Affinities of the Nucleocapsid Protein for Variants of SL3 RNA in HIV-1†

Andrew C. Paoletti, Michael F. Shubsda, Bruce S. Hudson,* and Philip N. Borer*

Department of Chemistry, Graduate Program in Structural Biology, Biochemistry, and Biophysics, Syracuse University, Syracuse, New York 13244-4100

Received June 17, 2002; Revised Manuscript Received September 23, 2002

ABSTRACT: Efficient packaging of genomic RNA into new HIV-1 virus particles requires that nucleocapsid domains of precursor proteins bind the SL3 tetraloop (G317-G-A-G320) from the 5′-untranslated region. This paper presents the affinities of 35 RNA variants of SL3 for the mature 55mer NC protein, as measured by fluorescence quenching of tryptophan-37 in the protein by nucleobases. The 1:1 complexes that form in 0.2 M NaCl have dissociation constants ranging from 8 nM (GGUG) to 20 μM (GAUA). The highly conserved (GGAG) sequence for the wild type is not the most stable (K_d = 28 nM), suggesting that other selective pressures beyond the stability of the complex must be satisfied. The leading requirement for strong interaction is for G320, followed closely by G318. Replacing either with U, A, or C reduces affinity by a factor of 15−120. NC-domains from multiple proteins combine to recognize unpaired G_2-loci, where two guanines are in close proximity. We have previously measured affinities of the NC protein for the important stem-loops of the major packaging domain [Shubsda, M. F., Paoletti, A. C., Hudson, B. S., and Borer, P. N. (2002) *Biochemistry 41*, 5276−82]. Comparison with the present work shows that the nature of the stem also modulates NC−RNA interactions. Placing the G_2-loci from the apical SL2 or SL1 loops on the SL3 stem increases affinity by a factor of 2−3, while placing the SL4 loop on the SL3 stem reduces affinity 50-fold. These results are interesting in the context of RNA−protein interaction, as well as for the discovery of antiNC agents for AIDS therapy.

The general description of packaging in HIV-1 and many other retroviruses is known (*1*), although some important details are still missing. Gag and gag-pol precursor proteins bear nucleocapsid (NC)[1] domains with two zinc-fingers that are crucial in selecting genomic RNA for packaging. About 1500−2000 of the precursors assemble at the inner membrane of the forming virion (*2*). The NC domain is later processed to a mature protein, with the 55 residue NCp7protein in HIV-1 illustrated in Figure 1. The mature NCp7 has essential roles in many other viral processes as well. Drugs to attack the nucleocapsid have the potential to interfere with critical functions at many stages of the viral life cycle (*3*−*9*).

NC-domains bind to a RNA region in the 5′-leader, referred to as the packaging signal, selecting two unspliced RNA molecules per virion. The packaging signal is diffuse in HIV-1, with several RNA−NC interactions responsible for discrimination of genomic RNA from the huge excess of other RNA molecules present in the cell. The major packaging domain resides in a region containing four stem-loops, SL1-4, located between nucleotides 240−360 (*10*, *11*). SL3 is the primary determinant of specificity, but other regions also contribute (*12*−*15*). Each loop in the major packaging domain has two or three unpaired G-residues and is a favorable candidate for interaction with NC zinc-fingers; two G-residues in close proximity appear to be required for strong interaction with the fingers (*16*−*21*). We refer to these as G_2-loci. The SL3 tetraloop is weakly ordered in the absence of protein (*22*, *23*) but makes specific interactions in the complex with NCp7 (*16*). The construct used in this work is shown in Figure 2, with the significant G-residues being the second and fourth bases in the loop. These correspond to positions 318 and 320 in the 9.2 kB HIV-1 genome.

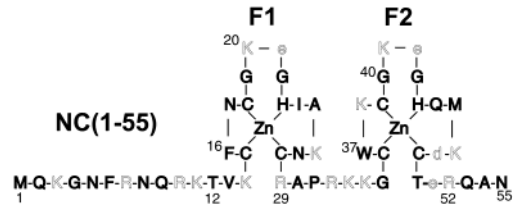

FIGURE 1: HIV-1 NCp7 protein. Residues carrying a charge at neutral pH are shown in outlined letters (positive = capitals, negative = lower case). Each cysteine side chain is present as a thiolate, so Zn_2·NCp7 (1−55) carries a +9 charge at neutral pH.

A high-resolution structure of SL3 with NCp7 provides many details of the interactions that determine selectivity (*16*) (Figure 3). G318 and G320 use their O6 and N1H atoms to make hydrogen bonds to backbone amides and carbonyls, each in a different zinc finger. The H-bond and hydrophobic

† Supported in part by Grant 02740-30-RGT from the American Foundation for AIDS Research (P.N.B. and B.S.H.), NIH Grant GM32691 (P.N.B.), and Syracuse University.

