1042–1049 *Nucleic Acids Research, 1997, Vol. 25, No. 5*

A high affinity binding site for the HIV-1 nucleocapsid protein

J. Andrew Berglund, Bruno Charpentier and Michael Rosbash*

Howard Hughes Medical Institute and Departments of Biology and Biochemistry, Brandeis University, Waltham, MA 02254, USA

Received October 28, 1996; Revised and Accepted January 10, 1997

ABSTRACT

The nucleocapsid protein (NC) of HIV-1 is a small zinc finger protein that contributes to multiple steps of the viral life cycle, including the proper encapsidation of HIV RNA. This is accomplished through an interaction between NC and a region at the 5′-end of the RNA, defined as the Psi element. However, the specificity of NC for Psi or for RNA in general is not well understood. To study this problem, we used SELEX to identify high affinity RNA ligands that bind to NC. A 'winner' molecule (SelPsi), as well as a subregion of Psi RNA, were further characterized to understand the interaction between NC and SelPsi and its relationship to the interaction between NC and Psi. The comparison makes predictions about the sequence and structure of a high affinity binding site within the HIV-1 Psi element.

INTRODUCTION

Nucleocapsid protein (NC) is involved in encapsidation and other steps in the viral life cycle of HIV. It is a proteolytic fragment of the *gag* gene product, contains two zinc finger motifs (Cys-X_2-Cys-X_4-His-X_4-Cys) and is rich in arginines and lysines. Proteins closely related to NC are found in all retroviruses. The NC domain of Gag plays a role in both encapsidation and dimerization of the viral RNA (1–4). *In vitro*, NC has been shown to participate in both annealing of $tRNA^{Lys}$ to the primer binding site and synthesis of pro-viral DNA (5–7). NC has also been shown to be a potent chaperone for general nucleic acid folding and unfolding (8–10); this activity is almost certainly important for the many roles NC plays in the HIV life cycle.

The encapsidation and dimerization signals are closely linked and are present at the 5′-end of the HIV RNA (3). The RNA sequence important for encapsidation is referred to as the Psi element. NC binding to Psi is dependent on a functional N-terminal zinc finger element (11). As expected, mutations or deletions within Psi or NC have a severe effect on the ability of the virus to encapsidate viral RNA; viral particles are still formed but little or no viral RNA is present (2,12). In addition, mutations within Psi have been shown to reduce the efficiency of dimerization *in vitro* (3,13–15). The close proximity of the encapsidation and dimerization signals suggests that the two processes are linked, but this remains to be clarified.

In one model, the secondary structure of Psi has been shown to contain four stem–loops (Fig. 2b; 15,16). In many different studies, all four stem–loops have been proposed to be important either for encapsidation or more directly as NC binding sites (2,12,15–17). The first stem–loop has also been shown to be a major determinant for RNA dimerization and has been proposed to be involved in a 'kissing' loop model (14). Stem–loops 3 and 4 are of particular interest, because this region is downstream of the major 5′-splice site and would contribute to proper selection of unspliced RNA rather than spliced RNA for packaging. NC coats the viral RNA and ~2000 NC molecules are associated with the dimer packaging substrate (18). The binding site for an individual NC molecule is 8 nt (19) and the binding of NC is cooperative (9,20).

How unspliced viral RNA is selected for packaging is not fully understood. Although NC is a non-specific nucleic acid binding protein, several groups have shown that NC or Gag specifically interacts with Psi RNA (11,15,17,20–22). However, this specificity is modest, as the observed difference in affinity was only between 10- and 100-fold when antisense Psi RNA was used as the negative control (11,15). The highest affinity binding has been localized to ~200 nt within Psi (15), but further deletions decrease NC binding substantially, making it difficult to define a smaller high affinity site. Three of the stem-loops (1, 3 and 4; see Fig. 2b) within the ~200 nt fragment were individually tested for binding by Clever and Parslow; all three bind to NC with a 4-fold reduction in affinity compared with the 200 nt intact Psi fragment (15). But this would appear to pose a specificity problem, because many other RNAs bind to NC with only a 2-fold further reduction in affinity compared with these three stem–loops (15).

To explore NC specificity further, we performed a SELEX experiment (23,24) and characterized the interaction of NC with the winner RNA, SelPsi. This was done by structure probing, footprinting and binding of both SelPsi and a mutant version of SelPsi. In addition, parallel binding and footprinting studies were carried out on RNAs derived from the HIV-1 Psi element. The data point to a high affinity binding site within the HIV-1 Psi region. This is due in part to the resemblance of SelPsi (Fig. 2a) to the Psi stem–loop 3 region, including the single-strand region 5′ of the stem (Fig. 7).

*To whom correspondence should be addressed. Tel: +1 617 736 3160; Fax: +1 617 736 3164; Email: rosbash@binah.cc.brandeis.edu

MATERIALS AND METHODS

Preparation of nucleocapsid protein

The clone for NC was a kind gift from Zenta Tsuchihashi and Patrick Brown and the protein was purified as reported (8). This clone comprises the first 71 amino acids of the *gag* gene product plus six histidines and an extra methionine at the N-terminus for nickel column purification.

