autoradiogram for a dilution experiment is shown in Fig. 2. Scans of selected lanes are shown in Fig. 3. In all the dilution experiments, the gels show similar band patterns. Lanes for low concentrations of SL3 and NCp7 show a single band with the same mobility as the band in the lane for SL3 only, so this band must be unbound RNA. The high-concentration lanes show a narrow band with a considerably lower mobility than the unbound RNA band, which is assigned as a stable protein-RNA complex. (Sometimes the band for unbound RNA is also present, but with low intensity.) At intermediate concentrations, one sees a third feature, a broad band or streak between the unbound RNA and the slow moving complex, which is assigned as an intermediate between the stable complex and unbound RNA. The average electrophoretic mobility (position) of the streak *decreases* as the concentration of NCp7/SL3 *increases*, Fig. 3. Fig. 4 gives the fractional intensities of the free RNA, the streak and the slow moving complex as a function of protein/RNA concentration.

Various models of SL3 binding to NCp7 were used to calculate fractional intensities for comparison with experiment. At first, the streak was ignored, or considered to represent the dissociation of the complex in the gel, and only the slowest band and the band due to unbound RNA were considered. It was impossible to fit the fractional intensities as a function of concentration by assuming that the slow moving SL3:NCp7 complex was 1:1, 2:1, or 1:2. Reflection shows that the streak cannot be due to the decomposition of the slowest moving complex in the gel because, if it were, it would still be present at the highest concentrations, when only the slowest moving complex is present. Furthermore, the fact that the intensity of the streak *decreases* as the intensity of

the slowest moving species *increases* (Fig. 4) shows that both features represent complexes of SL3 and NCp7, and that the complex appearing at highest concentration has a larger molecularity.

Subsequent models included two different complexes involving SL3 and NCp7. The streak was assigned as a complex, $P_i R_j$, and the slowest migrating species as $P_m R_n$, where $m + n > i + j$. Although it seems reasonable for the larger complex to be a multiple or aggregate of the smaller, models for which this was not the case were also considered. For example, the simplest models assume SL3:NCp7 = 1:1 for the streak and 2:1 or 1:2 for the large complex. It is impossible, however, to fit the fractional intensities with these models. If the streak is assumed to be 1:1 complex and the slowest moving band 2:1, the calculated intensity of the former does not decrease as concentration increases. If the slow band is assumed to be 1:2 (one RNA and two proteins), one cannot explain why the amount of free RNA is reduced almost to zero at high concentrations. If the streak is assumed to be 1:2 and the slow band is due to a larger complex, one cannot come close to fitting the fractional intensities. It is necessary to model the streak as a 2:1 complex.

With a 2:1 complex for the streak, we assumed that the slowest moving band is a 4:2 complex. This model shows the experimentally observed decrease in the 2:1 complex intensity at high concentrations, but the calculated intensity of the 4:2 complex does not increase fast enough with concentration to match the experimental data. Assuming that this complex is 8:4, i.e. formed from four of the smaller complexes, corrects the error. Although 16:8 works as well as 8:4, the choice of 8:4 is more reasonable. After trying other models with different molecularities for both complexes, we concluded that the small complex (streak) must involve two RNAs plus one protein, and that the large complex (slowest migrating band) probably contains eight RNAs and four proteins.

No equilibrium model can explain the persistence of the free-RNA spot at high concentrations. Our explanation for this is that some of the RNA is incapable of binding to protein. The amount of 'non-binding' RNA can be estimated

from the free-RNA spot intensity at the highest concentrations (it eventually levels off with concentration), and the data can be corrected by subtracting the non-binding RNA from the free-RNA spot intensity at each concentration. On fitting the corrected data to models, the conclusions about the molecularities of the complexes are unchanged, and in the values of the equilibrium constants change little. There is a small improvement in the goodness of fit of calculated fractional intensities to experimental values.

The proposed model raises several questions: (1) Why does the small complex appear as a streak and not a narrow band? (2) Why are complexes intermediate between 2:1 and 8:4 not seen? (3) Why does the position of maximum intensity in the streak move, as concentration increases, in the direction away from the free-RNA band toward the band for the large complex? All can be answered by postulating that the streak is formed from bands representing two or more species that have different mobilities and different molecularities, such as the 2:1 and 4:2 complexes. The relative intensities of the species would vary with concentration. Alternatively, the streak may represent an interconverting mixture of these complexes, in which case simulations show [41] that the individual bands would be replaced by a single broad band with maximum intensity at a position dependent on the relative amounts of the species. In view of the breadth of the streak, the latter possibility is favored.