* Corresponding authors. (P.N.B.) E-mail: pnborer@syr.edu; tel: 315 443-5925. (B.S.H.) E-mail: bshudson@syr.edu; tel: 315 443-5805; fax: 315 443-4070.

[1] Abbreviations: NCp7, the 55 amino acid nucleocapsid protein of HIV-1, Human Immunodeficiency Virus; SL1, SL2, SL3, and SL4, short stem-loop segments of RNA from the 5′-leader region of HIV-1; NC, nucleocapsid; SDS−PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; PEG-8000, poly(ethylene glycol) average molecular weight 8000; concentrations are denoted by bold italic font (e.g., $\boldsymbol{P}_t$ indicates the total concentration of protein); W, tryptophan; W37, tryptophan residue 37 of NCp7; μ, ionic strength.

10.1021/bi026307n CCC: $22.00
Published on Web 11/27/2002

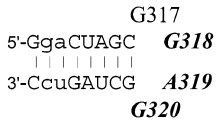

FIGURE 2: 20mer SL3 construct used in this study. The native tetraloop is shown; bases 318−320 were varied. A 16mer with a shortened stem was also examined in which the residues shown in lower case are absent.

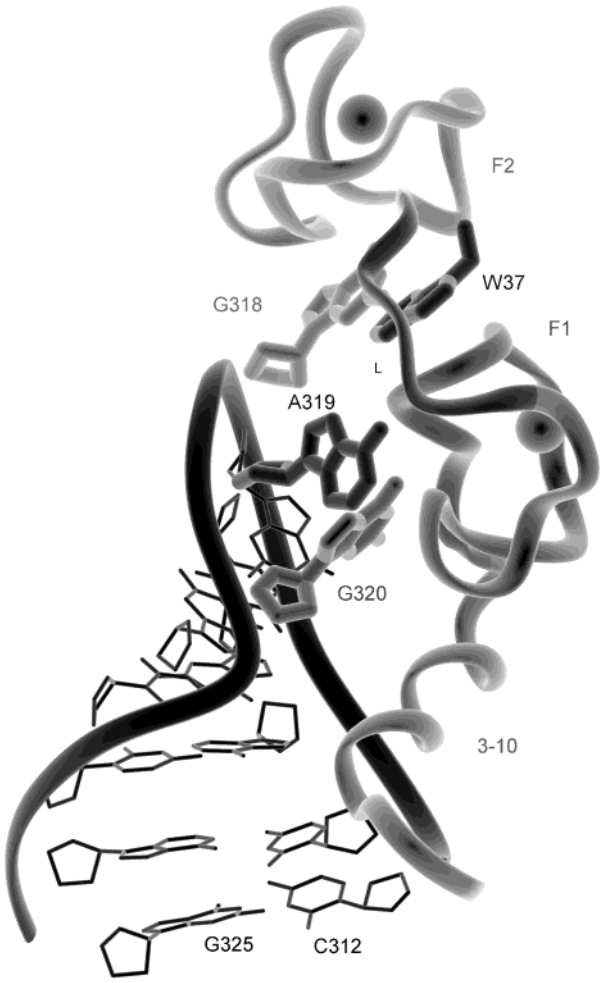

FIGURE 3: Detailed view of the SL3−NCp7 complex (*16*). G318 and G320 bind the two zinc-fingers in very similar fashion. W37 of the second finger stacks on G318; zinc ions are denoted as spheres (adapted from PDB 1A1T).

contacts involving these G-bases are nearly equivalent in each of the fingers. The figure also illustrates the intimate stacking between the G318 base and Trp-37 of the protein. This provides a sensitive assay for binding via the quenching of W37 fluorescence by the nucleobase. We used the assay to show that each of the G_2-loci in oligomer constructs of SL1-4 form 1:1 complexes with NCp7 at 0.2 M NaCl (*21*). Dissociation constants ranged from 23 to 320 nM for the wild-type loops. We also showed that a 154mer RNA containing all four stem-loops binds at least three NCp7 proteins. Nonspecific interactions between these highly charged molecules contribute heavily to affinity at lower ionic strength where most previous studies had been conducted (*10, 18, 20, 24*).

A next important step is to vary the sequence of the key loop bases in SL3. There are 256 possible combinations of four residues in the SL3 loop that could be examined for differences in affinity with NCp7. However, G317 does not make a hydrogen bond or hydrophobic contacts with the protein in the SL3−NCp7 complex (*16*), nor does the corresponding base, G289, in the SL2−NCp7 complex (*25*). Therefore, only the 64 possible variants of the loop positions GNNN using A, C, G, and U were considered. Affinities for 35 of these sequences are presented here.