SELEX experiment

The protocol followed for the SELEX experiment is similar to that published (23). A random pool of RNA was generated with the following sequence GGAGACAGUCCGAGC(N)$_{40}$GGGUCAAUGCGUCAUAGGAUCCCGC. This was done by PCR using the three oligonucleotides GCGGAATTCTAATACTACTCACTATAGGAGACAGTCCGAGCC (5′ oligonucleotide containing an *Eco*RI site and the T7 promoter), GGAGACAGTCCGAGCC(N)$_{40}$GGGTCAATGCGTCATA (template for PCR reaction containing 40 randomized nucleotides) and GCGGGATCCTATGACGCATTGACCC (3′ oligonucleotide containing a *Bam*HI site). The template generated from the PCR reaction was used for *in vitro* transcription as described (25). The RNA was purified on a 20% polyacrylamide gel, ethanol precipitated, run through a BioRad P6 spin column, heated at 95°C for 2 min and then put on ice (annealing reaction). In the first round of *in vitro* selection, NC was at a final concentration of 50 nM and the pool 0 RNA was at a final concentration of 500 nM. After every other round the concentrations of both the RNA and NC were lowered until in the final round of selection (ninth round) NC was at 1.25 nM and pool 8 RNA was at 100 nM. The concentration of tRNA present in the first six rounds of selection was 400 nM; for the last three rounds the tRNA concentration was raised to 600 nM. The binding buffer consisted of 50 mM Tris, pH 7.5, 100 mM NaCl, 1 mM $MgCl_2$, 30 µM ZnCl and 10 mM DTT. Pool 0 RNA was pre-filtered through a 0.45 µm nitrocellulose centrifugal filter unit (Schleicher & Schuell) to select against RNAs that bound selectively to the filter. This pre-filtering was also done every other round. The RNA and NC were incubated at room temperature for 20 min in binding buffer. The filter unit was pre-wetted with H_2O; the RNA–NC was then filtered, followed by a wash with 1 ml binding buffer. The bound RNA was eluted from the filter using 7 M urea and heated for 3 min at 95°C. After phenol extraction and ethanol precipitation, half of the RNA was used for reverse transcription. Reverse transcription was done at 45°C for 30 min using MMLV. After the ninth round of selection PCR products were digested with *Bam*HI and *Eco*RI and then inserted into pSP64 (Promega). Sequencing was done using the fmol™ DNA sequencing system (Promega).

Filter binding assay

All RNAs except Psi 27 were made by *in vitro* transcription from PCR products using [α-^{32}P]UTP to internally label the RNAs. Psi 27 was made using Perseptive Expedite RNA amidites on an Expedite 8909 Oligonucleotide Synthesizer and then kinased with [γ-^{32}P]ATP (26). RNAs were gel purified after labeling and treated in a similar manner as the selection pools of RNA. The approximate concentration of RNA in a filter binding experiment was 50 pM. NC was serially diluted in 1× filter binding buffer; this was done for all dilutions of NC. RNAs were annealed as in the selection experiment. Radioactive RNA, NC and tRNA (50 nM) were incubated in 1× binding buffer for 30 min on ice and then filtered at room temperature through a 0.45 µm nitrocellulose filter (Schleicher & Schuell). Radioactivity was quantitated using a BioRad Molecular Imager. The retention of free RNA (always <5%) was subtracted from all data points. Different RNA–NC complexes were retained with different efficiencies on the nitrocellulose: for the RNAs studied here the efficiency of retention ranged from 40 to 90%. This has been observed with other RNA–protein complexes in filter binding studies (27 and references therein). The smaller structured RNAs had the lower efficiencies of retention; for example SelPsi was retained at 40%. The fraction bound was taken from the plateau of binding and the plateau was assigned as 1.0. Plots were made with the program Microcal Origin (Microcal Software Inc.).

Sequences from the SELEX experiment

#1	**UGUUGGUGGUGCCG**UUCGGGCUGCAC	(3)
#2	UCCCGGUGCGUGCUGUGUGUGGAGUCUGGGCUGUCGGU	(2)
#3	UCGUGGUAGCCCACUUGUGGGAGUAGUCCUCGCUGUCUA	(2)
#4	AACGGGUGUGCUUGCUGACCGUGUGGUUGCGACUGUGCCU	(2)
#5	GUUUUGUGGUAGUCCCAUUCGGAGUAGUCAUCGGCCUG	
#6	UCCACGGUGCGUGCUGUGUGUGGAGUCUGGGCUGUCGGUG	
#7	CUGUUCGUAUACACGAAUGUAGCUUAAAGUGCCUAGGCC	
#8	CAUAGUGCUUGCUGUCUGGGCCUGGCAUCUCUAGCUGU	
#9	CUCU**UGUUGGUGGUGCCG**AGUGGUUCGUGGUUGGCCUCUG ***	
#10	GUUUUGUGGUAGCCCAUUCGGGAGUAGUCAUCGGUCUGU	(2)
#11	UACUUGUUCUGCCAUCAACUACUGCCCAUUAACUUUCCUC	(2)
#12	GUAAGCUUGGUAGCCGUUGUGGAGUAGUCGACUGUUUGUU	(2)
#13	GUUUUGUGGUAGCCCAUUCGGGAGUAGUCAUCGGCCUGU	(2)
#14	UGCCGGUGCGUGCUGAGUGGACGGGGUUCGUGUCUCAGC	

Figure 1. *In vitro* selection results. After nine rounds of *in vitro* selection 23 individual clones were sequenced. Only the randomized regions are shown. Many of the clones are missing nucleotides from the randomized region; some clones have as many as 14 nt deleted. The number of times each clone was isolated is in brackets. Bold and underlined nucleotides indicate a region of nucleotides conserved between clones 1 and 9. SelPsi (Fig. 2) was derived from the sequence indicated by three stars; the only differences between SelPsi and 9 are that the adenosine adjacent to the conserved region is changed to a uridine and the sequence of the stem has been changed for better *in vitro* transcription and stability.

Structure probing and footprinting

SelPsi and Psi 76 were treated with calf intestine alkaline phosphatase for 20 min at 37°C (26), phenol extracted and ethanol precipitated with glycogen as carrier. Both RNAs were kinased with [γ-^{32}P]ATP and purified in the same manner as the RNAs for the filter binding experiments. Before RNase digestion, RNAs were incubated at 65°C in SSC buffer (17.5 mM sodium citrate, pH 7.0, 150 mM sodium chloride) for 5 min then placed on ice for a few minutes; they were then incubated at room temperature for 10 min in the presence of NC in 1× binding buffer. RNase A (5 Prime→3 Prime Inc.), RNase T1 (Ambion) or RNase V1 (Pharmacia) was added to the mix, which was then incubated at 37°C for 5 min. RNase T1 is specific for single-stranded guanosines, RNase A for single-stranded pyrimidines and RNase V1 for double-stranded regions. To stop the reaction an equal volume of phenol was added. After precipitation with ethanol, RNAs were resuspended in formamide dye and separated on 12.5% denaturing gels.