A three-complex model was considered, involving, in addition to K_1 and K_2 , the equilibrium constant K_3 , corresponding to formation of a complex $P_r R_s$,

$$\frac{[P_r R_s]}{[P]^r [R]^s} = K_3 \quad (4)$$

The fractional intensity of the streak was calculated as

$$f_3 = \frac{j[P_i R_j] + r[P_r R_s]}{[R] + j[P_i R_j] + n[P_m R_n] + r[P_r R_s]} \quad (5)$$

Since this model includes an extra complex, it

necessarily improves the fit of calculated fractional intensities to experiment from three additional parameters, r , s , and K_3 , but the improvement is not significant. Also, one can change K_3 substantially without significantly changing the goodness of fit, provided that K_1 is also changed.

Therefore, two-complex models are used to interpret all of the dilution-gel data, with the understanding that the value of K_1 is an effective value. With the complex P_iR_j corresponding to the broad peak and the complex P_mR_n to the narrow, slowest-running, peak, it is found that $i = 1$, $j = 2$, $m = 4$, $n = 8$ works best. Calculated fractional intensities derived from this model are shown in Fig. 4.

The total intensity in a given lane was plotted vs. RNA concentration for this first dilution experiment (of seven carried out). The trend line had a slope of zero within the calculated precision, indicating that the same amount of RNA entered the gel from all samples in the dilution experiment. However, only two of the seven dilution experiments gave plots of total intensity vs. concentration that showed no trend, the other five experiments showing significant decreases in total intensity (up to 90%) with increasing concentration. This indicated loss of RNA. Since it occurred only when both RNA and protein were present, the material being lost was an insoluble protein-RNA complex. It remained either in the reaction tube (confirmed by scintillation count-

ing) or in the loading well, and did not enter the gel proper. Since the *relative* spot intensities for all dilution experiments behaved in the same way, the solution above the insoluble complex contained the *equilibrium* distribution of species.

However, the loss of RNA and protein led to a decrease in the total RNA (R_T) and total protein (P_T) concentrations in solution from their initial values, which were used to calculate free RNA and free protein in the equilibrium expressions. Therefore, it was necessary to decrease these quantities by the amounts of protein and RNA lost to the material that did not enter the gel. This was done by fitting the plot of total intensity vs. concentration to an exponential and generating a correction factor $b_k = I_k/I_T$, where I_k is the total intensity for lane k calculated from the exponential and I_T is the calculated intensity for $R_T = 0.1 \mu\text{M}$ (estimated concentration of RNA for the lane with no unlabeled RNA and no protein). The corrected total RNA concentration for lane k , R_T' , is given by $R_T' = b_k R_T$ and the corrected total protein concentration for lane k , P_T' , is given by $P_T' = 1/2(b_k + 1)P_T$. The latter correction assumes that the complex responsible for the loss of RNA contains RNA and protein in the ratio 2:1, the ratio found for the small complexes. Other ratios were tested for use in this correction, such as 1:1, which was found to remove too much of the RNA for the analysis of

Table 1
Binding constants obtained from dilution experiments^a

Run	[SL3] = [NCp7] (μM) ^b	K_1 (M^{-2}) ^c	K_2 (M^{-3}) ^c	rms dev. ^d	% decrease ^e
1	0.42–383	3.7×10^{12}	7.1×10^{16}	0.052	~ 0
2	0.71–233	1.5×10^{11}	2.7×10^{14}	0.11	~ 0
3	0.25–230	9.9×10^{10}	5.5×10^{15}	0.13	62
4	0.6–45	6.0×10^{11}	3.1×10^{16}	0.092	84
5	0.6–45	1.2×10^{11}	1.0×10^{17}	0.12	90
6	0.6–45	1.3×10^{11}	1.5×10^{16}	0.079	64
7	0.6–45	3.2×10^{11}	9.5×10^{16}	0.10	64

^aSL3 RNA Interacting with the nucleocapsid protein, NCp7.

^bThe concentrations have not been corrected for the loss of RNA as described in the text.

^cTemperature = 5°C, ionic strength = 28 mM.

^dRoot-mean-square deviation between observed and calculated fractional intensities.