EXPERIMENTAL PROCEDURES

Materials. The 55 amino acid NCp7 protein and RNA stem-loops (Dharmacon Research Inc., Lafayette, CO) were prepared, and concentrations were estimated as described previously (*21*). It was found that the protein can be stored for several weeks without significant degradation in a 25 mM acetate buffer, pH 5, with 10 mM dithiothreitol, 0.05 mM EDTA, 25 mM NaCl, and 0.1 mM $ZnCl_2$ at 4 °C. As a positive control on the quality of the protein sample, fluorescence titrations with SL3 were repeated daily.

Protein Binding Assay. Assay conditions were the same as described previously (*21*). The fluorescence titration curves were fitted to a model assuming 1:1 stoichiometry for the ratio of NCp7 to bound nucleic acid. The titration curves were fitted to the following equation:

$$(I - I_\infty)/I_0 = \{-(\boldsymbol{R}_t - \boldsymbol{P}_t + K_d) + [(\boldsymbol{R}_t - \boldsymbol{P}_t + K_d)^2 + 4\boldsymbol{P}_t K_d]^{1/2}\}/2\boldsymbol{P}_t$$

where $\boldsymbol{R}_t$ and $\boldsymbol{P}_t$ are the total RNA and protein concentrations, respectively, I_0 is the intensity at $\boldsymbol{R}_t = 0$, and I_∞ is the limiting intensity at saturation. (The negative sign on the first term on the right was omitted in our previous paper (*21*), although the calculations were done correctly. The right side of the equation is just the fraction of free protein.) In this work we set $I_\infty = 0$ except for tight-binding complexes (see Results). K_d was optimized for each titration to minimize the sum of the squared deviations between the theoretical and the experimental curves (*21*). The NCp7 concentration was fixed at 0.3 μM, and titrations were extended at least to $\boldsymbol{R}_t/\boldsymbol{P}_t = 2$; these fluorescence titrations are very sensitive in the nanomolar to micromolar concentration range.

RESULTS

SL3 Tetraloop Variants. Titration curves are shown in Figure 4 for SL3(GNNN) variants where positions 318−320 were AUA, GAC, CUG, UUG, GGU, GAG (wild type), and GUG. The affinities of these variants cover more than 3 orders of magnitude, with K_d values ranging from 8 nM to 20 μM.

Dissociation constants are reported in Table 1, assuming a 1:1 complex. No evidence for other stoichiometries is indicated by the data at 0.2 M NaCl (*21*). However, complexes do form with non-1:1 stoichiometries between these highly charged molecules at low salt concentrations (unpublished results). The binding isotherms for low-affinity complexes were not extended to saturation because the nature of the complex may change in the presence of a large excess of RNA. Therefore, K_d values greater than 1 μM are reported to only one digit accuracy.

There is also considerable uncertainty for tight-binding complexes. Only a few points in these titrations deviate from the infinite K_d^{-1} line, and the derived K_d values are more

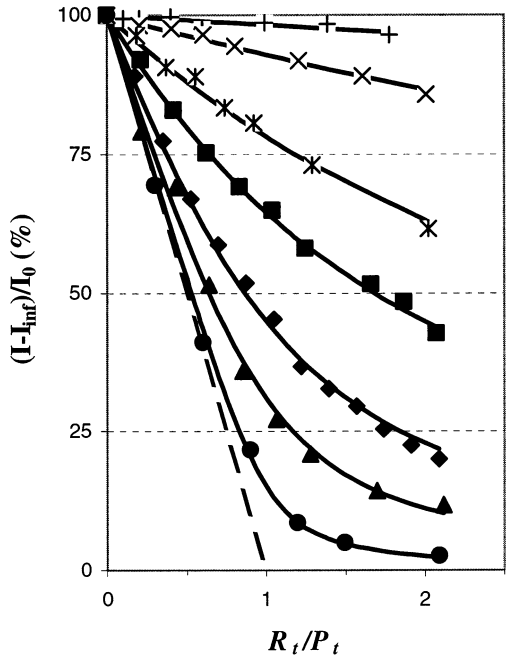

FIGURE 4: Fluorescence titrations of NCp7 (at concentration, P_t = 0.3 μM) with seven GNNN-loop variants of SL3. Total RNA concentration, R_t, increases left to right. Optimized fits (solid lines) for 1:1 complexes have K_d values listed in Table 1 (see Experimental Procedures and Results for details of fitting). Data points are as follows: +, NNN = AUA; ×, GAC; *, CUG; square, UUG; diamond, GGU; triangle, GAG; and circle, GUG. The dashed line is for a 1:1 complex with an infinite binding constant (K_d^{-1}).

affected than others by the assumption of zero residual fluorescence. The tight-binding GUG loop is a good example. If one assumes that $I_\infty = 0$, the best-fit $K_d = 10$ nM, and the points at the end of the titration lie above the best-fit line. The curve displayed in Figure 4 assumes that the complex has residual fluorescence at 2% of I_0. In this analysis, K_d differs by 20% (8 nM); when its titration was extended to $R_t/P_t = 3$, the residual intensity is ~3% of I_0. This is a reasonable correction: at $K_d = 8$ nM, $P_t = 0.3$ μM and $R_t/P_t = 3$; the fraction of free protein should be 1%. As can be seen in Figure 4, this correction produces an excellent fit. Thus, we have corrected the other tight-binding loops assuming 2% residual fluorescence in their complexes. This correction has very little effect when $K_d > 30$ nM.