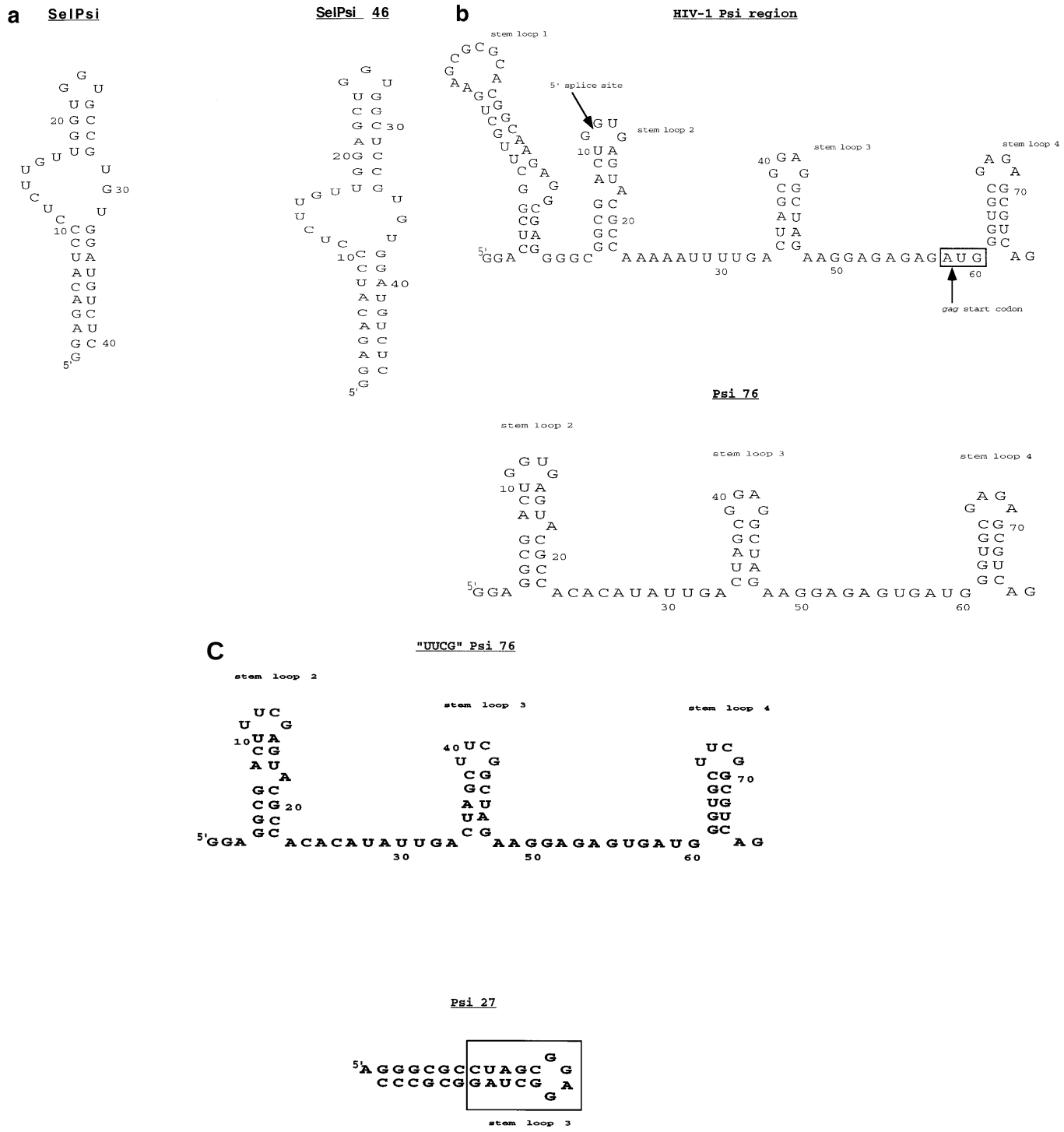

Figure 2. Predicted secondary structures of RNAs. (**a**) Predicted secondary structures of SelPsi and SelPsi 46 (Mulfold). (**b**) Secondary structure of the HIV-1 Psi region (13,14) and predicted secondary structure of Psi 76 (Mulfold). (**c**) Predicted secondary structures of UUCG Psi 76 and Psi 27 (Mulfold).

RESULTS

SELEX and characterization of SelPsi

A randomized region of 40 nt was used with the expectation that a region of this size plus flanking sequences would be sufficient for selection of high affinity RNA ligands for NC. The flanking sequences, used for reverse transcription and PCR, were designed to avoid self-complementarity and to minimize primer dimer formation as well as secondary structures that might bias the selection. The RNA was heated at 95°C for 2 min and then placed immediately on ice, to favor intramolecular structures. In addition, relatively low concentrations of NC (50–1.25 nM) were used for the selection and the RNA was always in large excess

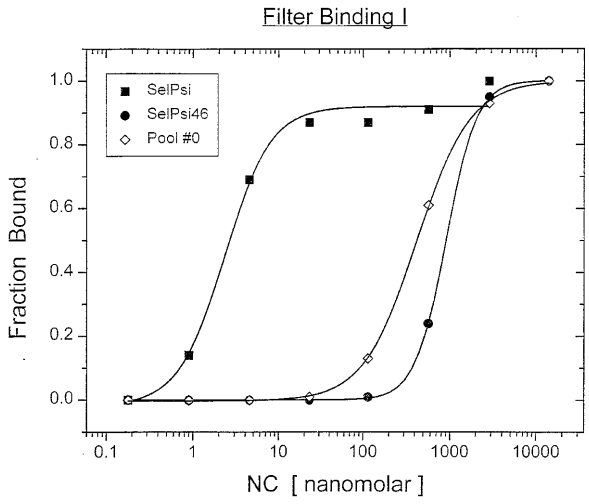

Figure 3. Binding data I. Filter binding for NC with SelPsi, SelPsi 46 and pool 0. All of the filter binding experiments were done at least twice.

over NC. These steps were taken to favor the interaction of single NC molecules with single RNA molecules.