^eThe percent of total intensity lost at the highest concentrations of protein and RNA. The values of K_1 and K_2 given have been obtained using RNA and protein concentrations that have been corrected for this loss (see text).

the gels to work. It should be emphasized that, although the correction improves agreement between values of K from different experiments, it is unimportant for establishing the species molecularities: without the correction all experiments returned the same molecularities. The values of K for the various dilution experiments using this correction are found in Table 1.

We now consider the titration gels. Fig. 5 shows an autoradiogram of such a gel. In Fig. 6, fractional intensities of the three observed spots are plotted as a function of protein concentration (RNA concentration is fixed). At low protein concentrations, only the spot for free RNA appears, as expected. For protein concentrations approaching the RNA concentration, one sees only a narrow spot, at the position of the large complex. The broad spot between the two other spots appears with very low intensity, and on some titration gels it is not observed at all.

Attempts were made to fit the data from a titration gel to a one-complex model, ignoring whatever intensity appeared between the two

spots. The assumption that the complex was 1:1 led to fractional intensities that were far from matching those experimentally observed. Assuming a model of *two* RNAs to *one* protein led to much better agreement, as did the 4:2 and 8:4 models. In fact, there was not much difference in the fits obtained from these three models.

The two-complex model can also fit the titration gel data, and with the same molecularities as used for the dilution gel data. Models with other molecularities were tried, but, as for the dilution gels, they could not fit the data at all. The best-fit values of the binding constants K_1 and K_2 for the titration experiments using the two-complex model are given in Table 2. They are substantially different from those determined by fitting the dilution gels. In particular, K_1 must be much smaller for the titration gels than for the dilution gels in order to predict the very low intensities of the small complex which are observed. Because these intensities are so small, a precise value for K_1 cannot be determined, and this causes the value of K_2 to be less precise as well.

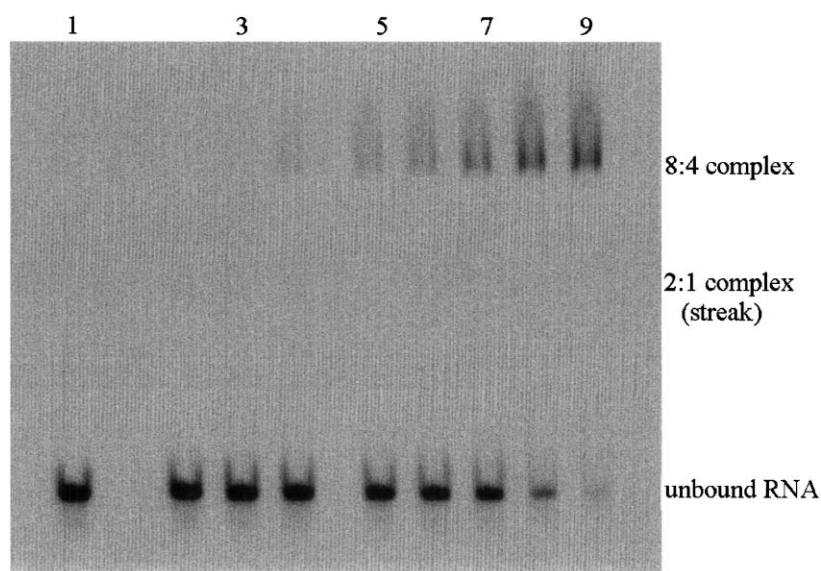


Fig. 5. The autoradiogram of a titration experiment. The direction of migration is from the top to the bottom. The fastest moving spot is unbound RNA and the slowest moving spot is the 8:4 complex. The 2:1 complex, while difficult to observe, appears as a streak between the two major species. The RNA concentration is 68 μM for all lanes and the NCp7 concentrations, in μM , increasing from left to right are (lane numbers in parentheses): 0 (1), 0.75 (2), 7.5 (3), 15 (4), 23 (5), 30 (6), 38 (7), 53 (8), and 68 (9). The concentrations have not been corrected for the 25% loss of RNA and NCp7 as described in the text.