Table 1 also reports affinities for NCp7 relative to SL3 in 0.2 M NaCl. The table is arranged so the strongest interactions are near the top. The first of the four groups contains data for all 16 GNNG loops. Note that all of the GGNG loops (top row) bind strongly. The other possible loop sequences cause a considerable reduction in affinity. Data were collected for 26 of 32 possible variants where position 320 is G or U. This was the most important of the tetraloop positions, with the base, G, contributing the largest amount to NCp7 affinity, followed by U. Given that A and C offer little toward stabilizing the complex, it was deemed unnecessary to measure K_d for the missing entries in Table 1.

Other RNA and DNA Oligomers. Several other nucleic acids were also analyzed. These include a 16mer SL3−RNA construct, without the GA/UC doublet near the terminus of the 20mer stem (Figure 2). The 16mer and 20mer molecules have the same K_d within experimental error (~30 nM). Two 20mer DNA molecules were also studied; the exact analogue,

Table 1: K_d (nM) and Relative Affinity (%) for 35 SL3 Variants[a]

G320

	G319		U319		A319		C319	
G318	18	***200***	8	***400***	28	***100***	65	***40***
U318	150	***20***	350	8	770	4	280	10
A318	140	***20***	290	10	420	7	210	10
C318	360	***8***	840	3	810	3	310	9

U320

	G319		U319		A319		C319	
G318	110	***30***	740	4	940	3	3000	1
U318	480	***6***	2000	2	3000	1	4000	0.6
A318	330	***8***	− −		− −		− −	
C318	580	***5***	− −		− −		− −	

A320

	G319		U319		A319		C319	
G318	− −		1000	2	1000	2	− −	
U318	− −		− −		− −		− −	
A318	2000	***1***	20 000	0.2	5000	0.5	3000	0.9
C318	− −		− −		− −		− −	

C320

	G319	U319		A319		C319
G318	− −	2000	1	4000	0.8	− −
U318	− −	− −		− −		− −
A318	− −	− −		3000	1	− −
C318	− −	− −		− −		− −

[a] Each cell contains K_d (left) and relative affinity (RA) (right). RA = 100% for G318-A319-G320; K_d was not measured for cells marked with − −. Residues 320 and 319 are shown at the top of each group, with 318 at the left; all sequences contain G317. RA values for loops containing G-N-G are shown in bold italic font. Entries where $K_d > 1000$ have been rounded to one significant digit; uncertainties are estimated at ± 20−30% for the other entries (*21*). All RA values were calculated from the best-fit K_d, then rounded to one significant digit.

SL3−DNA, had a substantially higher K_d than SL3−RNA (RA = 12%), and when an abasic site (s) was introduced at position 319, RA decreased to 4%. The single-stranded sequences, d(AACGCC), d(s_3GAG), d(s_5GAG), and d(s_6GAG), all had RA < 1%. Finally, previous work showed that the UNCG tetraloops, UUCG and UACG, exhibit affinities <1% of SL3−RNA (*21*).

DISCUSSION

High Affinity. The dominant requirement for high affinity is for the G318-N-G320 sequence (Table 1). The most important of these is G320, even though it is not the base that stacks on W37 (Figure 3). Replacing this base in the G-A-G320 wild type with the others leads to a dramatic decrease in relative affinity: G-A-U320, 3%; G-A-A320, 2%; and G-A-C320, 0.8% (Table 1). The second important requirement is for G318. Here U318-A-G, 4%; A318-A-G, 7%; and C318-A-G, 3%. The effects of substitutions at position 319 are highly variable. It is interesting that two of the G-N319-G loops (where N = U, G) bind more strongly than the wild-type G-A-G sequence. From the standpoint of drug design/discovery, SL3(GGUG) (with a higher affinity than the wild type) already represents an attractive first step, K_d ~8 nM. Probably the native interaction with GAG has evolved to satisfy other pressures beyond the stability of the complex with NC, an idea that is reinforced by sequence conservation.

A survey of HIV-1 sequences deposited in Genbank shows that few mutations are tolerated in SL3. There are 12 G317A, zero G318N, two A319U, one G320A, and no other mutations among 489 SL3-containing sequences (Y. Lin, and P. N. Borer, unpublished). The accommodation of A-residues at 317 corresponds with that base being stacked upon the base-paired stem (Figure 3) while not making specific H-bonding contacts with the protein (*16*). G- and A-stacking should be similar in geometry and loop-stabilization. In the G_2-locus for NC binding, G318 is absolutely conserved, and only one instance of a G320A change has been recorded. The rate of sequencing errors is probably higher than this 1/489 occurrence. The two A319U changes may also be less than the likelihood of errors in the database, but it is intriguing that these changes would enhance the stability of SL3−NC complexes.