The results from the SELEX experiment indicate that NC favors stretches of guanosines and uridines. After nine rounds of selection, the pool 8 RNA bound to NC with ~50-fold higher affinity than pool 0 RNA (data not shown). Twenty three clones were sequenced from pool 8. Nine were duplicates, giving a total of 14 different sequences. Almost all of these sequences contain long stretches of guanosines and uridines (Fig. 1). Two clones, 1 and 9, contain a conserved stretch of 14 nt that are guanosine and uridine rich. (In Figure 1 this conserved sequence is in bold and underlined.)

Twelve of the 14 clones were tested individually for NC binding. Of these, 1 and 9 bound to NC with an affinity that is significantly better than that of the rest of the clones (data not shown). Both clones are predicted to fold into secondary structures that contain the conserved 14 nt at the end of a long stem–loop. Clone 9 was easily reduced to a smaller RNA, SelPsi (Fig. 2a), that contains only 40 nt and retains high affinity NC binding. Reducing the size of the RNA resulted in a few changes to the sequence compared with the initial clone; the stem was shortened and changes were made in the stem to optimize both transcription and stability of SelPsi. In addition, a single adenosine after the conserved region was changed to uridine. SelPsi binds to NC with an apparent K_d of 2.3 nM, which is 200-fold stronger than pool 0 RNA (Fig. 3).

To characterize how NC recognizes SelPsi, structure probing and footprinting studies were performed with RNases A, T1 and V1 (Fig. 4). Although the general secondary structure predicted by Mulfold (28) and the results determined by structure probing are similar, they are not exactly the same. The probing suggests that there are unusual structural features in the upper half of the molecule. Although these may be quite complicated, we propose two additional base pairs within the bulge, C11:G30 and U17:G28 (dashed lines in Fig. 4b represent these base pairs). These are not predicted by Mulfold and are based on insensitivity to RNases A and T1. Unusual features are also suggested in the upper part of the stem, as this region is susceptible to RNases and is therefore probably rather unstable, and C6 within the lower part of the stem is cleaved by RNase A.

To determine which regions are affected when NC binds to SelPsi, footprinting studies were done at three different concentrations of NC (Fig. 4a). Most of the single-strand guanosines and pyrimidines are well protected by NC. Almost all of these are present in the upper part of the stem consisting of the pyrimidine bulge, the short unstable stem and the G-rich loop. The lower part of the stem probably does not interact directly or strongly with NC. In the presence of protein it is more sensitive to RNase V1 at G36 and unchanged at A34 and U35. RNase A cleavage of C6 is enigmatic, because all other nucleotides within this part of the stem are insensitive; even G36, the base pairing partner of C6, is not sensitive to RNase T1. It is possible that the helix at this position is kinked in some unusual manner that makes C6 RNase A sensitive. In the presence of NC C6 is partially protected and G36 becomes hypersensitive to RNase V1, which suggests a further conformational change in this region. In summary, we suggest that NC binds predominantly to the upper half of SelPsi, which alters the conformation of the lower stem.

Table 1. Summary of filter binding data

RNA	Apparent K_d (nM)
SelPsi	2.3
Pool 0	420
SelPsi 46	905
HIV Psi	58
Psi 76	138
UUCG' Psi 76	332
Psi 27	1450

The apparent K_ds are taken from the plots at half-maximal binding using the program Microcal Origin (Microcal Software, Inc.).

The footprinting data explain certain aspects of the NC–SelPsi interaction, but they do not clarify why SelPsi binds to NC with higher affinity compared with other RNAs (Table 1). The SelPsi–NC apparent K_d of 2.3 nM is 25-fold stronger than the affinity of NC for HIV Psi. Although NC favors stem–loops (stem–loops 1, 3 and 4 from HIV Psi; 15), SelPsi binds with much higher affinity (100-fold) than the individual stem–loops isolated from the HIV-1 Psi element. A partial explanation might be the unusual structure of the upper half. To test this prediction, we made a mutation that should both stabilize and lengthen the short upper stem proposed to exist within the upper half of SelPsi (Fig. 2a). This mutant RNA, SelPsi 46 (Fig. 2a), binds to NC with a 400-fold reduced affinity compared with SelPsi (Fig. 3). In fact, SelPsi 46 binds to NC 2-fold more weakly than pool 0. There are a number of possibilities for how this mutation reduces the binding affinity in such a dramatic fashion (see Discussion), but the simple conclusion is that strong NC–SelPsi binding is a function of detailed features of SelPsi sequence and structure.

HIV Psi

To better understand how NC interacts with HIV-1 Psi, different RNAs derived from this region were also used for binding and footprinting studies. The HIV-1 Psi region encompassing nt 632–837 (fragment B from Clever and Parslow) was used for the filter binding experiment to obtain an apparent K_d of 58 nM, identical to the value obtained by Clever and Parslow for this RNA (15). This fragment has been shown to bind NC with similar affinity compared with any other larger region of HIV RNA and