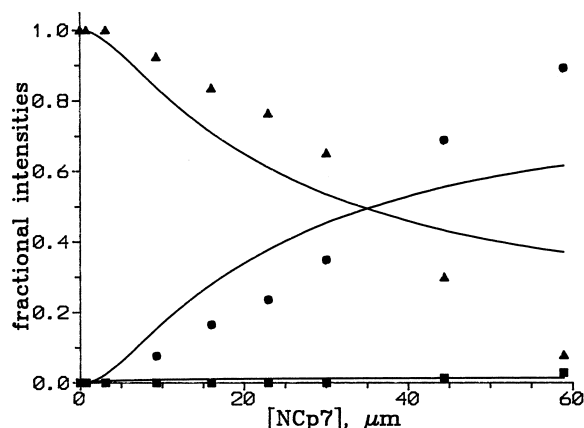


Fig. 6. The fractional intensities of the species shown in the autoradiogram in Fig. 5 plotted as a function of the NCp7 concentration, after correcting for the loss of RNA and NCp7. Triangles represent the unbound RNA, circles represent the 8:4 complex, and squares represent the 2:1 complex. The lines show calculated fractional intensities for $K_1 = 2.5 \times 10^7 \text{ M}^{-2}$ and $K_2 = 2.1 \times 10^{20} \text{ M}^{-3}$.

Of the three titration experiments, one showed no decrease in total intensity in a lane with protein concentration. For the others, an exponential fit of total intensity vs. P_T was used to correct for loss of RNA and protein. As for the dilution gels, R_T for lane k was multiplied by $b_k = I_k/I_T$ where I_k is the total intensity calculated for the concentration of lane k and I_T is the total intensity calculated for the no-protein lane. The corrected total protein concentration was $P_T' = 1/2(b_k + 1)P_T$.

5. Discussion

We have shown how analysis of electrophoretic gels can be used to obtain information about the binding of the SL3 segment of HIV RNA to the nucleocapsid protein NCp7. Radiolabeled RNA is allowed to come to equilibrium with NCp7 in solution and samples are electrophoresed in a non-denaturing polyacrylamide gel to separate species of different mobilities. Two kinds of experiments were performed: in dilution gels, different lanes correspond to solutions with different protein and RNA concentrations, but with $P_T = R_T$; in titration gels, R_T is the same in all lanes but P_T varies from lane to lane, with $P_T \leq R_T$.

Optical density plots derived from measurements of an autoradiogram of the gel show a number of features, corresponding to different RNA-containing species. The intensities of these features are proportional to the relative amounts of these species in the gel. If reaction kinetics in the gel are slow, these reflect the amounts of the species in the solution loaded into the gel wells. Assuming a model (chemical formula) for the species in solution, one calculates spot intensities and tries to fit them to those observed by varying the values of equilibrium constants. Since most models do not permit a good fit, only a few models are acceptable.

In all gels, one observes two well-defined bands, one of which can be unambiguously assigned to unbound RNA and the other to a RNA-protein

Table 2
Binding constants obtained from titration experiments^a

Run	[SL3] (μM) ^b	[NCp7] (μM) ^b	K_1 (M^{-2}) ^c	K_2 (M^{-3}) ^c	rms dev. ^d	% decrease ^e
1	136	0–103	5.0×10^6	8.0×10^{17}	0.060	15
2	68	0–68	2.5×10^7	2.1×10^{20}	0.11	25
3	68	0–68	5.0×10^7	8.4×10^{19}	0.18	~ 0

^aSL3 RNA Interacting with the Nucleocapsid Protein, NCp7.

^bThe concentrations given have not been corrected for the precipitation phenomenon described in the text.

^cTemperature = 5°C, ionic strength = 55 mM.

^dRoot-mean-square deviation between observed and calculated fractional intensities.

^eThe percent of total intensity lost at the highest protein concentration. The values of K_1 and K_2 given have been calculated using total RNA and protein concentrations that have been corrected for loss (see text).

complex. In dilution gels, and in some titration gels, one finds a broad band between these two. Due to the way its intensity changes with protein and RNA concentration, it must represent one or more RNA-protein complexes of smaller molecularity than that responsible for the slowest-moving band. That fact that the intermediate band is broad suggests that it involves several such complexes, possibly interconverting in the gel.

Models involving three or four complexes of different molecularities were considered. Of course, the additional parameters involved allow these models to give better fits to experimental data, but the improvement is not enough to be significant. Thus, we considered two-complex models, with the understanding that K_1 and K_2 represent effective or average values, as do the molecularities of the complexes involved. We assumed that the faster-moving complex, represented by the broad streak, involved fewer protein and RNA molecules than the slower-moving one.

The simplest choice for the former is 1:1, but no model incorporating this assumption could fit intensities from any of our gels. It was necessary to assume two RNAs plus one protein to get a good fit. Although many choices were tried for the slower complex, the best fits were found assuming it was some multiple of the faster complex, with the best of these being eight RNAs plus four proteins. This fit all the dilution and titration gels, although the values found for the equilibrium constants K_1 and K_2 were different for gels of different types. This may be related to the fact that the low intensities found for the broad band on titration gels made it impossible to determine K_1 with any accuracy, and this in turn made the determination of K_2 difficult. Values of the equilibrium constants as determined from various gels are given in Tables 1 and 2.