Low Affinity. Given the very low affinities of some of the constructs (e.g., GAUA, 0.2%; GAAA, 0.5%; and GUCU, 0.6%), nonspecific binding to the stem and the loop must contribute little to stability of the complexes at 0.2 M NaCl. Beyond the lack of direct quenching of Trp-fluorescence by SL3(GAUA), we have previously reported that it does not compete with wild-type SL3(GGAG) for the W37 binding site (*21*). Among the weakly binding variants, most fall in the category where all of the NNN sites have C, A, or U (seven of the variants collected in Table 1, RA < 2%). This group, which has only G317 as a G-base, includes 20 of the 29 variants not assayed, so it is unlikely that any of those will have substantial affinity for NCp7 and will, for this reason, be uninteresting in the context of drug design. By contrast, there are 17 constructs with RA ≥ 5%, and all of them have at least two G-bases. Thus, a large part of the binding free energy at physiological ionic strength must be because of interaction of the protein with the loops containing G_2-loci. Others have noted that affinities of NCp7 for single-stranded DNA oligomers have the order of G > U > A ≅ C (*18*, *19*).

Intermediate Affinity. Our data also show that combinations of two G-residues or G- and U-residues can lead to complexes with moderately high stabilities. (The array of H-bond donors and acceptors on G are more similar to U than the other bases; U-O4 and U-N3H are in nearly equivalent positions to the G-O6 and G-N1H that make H-bonds with atoms in the zinc-fingers (*16*, *26*).) It has been reported that G-N-G oligomers and $(dT\text{-}dG)_n$ sequences bind to the mature NC more tightly than do other DNA sequences (*18*, *19*). Furthermore, addition of $(dT\text{-}dG)_n$ to NCp7 inhibits the reaction of Cys-49 with *N*-ethylmaleimide and other sulfhydryl reactive agents (*27*). Several other G/U-containing loops merit discussion.

Table 1 highlights the affinity values in bold italic font for GNNN sequences that contain GNG sequences. Obviously, this includes the tight binding GGNG loops in the first row of the table. In addition, the first column has GNGN, and all but one of these entries has affinity ≥5% of the wild-type loop. Three of these have affinities ≥20%, and this may be a composite of G_2- or UG-loci, viz., G-N-G (GUGG, GAGG, and GGGU), U-N-G (GUGG), G-G-U (GGGU), and G-N-N-G (GUGG, GAGG). Note also that each of the top 16 entries in the table has the possibility of forming a weakly bound G-N-N-G complex. These and several other intermediate affinity loops include G-G, U-G, G-U, or U-U adjacencies that may also contribute to stability. If so, the RNA loop−protein complex may jump between several weakly bound states, residing primarily in the strong GGNG sites when they are available.

Other Loops and Stems. Other interesting sequences contain stable UNCG or GNRA tetraloops (*28*−*30*) that lower the free energy of the unbound RNA. We have previously reported the low affinities of UUCG and UACG loops (*21*). SL3(GAGA) is interesting for two reasons. First, to have high affinity the sequence should contain a G-N-G locus, along with GUGG, GAGG, and GGGU (previous paragraph), yet its affinity is only 1% of the wild type. This may be attributable to forming the special GNRA loop where there is a sheared G-A base pair occupying important H-bond sites on G317, and G319 is stacked between A318 and A320 (*30*, *31*). A second reason is that the GAGA loop also forms in SL4, where we showed that a 14mer construct binds NCp7 with K_d = 320 nM in 0.2 M NaCl (*21*). Placing the GAGA loop on the SL3 stem increases K_d to ∼2000 nM. Another report with two SL4 constructs in 0.05 M ionic strength reported low affinity for NCp7 (*32*). Together these reports suggest that the nature of the stem can modulate the stability of loop−NCp7 complexes. This might occur by specific H-bonding interactions of the protein with stem base pairs near the loop (*16*) or by providing a more favorable path to the NC-bound loop form in less stable stems (*31*). This path may be different by only subtle energy effects—a factor of 7 difference in K_d corresponds to <1 kcal/mol in free energy.

Another example of the modulation of loop stability by the stem is provided by comparing SL2(GGUG) (*21*) to SL3-(GGUG). Our previous report gave K_d = 23 nM for SL2, and Table 1 indicates 8 nM for SL3(GGUG). Now, placing the GGUG loop on the SL3 stem decreases K_d by a factor of 3. The major difference between the SL3(GGAG) and the SL2(GGUG) complexes is that the 3-10 helix moves from one side of the stem to the other (*16*, *26*). Again, the difference in K_d values is not large, corresponding to less than 1 kcal/mol in $\Delta G°$.