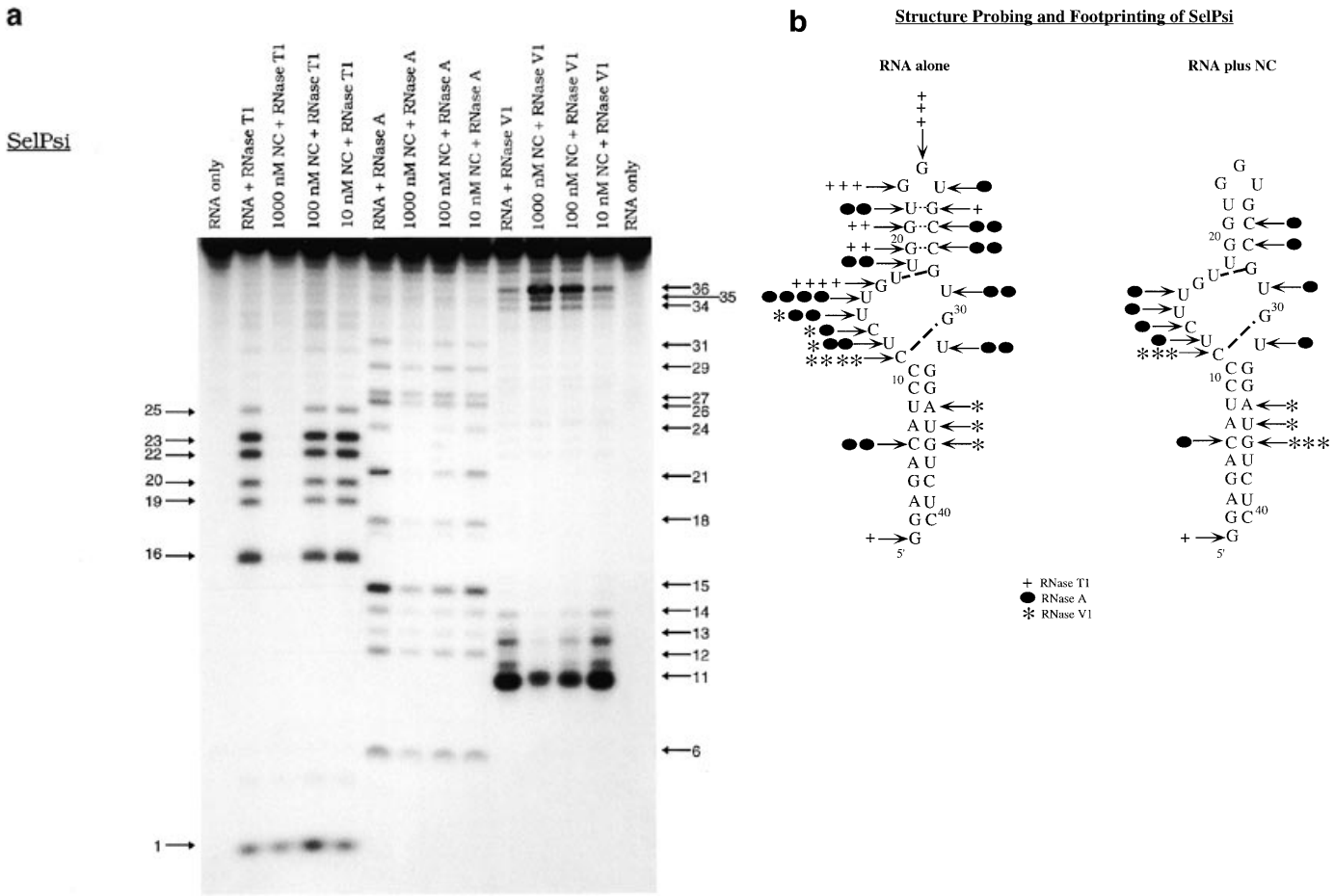

Figure 4. Footprinting of SelPsi RNA. (**a**) SelPsi was probed using RNases T1 (7.5 U/ml), A (0.001 U/ml) and V1 (0.056 U/ml), in both the absence and presence of NC. Three different concentrations of NC were incubated with SelPsi RNA before addition of RNases, as indicated at the top of the gel. All samples also contained 100 nM tRNA. The numbered arrows indicate the nucleotides as numbered in Figures 2a and 4b. **(b)** Summary of the structure probing (RNA alone) and footprinting (RNA plus NC) data. The different number of symbols next to each nucleotide represents the varying levels of RNase reactivity. The structure probing revealed two new base pairs (C11:G30 and U17:G28, represented by thick dashed lines) and showed that the upper stem is not stable (represented by thin dashed lines). In the presence of NC, only the upper half of SelPsi is protected.

it is therefore assumed to contain at least one high affinity NC binding site (15). A smaller RNA more amenable to footprinting and mutational studies was designed; Psi 76 (Fig. 2b) contains the region including stem–loops 2, 3 and 4 from the Psi region. We choose this region of Psi because two different studies have shown it to be important for packaging (2,16). Psi 76 binds NC with an apparent K_d of 138 nM, a slightly greater than 2-fold reduction in affinity compared with the larger Psi region (Fig. 5). Psi 76 has five point mutations compared with HIV-1 RNA (Fig. 2b). These point mutations allow better *in vitro* transcription and also eliminate a possible dimerization signal in the single-stranded region between stem–loops 2 and 3 (17).

Footprinting studies in the absence of NC are in good agreement with the predicted secondary structure (15,16) and also agree with the structure probing results of Clever and Parslow for this region of Psi (15). In the presence of NC, all of the single-stranded regions are protected, whereas parts of the double-stranded regions remain sensitive to RNase V1 (Fig. 6b). Interestingly, only in the presence of NC does RNase V1 cleave in stem–loop 4. In the absence of NC, the nucleotides proposed to base pair in stem–loop 4 are partially cleaved by RNase T1, implying that stem–loop 4 is not fully stable (Fig. 6b); this was

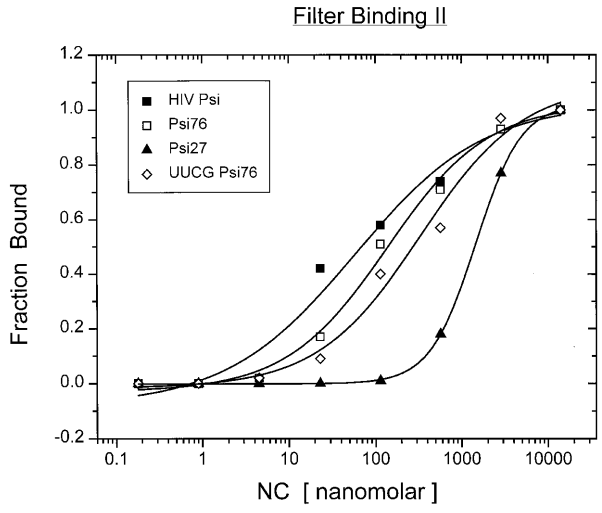

Figure 5. Binding data II. Filter binding for NC with the HIV-1 Psi region (clone from Clever and Parslow), Psi 76, UUCG Psi 76 and Psi 27.