A problem is that species too large to enter the gel may form in concentrated solutions. This is probably the reason that total intensity in a lane decreases with increasing concentration of protein or RNA in some dilution gels. It may also explain why it was not possible to run RNA titration gels, with protein concentration fixed and that of RNA varied; lanes with P_T greater

than R_T showed little or no intensity, indicating loss of most of the RNA.

When total intensity decreased with RNA and protein concentration, we corrected the values of R_T and P_T , assuming RNA and protein were lost in a 2:1 ratio. It should be emphasized that we obtain the same molecularities for the complexes whether or not a correction for loss of RNA is made. The values of the constants K_1 and K_2 are, of course, changed when the correction is made. The large difference between the values of these constants determined from dilution and titration gels may come from the assumption that the amount of lost protein was half the amount of lost RNA. Assuming instead that significantly more protein is lost would affect the calculations for the titration gels much more than for the dilution gels.

The interaction of NCp7 with SL3 was previously studied using NMR [26]. The system was analyzed as a 1:1 complex between protein and RNA, although the stoichiometry of the complex was not directly determined from the NMR data. Rather, stoichiometry was found by separating the species in the NMR sample using gel filtration chromatography and assigning molecular weights by comparing mobilities of species in the sample with those of standards having known molecular weights.

The temperatures used in this study and those employed in the NMR investigation were not the same. Our experiments were mostly carried out at 5°C, a temperature lower than that used in the NMR work. We found that the 8:4 complex (SL3:NCp7) exists at concentrations approaching those used in the NMR work but that it slowly decomposes in the gel if the running temperature is 37°C. Its stability in bulk solution should be much lower than in the gel, so it may be that the NMR study analyzed lower molecular weight species (giving narrow NMR lines) which were in exchange with the 8:4 complex. It should also be noted that the buffer conditions for the NMR and the gel shift experiments are different. The equilibration buffer for the gel shift experiments lacks the 25 mM NaCl used in the NMR experiments, which means the ionic strength is lower. Also, the

NMR sample is equilibrated and observed in the same buffer, but the gel sample is able to mix with $1 \times$ TBE in the gel well after equilibration. In addition to changing the ionic strength, the TBE buffer may influence the equilibrium [45].

We inferred the formation of insoluble species at the higher concentrations of most of our gel studies. Different preparations of RNA were used in different studies, but the presence of insoluble material did not correlate with the source or age of SL3. The situation may be similar to supersaturation of a solution in that the formation of a solid may be thermodynamically favored but the kinetics for formation are influenced by environmental factors that are difficult to control. Since the authors of the NMR work [26] did not report observing insoluble material, it may be that precipitation was favored by the lower temperature of our gel studies.

The results of our study also contrast with those recently published by Mély and coworkers [27] who used fluorescence to study the binding of the 72-amino-acid variant of NCp7. These investigators reported stoichiometries, binding constants and specificities of interactions of the DNA analogue of SL3, other DNA hairpins, and various single-stranded oligomers. They measured the fluorescence intensity of two tryptophan residues of the protein as nucleic acid was added to the system. Analysis of the fluorescence data using Scatchard methods showed that SL3 DNA reacts with NCp7 to form a 1:1 complex with a binding constant of $\sim 10^7 \text{ M}^{-1}$. That fact that a different system (DNA and the 72-amino acid variant of NCp7) and different experimental conditions (20°C) were used may explain the discrepancies between our work and this one [27].

Another recent fluorescence study by Urbaneja et al. [46] leads to conclusions contradicting both Mély et al. [27] and us. This study shows a 1:2 RNA/protein stoichiometry with a dissociation constant K_d of 10 nM in sodium phosphate at pH 7.0. They also suggest that Trp-37 is part of the binding site because of the $\sim 90\%$ tryptophan fluorescence quenching when nucleic acid was bound. The binding is suggested to be in the loop of the SL3 hairpin since DNA hairpin oligonu-

cleotides without loop bases bound with 1:1 stoichiometry and $K_d = 50 \text{ nM}$.