The apical loop of SL1 is another interesting example. It has a nine-residue loop containing a GCGCGC sequence, therefore possessing two G_2-loci. Our previous work showed that SL1 constructs bind NCp7 in a 1:1 complex with K_d = 100 nM (*21*). Thus, protein binding to one of the G_2-loci excludes binding the other, as might be expected. Table 1 shows that SL3(GGCG) has a significantly lower K_d value (65 nM), indicating that the GCG-locus on the SL3 stem is a more favorable environment for binding.

Considering the secondary effects described in the previous few paragraphs, it is perhaps not surprising that attempts to derive additive sequence-dependent free energy contributions for the SL3(GNNN) complexes with NCp7 have not succeeded. It is likely that still other effects may contribute to the stabilities. The overwhelming criterion for high stability complexes is the presence of an unpaired G_2-locus capable of adopting a structure similar to that in Figure 3.

Other RNA and DNA Oligomers. The 16mer SL3−RNA construct (Figure 2) is interesting because it has K_d = 35 nM, virtually identical to the 20mer that provides the basis for comparison in Table 1. This supports our assertion that nonspecific binding to the stem contributes little to stability in 0.2 M NaCl solutions. It also suggests that it may be possible to make even shorter SL3-mimics that could be

prototypes for antiNC agents. Further exploration in this area is warranted.

Two 20mer DNA molecules were also studied. Both the exact analogue, d[SL3(GGAG)], and d[SL3(GGsG)] have only 12 and 4% affinity relative to ribo-SL3(GGAG), respectively (s is an abasic site). Thus, the affinity for DNA oligomers appears considerably lower in 0.2 M NaCl as compared to tight-binding RNA loops. Several d(GNG)-containing single strands have been reported to possess relatively high affinity for NCp7 (*18*, *19*). However, none of the single-stranded sequences, d(AACGCC), d(s_3GAG), d(s_5GAG), and d(s_6GAG) possessed even micromolar affinity for NCp7 at 0.2 M NaCl.

Other Related Work. In a study by Fisher et al. (*18*), oligonucleotides were bound to NCp7 (1−55) and examined by surface plasmon resonance. The SPR experiments were conducted at ∼0.2 M ionic strength, μ. Aliquots of increasing NCp7 concentration were washed over DNA oligomers immobilized on SPR chips, and the increase in mass was determined. SPR was utilized to initially demonstrate that NCp7 strongly preferred binding a G-rich 28mer single strand over its complement and the double stranded DNA. Among other interesting sequences, they reported results for several homo-oligomers, (dT-dG)$_n$, and ribo-(U-G)$_4$. The alternating TG and UG molecules bound most tightly, at 1:1 stoichiometries for $n \leq 4$, at R_1P_2 stoichiometry for (dT-dG)$_5$, and at R_1P_4 for (dT-dG)$_{10}$. This is in line with other studies that suggest a site size of ∼6−8 nucleotides for NCp7 binding single-stranded nucleic acids (*33*−*36*). The affinity for (U-G)$_4$ was comparable to that for (dT-dG)$_n$.

Vuilleumier et al. (*19*) studied many short single-stranded DNA oligomers using fluorescence titrations very similar to those analyzed in our work. Most of the work was conducted with a 72-residue version of NCp7 having a similar sequence to Figure 1, adding 17 residues at the C-terminus and substituted at K3R, T12N, I24T, and K26R (this peptide would have a +12 charge at neutral pH); some of the work was also done with a truncated 12−53 version (+6 charge). Binding constants were reported from fluorescence titrations at $\mu = 0.12$ M, but they also conducted titrations at higher salt concentrations. The binding analysis specifically included a plateau value, Q_{max} (= 100 − I_∞), so that I_∞ varied over a range from ∼30 to 50% of I_0 for a collection of DNA and RNA hexanucleotides. Among 20 molecules examined, K_d ranged from 140 nM for d(TGTGCC) to ∼100 μM for A- and C-rich molecules. If the salt dependence is linear in log[Na^+] (*19*, *21*), one can estimate that K_d should be ∼900 nM for d(TGTGCC) at $\mu = 0.2$ M. The d(s_nGAG) molecules examined in the present report bind less well, $K_d > 1$ μM, than d(TGTGCC) but are in the same regime of stability as several others in the Vuilleumier et al. report. We chose not to pursue these micromolar affinity molecules as antiNC drug leads, preferring to focus on nanomolar derivatives of SL3.

Other reports have introduced RNA aptamers with low nanomolar affinities as promising leads toward antiNC agents (*37*−*41*). It is interesting that many of these aptamers have no obvious G_2-loci, and if they bind via the two zinc-fingers, combinations of unpaired G- and U-residues probably are the basis for interaction.