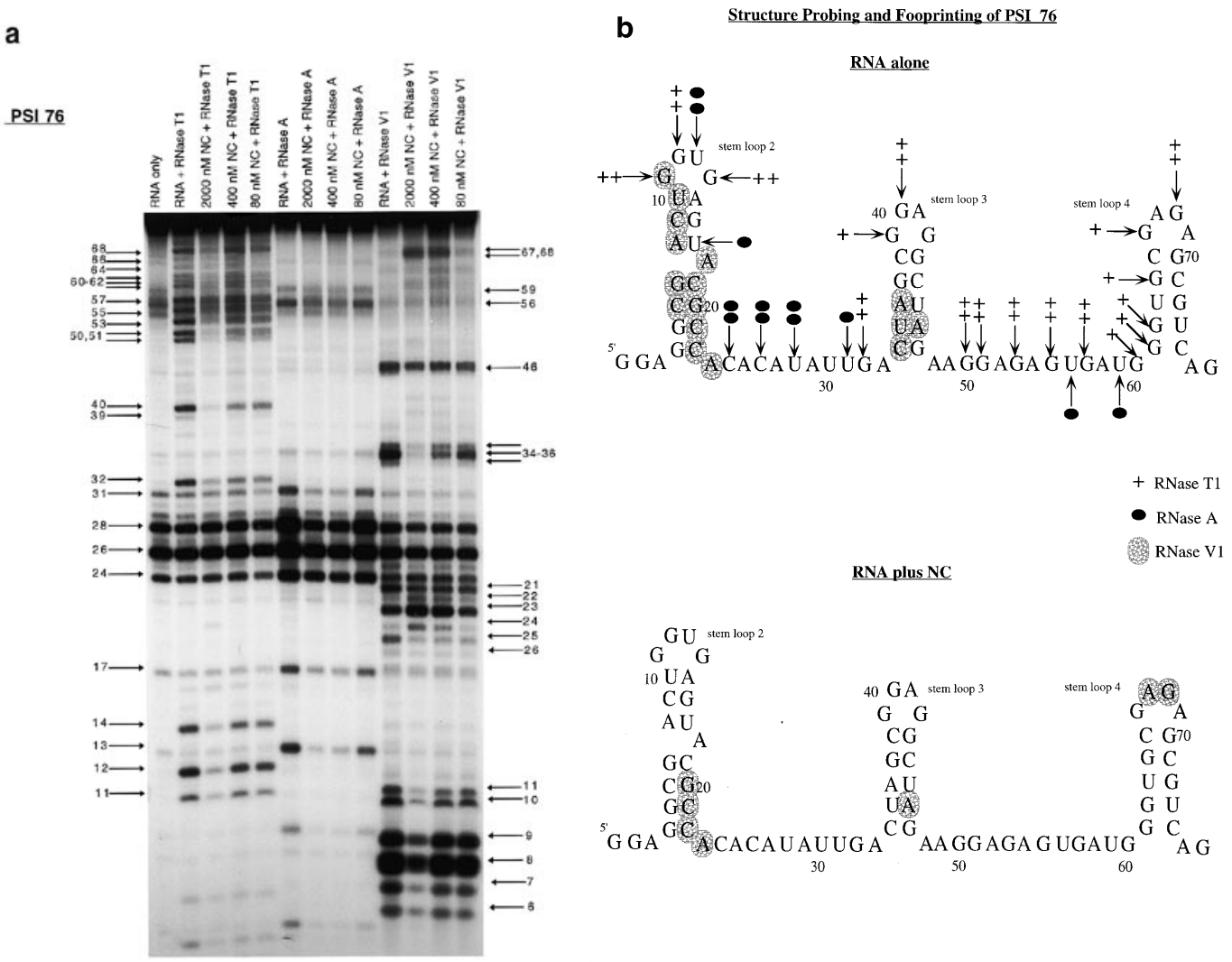

Figure 6. Footprinting of Psi 76. (**a**) Psi 76 was probed using RNases T1 (3 U/ml), A (1×10^{-4} U/ml) and V1 (0.05 U/ml). The numbered arrows represent the nucleotides as numbered in Figures 2b and 6b. The concentration of tRNA was 40 nM in all samples. (**b**) Summary of the structure probing (RNA alone) and footprinting (RNA plus NC) data. The structure probing data agree with the predicted structure of Psi 76. Only the reactivities that remain unchanged or are enhanced in the presence of NC are shown in the RNA plus NC panel. In the presence of NC, all of the single-stranded regions were protected. None of the stems appear to be disrupted and stem–loop 4 might even be stabilized, because A67 and G68 are cleaved by RNase V1 only in the presence of NC.

also observed by Clever and Parslow (15). When NC binds to Psi 76, the observed increase in stem–loop 4 cleavage by RNase V1 indicates that it is stabilized in the presence of NC. In summary, it appears that several NC molecules interact with Psi 76 but that the overall secondary structure is only minimally altered.

All three stem–loops within Psi 76 are purine rich and SelPsi contains two guanosines at the end of its loop; we assume that these purines, in particular the guanosines, are important for NC–Psi 76 binding. To test this hypothesis, we changed all three loops of Psi 76 into stable UUCG tetraloops (29; Fig. 2c). This change results in a nearly 3-fold decrease in binding compared with Psi 76 (Fig. 5). The structure of this RNA remained consistent with that of Psi 76, as shown by structure probing (data not shown). The decreased affinity confirms that the sequence of the loop nucleotides is involved in the higher affinity interaction between NC and Psi 76.