The structure of SL3 indicates that it can undergo a hairpin-duplex interconversion similarly to RNA sequences earlier studied by us [28]. Since the loop region of SL3 is not self-complementary, the bases in the loop, 5'-GGAG, would be expected to form a bulge in the duplex. Lack of base pairing in this region would destabilize the duplex compared to other RNAs with a self-complementary loop region in the monomer. Native gel electrophoresis shows that the monomer-dimer equilibrium constant for SL3 under conditions used for the protein binding work is $\sim 10^3 \text{ M}^{-1}$ (data not shown), five orders of magnitude smaller than the value of K obtained for a similar-length hairpin with a self-complementary loop [28]. Only a small amount of RNA dimer is present under the conditions used in the present studies, but complexation of duplex RNA to protein could drive the dimerization process, producing a complex having bound duplex RNA. This complex could subsequently oligomerize to form higher molecular weight species such as the 8:4 complex. Alternatively, two hairpin monomers (the major RNA form in solution) could react with one protein molecule stepwise to produce the 2:1 complex. Since we found no evidence for the 1:1 complex, the binding must be highly cooperative, i.e. the binding constant for a second SL3 molecule to NCp7 when a first is bound is much larger than for the first. Clearly, characterizing the bound RNA and determining the binding mechanism will require additional study.

It is difficult to extrapolate our findings to the state of packaged RNA in the virion of the HIV virus [1]. However, the oligomerization of the 2:1 complex suggests that individual RNA-protein particles contain functionalities that allow self-association. For long RNA strands bound to many NCp7 molecules, this kind of particle-particle interaction could be one of the mechanisms used to compact and organize RNA in the virion of the virus. We are examining the binding of NCp7 to longer Ψ -RNA segments to determine whether the packaging region of HIV, as it likely to exist in the virus, can be used as a target for drug action.

Acknowledgements

This work was supported in part by NIH grant GM32691 to P.N. Borer. We wish to thank Dr Lucia Pappalardo for samples of SL3 used in the study. Special thanks are due to Professor P.N. Borer for helpful discussions pertaining to this work.