The Vuilleumier et al. (*19*) report analyzed DNA analogues of SL3. Again, quenching was not complete at amounts of DNA that saturate the protein (I_∞ = 10−25%). Affinities were higher for the wild-type sequence, d[SL3(GGAG)] (K_d = 70 nM at $\mu = 0.12$ for the 1−72 protein), than the other loops examined (AAAG), (AGAA), and (AAAA), (60, 50, and 20% RA, respectively), but the loop sequence effects were much less dramatic than in our RNA studies. The 12−53 protein binds less tightly at 0.12 M salt, as might be expected, K_d = 1000 nM for d[SL3(GGAG)]. The salt-dependence was steep; at 0.2 M the data predicts that all of the molecules have $K_d > 18$ μM. It is difficult to compare this work with ours, given that the proteins have different sizes and charges, the nucleic acid molecules are different, and a large I_∞ value was included in the analysis.

CONCLUSION

Our work continues to explore the diversity of interactions between NCp7 and RNA loops using variants of SL3. Our previous paper surveyed the affinities of the wild-type interaction sites for NCp7 in the major packaging domain of the 5′-leader RNA (*21*). To design and evaluate antiNC drugs, it is useful to know the affinities the protein has for its natural substrates under physiological conditions and to probe the nature of G_2-loci by systematic variation of the sequence. The current results also add to our understanding of RNA−protein interactions.

We have established a close correlation between structural features and a rapid, sensitive technique to evaluate the diversity of RNA−NCp7 interactions. Given the 1:1 nature of the equilibrium and the small (∼2%) corrections for residual fluorescence in the complex, the signal in the assay is almost directly proportional to the population of the unbound protein. No large corrections are needed, as is common in filter-binding assays (e.g., refs *38* and *42*), to normalize the measured signal to a saturation level that is a small fraction of the RNA input. Nor is it necessary to add tRNA to suppress nonspecific binding, which would be an especially questionable practice since NCp7 has a documented affinity for single strands. Either of these practices would diminish confidence in K_d values derived for the NCp7−SL3 system.

It is critically important to control the salt concentration in attempting to measure NC−RNA affinity. Measurements below 0.2 M NaCl give unusual stoichiometries and titrations that vary substantially depending on the detailed protocol (ref *21* and unpublished results). Furthermore, nonspecific interactions between these highly charged molecules contribute heavily to the binding at low ionic strength where most previous studies have been done. By simply conducting the binding assay at 0.2 M NaCl (near physiological conditions; in blood the ionic strength is ∼0.18 M ignoring contributions of charged macromolecules (*43*)), we can assess the relative affinity of NCp7 binding to the loop G_2-loci.

REFERENCES

1. Coffin, J. M., Hughes, S. H., and Varmus, H. E. (1997) *Retroviruses*, Cold Spring Harbor Lab Press, Plainview, NY.
2. Vogt, V. M., and Simon, M. N. (1999) *J. Virol. 73*, 7050−5.
3. Darlix, J. L., Cristofari, G., Rau, M., Pechoux, C., Berthoux, L., and Roques, B. (2000) *Adv. Pharmacol. 48*, 345−72.
4. Fu, W., Gorelick, R. J., and Rein, A. (1994) *J. Virol. 68*, 5013−8.