In an attempt to isolate a single higher affinity site from Psi, we synthesized a 27 nt RNA containing the sequence of Psi stem–loop 3 (Fig. 2c; Psi 27). Psi 27 binds to NC with an apparent K_d of 1450 nM, the lowest affinity RNA in this study. A likely reason is that Psi 27 has a strong stem and is also missing the adjacent single-stranded regions present within the HIV Psi element. Indeed, the construct originally used to observe higher affinity stem–loop 3 binding (K_d 200 nM) contained the single-strand regions on either side of stem–loop 3 (15). Either these single-stranded regions are important for higher affinity binding or the longer stem inhibits strong binding to Psi 27. Taken together, these observations and others (15) suggest that NC not only recognizes the sequence within loops but also utilizes the adjacent single-stranded regions to optimize interactions with RNA.

DISCUSSION

This SELEX experiment is the first performed with NC and supplies important information about the NC–RNA interaction.

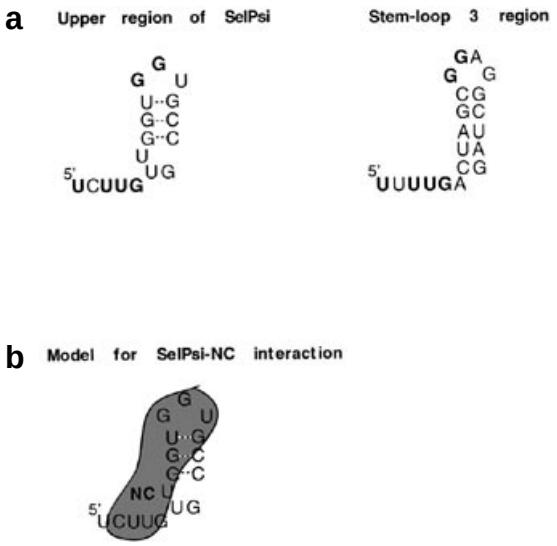

Figure 7. A conserved high affinity site in SelPsi and the HIV-1 Psi element and a model of the NC–SelPsi interaction. (**a**) Comparison of the upper half of SelPsi and the HIV-1 stem–loop 3 region. The conserved nucleotides are highlighted in bold. (**b**) A model of the NC–SelPsi interaction.

The selected high affinity RNA ligand also provides a lead candidate for inhibitor studies against HIV replication, because NC binds to SelPsi with 25-fold higher affinity than HIV Psi RNA (Table 1). High affinity RNA ligands for the HIV proteins Rev and Tat have been previously shown to inhibit HIV replication (30,31).

This and other work demonstrates that NC has a higher affinity for the HIV Psi RNA than for other RNAs (2,12,15–17). Although the reason for this higher affinity was unclear, the SELEX results suggest that NC favors stretches of guanosines and uridines. In addition, SelPsi forms a secondary structure that presents these favored nucleotides at the end of a long and stable stem–loop structure (Fig. 2a). SelPsi is, however, more than a simple stem–loop; it contains a stable lower stem, a bulge region, a short unstable stem and a 3 nt loop. The results from the footprinting experiments indicate that the stable lower stem is probably not in direct contact with NC. However, it is still involved in high affinity binding, because its absence decreases the binding affinity >100-fold (data not shown). The lower stem is likely involved in positioning or holding the upper half of SelPsi in the correct structure.

The footprinting indicates that the upper half of SelPsi is only modestly structured. Indeed, the addition of base pairs to its short unstable stem (SelPsi 46) decreased NC binding by 400-fold. This result indicates that either its stability or length is critical for high affinity binding. We favor the latter possibility, because changing two G:U base pairs to G:C (U17 and U21) within SelPsi did not change the NC binding affinity appreciably (data not shown). We suggest that an appropriate stem length serves to orient the relative positions of the loop and bulge so that NC can bridge the stem and interact optimally with both single-stranded elements (Fig. 7).

An issue that complicates interpretation of the NC–SelPsi interaction is that RNA dimerization appears to be linked to encapsidation (1,3,32). It is therefore not certain whether NC recognizes an RNA monomer or dimer for packaging. This issue was only indirectly addressed in these studies, as neither SelPsi nor Psi 76 appears to dimerize, as assayed by native gel electrophoresis (data not shown). Although NC may recognize the stem–loop 3 region or other high affinity sites in a dimer configuration, the simplest interpretation of our results is that NC initially binds to an RNA monomer and then may subsequently contribute to the dimerization process.

The upper half of SelPsi is very similar to the stem–loop 3 region within the HIV Psi element (Fig. 7). Both SelPsi and this region have short stems bracketed by two guanosines in the loop, as well as UYUUG (Y = pyrimidine) sequences on the 5′-side of the stem (Fig. 7). Based on the SELEX results and on prior encapsidation studies, we propose that this stem–loop 3 region is a high affinity site for NC within the context of HIV Psi. The integrity of these three stems is important for encapsidation efficiency (16) and the region encompassing stem–loop 3 is able to work as a packaging signal in a heterologous RNA (33). Taken together, the evidence indicates that the stem–loop 3 region is important for correct packaging of HIV RNA and is also a high affinity NC binding site. Other regions within Psi are also involved in packaging (16), suggesting that there are additional high affinity binding sites.

Although the upper portion of SelPsi and the stem–loop 3 region are similar, there are also differences that may explain the different binding affinities. The upper half of SelPsi is at the end of a strong stem; in contrast, the stem–loop 3 region is surrounded by single-stranded RNA. The short stem of SelPsi contains a bulged uridine and is rather unstable, whereas stem–loop 3 is canonical and stable. Also, the loops have some differences in sequence and in the number of nucleotides: SelPsi is GGU and stem–loop 3 is GGAG. Despite these differences, the data suggest that it is the presence of the lower stem in SelPsi that is predominantly responsible for the 25-fold difference in binding (because deleting the lower stem of SelPsi destroys the high affinity binding of SelPsi; data not shown).

Why isn't there a comparable high affinity structure within HIV Psi? A strong and relatively long stem like that present in the lower half of SelPsi might impede NC polymerization, because the lower stem might inhibit addition of subsequent NC molecules after the initial high affinity interaction. In contrast, the single-stranded RNA surrounding stem–loop 3 might be optimized for NC polymerization. The importance of polymerization would explain: (i) the lower affinity of Psi for single molecule binding; (ii) why other regions within Psi have been shown to be important for encapsidation; (iii) the possible presence of additional high affinity binding sites within Psi. The importance of polymerization also suggests an explanation for the modest specificity of the NC–Psi interaction compared with other well-studied protein–RNA interactions (see for example 27), namely high specificity might be incompatible with polymerization. A contribution of surrounding sequence and structure to polymerization, as well as only modest affinity and specificity to an initial 'high affinity' site, has also been proposed for Rev, another HIV RNA binding protein (34).