References

- [1] J. Coffin, in: R. Weiss, N. Teich, H. Varmus, J. Coffin (Eds.), *RNA Tumor Viruses*, 2, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1985, pp. 17–74.
- [2] J.G. Omichinski, G.M. Clore, K. Sakaguchi, E. Appella, A.M. Gronenborn, Structural characterization of a 39-residue synthetic peptide containing the two zinc binding domains from the HIV-1 p7 nucleocapsid protein by CD and NMR spectroscopy, *FEBS Lett.* 292 (1991) 25–30.
- [3] N. Morellet, N. Jullian, H. de Rocquigny, B. Maigret, J.L. Darlix, B.P. Roques, Determination of the structure of the nucleocapsid protein NCp7 from the human immunodeficiency virus type 1 by 1H NMR, *EMBO J.* 11 (1992) 3059–3065.
- [4] M.F. Summers, L.E. Henderson, M.R. Chance et al., Nucleocapsid zinc fingers detected in retroviruses: EXAFS studies of intact viruses and the solution-state structure of the nucleocapsid protein from HIV-1, *Protein Sci.* 1 (1992) 563–574.
- [5] C. Barat, V. Lullien, O. Schatz et al., HIV-1 reverse transcriptase specifically interacts with the anticodon domain of its cognate primer tRNA, *EMBO J.* 8 (1989) 3279–3285.
- [6] H. De Rocquigny, C. Gabus, A. Vincent, M.C. Fournie-Zaluski, B. Rogues, J.L. Darlix, Viral RNA annealing activities of human immunodeficiency virus type 1 nucleocapsid protein require only peptide domains outside the zinc fingers, *Proc. Natl. Acad. Sci. USA* 89 (1992) 6472–6476.
- [7] K. Sakaguchi, N. Zambrano, E.T. Baldwin et al., Identification of a binding site for the human immunodeficiency virus type 1 nucleocapsid protein, *Proc. Natl. Acad. Sci. USA* 90 (1993) 5219–5223.
- [8] J.L. Darlix, C. Gabus, M.T. Nugeyre, F. Cavel, F. Barre-Sinoussi, Cis elements and trans-acting factors involved in the RNA dimerization of the human immunodeficiency virus HIV-1, *J. Mol. Biol.* 216 (1990) 689–699.
- [9] D. Muriaux, H. de Rocquigny, B.P. Roques, J. Paoletti, NCp7 activates HIV-1Lai RNA dimerization by converting a transient loop-loop complex into a stable dimer, *J. Biol. Chem.* 271 (1996) 33686–33692.
- [10] R.D. Berkowitz, S.P. Goff, Analysis of binding elements in the human immunodeficiency virus type 1 genomic RNA and nucleocapsid protein, *Virology* 202 (1994) 233–246.
- [11] J. Clever, C. Sasseti, T.G. Parslow, RNA secondary structure and binding sites for gag gene products in the 5' packaging signal of human immunodeficiency virus type 1, *J. Virol.* 69 (1995) 2101–2109.
- [12] V. Tanchou, C. Gabus, V. Rogemond, J.L. Darlix, Formation of stable and functional HIV-1 nucleoprotein complexes in vitro, *J. Mol. Biol.* 252 (1995) 563–571.
- [13] R.D. Berkowitz, J. Luban, S.P. Goff, Specific binding of human immunodeficiency virus type 1 gag polyprotein and nucleocapsid protein to viral RNAs detected by RNA mobility shift assays, *J. Virol.* 67 (1993) 7190–7200.
- [14] F. Dib-Hajj, R. Khan, D.P. Giedroc, Retroviral nucleocapsid proteins possess potent nucleic acid strand renaturation activity, *Protein Sci.* 2 (1993) 231–243.
- [15] Z. Tsuchihashi, P.O. Brown, DNA strand exchange and selective DNA annealing promoted by the human immunodeficiency virus type 1 nucleocapsid protein, *J. Virol.* 68 (1994) 5863–5870.
- [16] N. Morellet, H. Demene, V. Teilleux et al., Structure of the complex between the HIV-1 nucleocapsid protein NCp7 and the single-stranded pentanucleotide d(ACGCC), *J. Mol. Biol.* 283 (1998) 419–434.
- [17] S.P. Stoylov, C. Vuilleumier, E. Stoylova et al., Ordered aggregation of ribonucleic acids by the human immunodeficiency virus type 1 nucleocapsid protein, *Biopolymers* 41 (1997) 301–312.
- [18] M. Lapadat-Tapolsky, H. de Rocquigny, D. VanGent, B. Roques, R. Plasterk, J.L. Darlix, Interactions between HIV-1 nucleocapsid protein and viral DNA may have important functions in the viral life cycle, *Nucl. Acids Res.* 21 (1993) 831–839.
- [19] S. Druillennec, A. Caneparo, H. de Rocquigny, B.P. Roques, Evidence of interactions between the nucleocapsid protein NCp7 and the reverse transcriptase of HIV-1, *J. Biol. Chem.* 274 (1999) 11283–11288.
- [20] Y. Mély, H. de Rocquigny, M. Sorinas-Jimeno et al., Binding of the HIV-1 nucleocapsid protein to the primer tRNA(3Lys), in vitro, is essentially not specific, *J. Biol. Chem.* 270 (1995) 1650–1656.
- [21] H. Al-Ghusein, H. Ball, G.L. Iglor et al., Chemically synthesised human immunodeficiency virus P7 nucleocapsid protein can self-assemble into particles and binds to a specific site on the tRNA(Lys,3) primer, *Biochem. Biophys. Res. Commun.* 224 (1996) 191–198.
- [22] E. LeCam, D. Coulaud, E. Delain et al., Properties and growth mechanism of the ordered aggregation of a model RNA by the HIV-1 nucleocapsid protein: an electron microscopy investigation, *Biopolymers* 45 (1998) 217–229.
- [23] J.L. Clever, D.A. Eckstein, T.G. Parslow, Genetic dissociation of the encapsidation and reverse transcription functions in the 5' R region of human immunodeficiency virus type 1, *J. Virol.* 73 (1999) 101–109.
- [24] J.A. Berglund, B. Charpentier, M. Rosbash, A high