5. Prats, A. C., Sarih, L., Gabus, C., Litvak, S., Keith, G., and Darlix, J. L. (1988) *EMBO J. 7*, 1777−83.
6. Druillennec, S., Caneparo, A., de Rocquigny, H., and Roques, B. P. (1999) *J. Biol. Chem. 274*, 11283−8.
7. Lener, D., Tanchou, V., Roques, B. P., Le Grice, S. F., and Darlix, J. L. (1998) *J. Biol. Chem. 273*, 33781−6.
8. de Rocquigny, H., Caneparo, A., Delaunay, T., Bischerour, J., Mouscadet, J. F., and Roques, B. P. (2000) *Eur. J. Biochem. 267*, 3654−60.
9. Laughrea, M., Shen, N., Jette, L., Darlix, J. L., Kleiman, L., and Wainberg, M. A. (2001) *Virology 281*, 109−16.
10. Clever, J., Sassetti, C., and Parslow, T. G. (1995) *J. Virol. 69*, 2101−9.
11. McBride, M. S., and Panganiban, A. T. (1996) *J. Virol. 70*, 2963−73.
12. Hayashi, T., Shioda, T., Iwakura, Y., and Shibuta, H. (1992) *Virology 188*, 590−9.
13. Hayashi, T., and Iwakura, Y. (1993) *Tanpakushitsu Kakusan Koso 38*, 779−83.
14. Clever, J. L., and Parslow, T. G. (1997) *J. Virol. 71*, 3407−14.
15. Das, A. T., Klaver, B., and Berkhout, B. (1998) *J. Virol. 72*, 9217−23.
16. De Guzman, R. N., Wu, Z. R., Stalling, C. C., Pappalardo, L., Borer, P. N., and Summers, M. F. (1998) *Science 279*, 384−8.
17. South, T. L., and Summers, M. F. (1993) *Protein Sci. 2*, 3−19.
18. Fisher, R. J., Rein, A., Fivash, M., Urbaneja, M. A., Casas-Finet, J. R., Medaglia, M., and Henderson, L. E. (1998) *J. Virol. 72*, 1902−9.
19. Vuilleumier, C., Bombarda, E., Morellet, N., Gerard, D., Roques, B. P., and Mely, Y. (1999) *Biochemistry 38*, 16816−25.
20. Maki, A. H., Ozarowski, A., Misra, A., Urbaneja, M. A., and Casas-Finet, J. R. (2001) *Biochemistry 40*, 1403−12.
21. Shubsda, M. F., Paoletti, A. C., Hudson, B. S., and Borer, P. N. (2002) *Biochemistry 41*, 5276−82.
22. Pappalardo, L., Kerwood, D. J., Pelczer, I., and Borer, P. N. (1998) *J. Mol. Biol. 282*, 801−18.
23. Zeffman, A., Hassard, S., Varani, G., and Lever, A. (2000) *J. Mol. Biol. 297*, 877−93.
24. Summers, M. F., Henderson, L. E., Chance, M. R., Bess, J. W., South, T. L., Blake, P. R., Sagi, I., Perez-Alvarado, G., Sowder, R. C., Hare, D. R., and Arthur, L. O. (1992) *Protein Sci. 1*, 563−74.
25. Amarasinghe, G. K., De Guzman, R. N., Turner, R. B., and Summers, M. F. (2000) *J. Mol. Biol. 299*, 145−56.
26. Amarasinghe, G. K., De Guzman, R. N., Turner, R. B., Chancellor, K. J., Wu, Z. R., and Summers, M. F. (2000) *J. Mol. Biol. 301*, 491−511.
27. Chertova, E. N., Kane, B. P., McGrath, C., Johnson, D. G., Sowder, R. C., II, Arthur, L. O., and Henderson, L. E. (1998) *Biochemistry 37*, 17890−7.
28. Molinaro, M., and Tinoco, I. (1995) *Nucleic Acids Res. 23*, 3056−63.
29. Allain, F. H., and Varani, G. (1995) *J. Mol. Biol. 250*, 333−53.
30. Jucker, F. M., Heus, H. A., Yip, P. F., Moors, E. H., and Pardi, A. (1996) *J. Mol. Biol. 264*, 968−80.
31. Kerwood, D. J., Cavaluzzi, M. J., and Borer, P. N. (2001) *Biochemistry 40*, 14518−29.
32. Amarasinghe, G. K., Zhou, J., Miskimon, M., Chancellor, K. J., McDonald, J. A., Matthews, A. G., Miller, R. R., Rouse, M. D., and Summers, M. F. (2001) *J. Mol. Biol. 314*, 961−970.
33. Dib-Hajj, F., Khan, R., and Giedroc, D. P. (1993) *Protein Sci. 2*, 231−43.
34. Khan, R., Chang, H. O., Kaluarachchi, K., and Giedroc, D. P. (1996) *Nucleic Acids Res. 24*, 3568−75.
35. Mely, Y., de Rocquigny, H., Sorinas-Jimeno, M., Keith, G., Roques, B. P., Marquet, R., and Gerard, D. (1995) *J. Biol. Chem. 270*, 1650−6.
36. Urbaneja, M. A., Kane, B. P., Johnson, D. G., Gorelick, R. J., Henderson, L. E., and Casas-Finet, J. R. (1999) *J. Mol. Biol. 287*, 59−75.
37. Allen, P., Collins, B., Brown, D., Hostomsky, Z., and Gold, L. (1996) *Virology 225*, 306−15.
38. Berglund, J. A., Charpentier, B., and Rosbash, M. (1997) *Nucleic Acids Res. 25*, 1042−9.
39. Lochrie, M. A., Waugh, S., Pratt, D. G., Clever, J., Parslow, T. G., and Polisky, B. (1997) *Nucleic Acids Res. 25*, 2902−10.
40. Clever, J. L., Taplitz, R. A., Lochrie, M. A., Polisky, B., and Parslow, T. G. (2000) *J. Virol. 74*, 541−6.
41. Kim, S. J., Kim, M. Y., Lee, J. H., You, J. C., and Jeong, S. (2002) *Biochem. Biophys. Res. Commun. 291*, 925−31.
42. Hall, K. B., and Stump, W. T. (1992) *Nucleic Acids Res. 20*, 4283−90.
43. Kratz, A., and Lewandrowski, K. B. (1998) *N. Engl. J. Med. 339*, 1063−72.

BI026307N