Can a factor of 10 in specificity account for selection of the correct RNA for encapsidation or are other factors involved? One possibility is that Gag has more specificity than NC, but this appears unlikely (15,21). Alternatively, subcellular localization of NC and unspliced HIV RNA might enhance the *in vivo* specificity for encapsidation. Finally, NC polymerization onto HIV RNA may contribute in an as yet poorly understood way to encapsidation specificity.

This rather poor specificity of NC for Psi is undoubtedly relevant to its other roles in the viral life cycle, in which it may

function as an RNA chaperone. In other words, high specificity may be incompatible with the catalysis of multiple annealing and folding reactions between other RNA sequence elements. Therefore, the multiple roles of NC may require a compromise between specificity on the one hand and promiscuous binding reactions on the other hand.

ACKNOWLEDGEMENTS

We thank our collegues, especially M.Moore, A.Rein and H.Colot, for critical reading of the manuscript. The work was supported in part by the National Institutes of Health to M.R. (GM 23549).

REFERENCES

1 Darlix,J.-L., Lapadat-Tapolsky,M., de Rocquigny,H. and Roques,B.P. (1995*) J. Mol. Biol.*, **254**, 523–537.
2 Aldovini,A. and Young,R.A. (1990) *J. Virol*., **64**, 1920–1926.
3 Darlix,J.-L., Gabus,C., Nugeyre,M.-T., Clavel,F. and Barre-Sinoussi,F. (1990*) J. Mol. Biol.*, **216**, 689–699.
4 Berkowitz,R.D., Ohagen,A., Hoglund,S. and Goff,S.P. (1995) *J. Virol*., **69**, 6445–6456.
5 de Rocquigny,H., Gabus,C., Vincent,A., Fournie-Zalusky,M.-C., Roque,B. and Darlix,J.-L. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 6472–6476.
6 Allain,B., Lapadat-Tapolsky,M., Berlioz,C. and Darlix,J.-L. (1994) *EMBO J*., **13**, 973–981.
7 Lapadat-Tapolsky,M., de Rocquigny,H., van Gent,D., Roques,B., Plasterk,R. and Darlix,J.-L. (1993) *Nucleic Acids Res*., **21**, 831–839.
8 Tsuchihashi,Z. and Brown,P.O. (1994) *J. Virol*., **68**, 5863–5870.
9 Khan,R. and Giedroc,D.P. (1992) *J. Biol. Chem*., **267**, 6689–6695.
10 Herschlag,D. (1995) *J. Biol. Chem*., **270**, 20871–20874.
11 Dannull,J., Surovoy,A., Jung,G. and Moelling,K. (1994) *EMBO J*., **13**, 1525–1533.
12 Lever,A., Gottlicher,M., Haseltime,W. and Sodroski,J. (1989) *J. Virol*., **63**, 4085–4087.
13 Awang,G. and Sen,D. (1993) *Biochemistry*, **32**, 11453–11457.
14 Paillart,J.-C., Skripkin,E., Ehresmann,B., Ehresmann,C. and Marquet,R. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 5572–5577.
15 Clever,J., Sassetti,C. and Parslow,T.G. (1995) *J. Virol*., **69**, 2101–2109.
16 McBride,M.S. and Panganiban,A.T. (1996) *J. Virol*., **70**, 2963–2973.
17 Sakaguchi,K., Zambrano,N., Baldwin,E.T., Shapiro,B.A., Erickson,J.W., Omichinski,J.G., Clore,G.M., Gronenborn,A.M. and Appella,E. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 5219–5223.
18 Coffin,J. (1985) In Weiss,R., Teich,N., Varmus,H. and Coffin,J. (eds), *RNA Tumor Viruses*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, Vol. 2, pp. 17–74.
19 Khan,R. and Giedroc,D.P. (1994) *J. Biol. Chem*., **269**, 22538–22546.
20 Berkowitz,R., Luban,J. and Goff,S. (1993) *J. Virol*., **67**, 7190–7200.
21 Berkowitz,R.D. and Goff,S.P. (1994) *Virology*, **202**, 233–246.
22 Luban,J. and Goff,S.P. (1994) *J. Virol*., **68**, 3784–3793.
23 Turek,C. and Gold,L. (1990) *Science*, **249**, 505–510.
24 Ellington,A.D. and Szostak,J.W. (1990) *Nature*, **346**, 818–822.
25 Milligan,J.F. and Uhlenbeck,O.C. (1989) *Methods Enzymol*., **180**, 51–62.
26 Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
27 Hall,K.B. and Stump,W.T. (1992) *Nucleic Acids Res*., **20**, 4283–4290.
28 Jaeger,J.A., Turner,D.H. and Zuker,M. (1989) *Methods Enzymol*., **183**, 281–306.
29 Molinaro,M. and Tinoco,I. (1995) *Nucleic Acids Res*., **23**, 3056–3063.
30 Lee,S.W., Gallardo,H.F., Gilboa,E. and Smith,C. (1994) *J. Virol*., **68**, 8254–8264.
31 Lee,S.W., Gallardo,H.F., Gaspar,O., Smith,C. and Gilboa,E. (1995) *Gene Ther*., **2**, 377–384.
32 Feng,Y.-X., Copeland,T.D., Henderson,L.E., Gorelick,R.J., Bosche,W.J., Levin,J.G. and Rein,A. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 7577–7581.
33 Hayashi,T., Shioda,T., Iwakura,Y. and Shibuta,H. (1992) *Virology*, **188**, 590–599.
34 Zemmel,R.W., Kelley,A.C., Karn,J. and Butler,P.J.G. (1996) *J. Mol. Biol*., **258**, 763–777.