- affinity binding site for the HIV-1 nucleocapsid protein, *Nucl. Acids Res.* 25 (1997) 1042–1049.
- [25] J.A. Loo, T.P. Holler, S.K. Foltin et al., Application of electrospray ionization mass spectrometry for studying human immunodeficiency virus protein complexes, *Proteins Struct. Func. Genet.* 2 (Suppl.) (1998) 28–37.
- [26] R.N. DeGuzman, Z.R. Wu, C.C. Stalling, L. Pappalardo, P.N. Borer, M.F. Summers, Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element, *Science* 279 (1998) 384–388.
- [27] C. Vuilleumier, E. Bombarda, N. Morellet, D. Gerard, B.P. Roques, Y. Mély, Nucleic acid sequence discrimination by the HIV-1 nucleocapsid protein NCp7: a fluorescence study, *Biochemistry* 38 (1999) 16816–16825.
- [28] M.F. Shubsda, M.P. McPike, J. Goodisman, J.C. Dabrowiak, Monomer-dimer equilibrium constants of RNA in the dimer initiation site of human immunodeficiency virus type 1, *Biochemistry* 38 (1999) 10147–10157.
- [29] J. Grodberg, J.J. Dunn, ompT encodes the Escherichia coli outer membrane protease that cleaves T7 RNA polymerase during purification, *J. Bacteriol.* 170 (1998) 1245.
- [30] J.F. Milligan, O.C. Uhlenbeck, Synthesis of small RNAs using T7 RNA polymerase, *Methods Enzymol.* 180 (1989) 51.
- [31] G. Kallansrud, B. Ward, A comparison of measured and calculated single- and double-stranded oligodeoxynucleotide extinction coefficients, *Anal. Biochem.* 236 (1996) 134–138.
- [32] P.N. Borer, Ph.D. Dissertation, University of California, Berkeley, CA, 1972.
- [33] T. Sugiyama, H. Fraenkel-Conrat, Identification of 5' linked adenosine as end group of TMV-RNA, *Proc. Natl. Acad. Sci. USA* 47 (1961) 1393–1397.
- [34] F.B. Howard, C.Q. Chen, P.D. Ross, H.T. Miles, Hairpin formation in the self-complementary dodecamer d-GGTACGCGTACC and derivatives containing GA and IA mispairs, *Biochemistry* 30 (1991) 779–782.
- [35] P.J. Tummino, J.D. Scholten, P.J. Harvey et al., The in vitro ejection of zinc from human immunodeficiency virus (HIV) type 1 nucleocapsid protein by disulfide benzamides with cellular anti-HIV activity, *Proc. Natl. Acad. Sci. USA* 93 (1996) 969–973.
- [36] J. Sambrook, E.F. Fritsch, T. Maniatis, *Molecular Cloning: A Laboratory Manual*, 2nd ed, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.
- [37] J.R. Cann, Theory of electrophoresis of hybridizing enzymes with kinetic control: implications for population genetics of electrophoretic markers, *J. Theor. Biol.* 127 (1987) 461–477.
- [38] J.R. Cann, Phenomenological theory of gel electrophoresis of protein-nucleic acid complexes, *J. Biol. Chem.* 264 (1989) 17032–17040.
- [39] J.R. Cann, Theory and practice of gel electrophoresis of interacting macromolecules, *Anal. Biochem.* 237 (1996) 1–16.
- [40] C. Kleinschmidt, K. Tovar, W. Hillen, Computer simulations and experimental studies of gel mobility patterns for weak and strong non-cooperative protein binding to two targets on the same DNA: application to binding on Tet repressor variants to multiple and single *tet* operator sites, *Nucl. Acids Res.* 19 (1991) 1021–1028.
- [41] M.F. Shubsda, J. Goodisman, J.C. Dabrowiak, Characterization of hairpin-duplex interconversion of DNA using polyacrylamide gel electrophoresis, *Biophys. Chem.* 76 (1999) 95–115.
- [42] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, *Numerical Recipes in Fortran: the Art of Scientific Computing*, 2nd ed, Cambridge University Press, Cambridge, 1992.
- [43] T.L. South, P.R. Blake, R.C. Sowder, L.O. Arthur, L.E. Henderson, M.F. Summers, The nucleocapsid protein isolated from HIV-1 particles binds zinc and forms retroviral-type zinc fingers, *Biochemistry* 29 (1990) 7786–7789.
- [44] N.B. McDonnell, R.N. DeGuzman, W.G. Rice, J.A. Turpin, M.F. Summers, Zinc ejection as a new rationale for the use of cystamine and related disulfide-containing antiviral agents in the treatment of AIDS, *J. Med. Chem.* 40 (1997) 1969–1976.
- [45] N.C. Stellwagen, K. Strutz, C. Gelfi, P.G. Righetti, DNA and buffers: the hidden danger of complex formation, *Biophys. J.* 76 (1999) A323.
- [46] M.A. Urbaneja, J.R. Casas-Finet, A. Ozarowski, A.H. Maki, A luminescence study of the binding of HIV-1 nucleocapsid protein NCp7 to stem loop sequences, *Biophys. J.* 78 (2000) 301A.