

# Identification of a binding site for the human immunodeficiency virus type 1 nucleocapsid protein

(RNA binding/dimerization)

KAZUYASU SAKAGUCHI\*, NICOLA ZAMBRANO\*, ERIC T. BALDWIN†, BRUCE A. SHAPIRO‡,  
JOHN W. ERICKSON†, JAMES G. OMICHINSKI§, G. MARIUS CLORE§, ANGELA M. GRONENBORN§,  
AND ETTORE APPELLA\*¶

\*Laboratory of Cell Biology and ‡Laboratory of Mathematical Biology, National Cancer Institute, §Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD 20892; and †Structural Biochemistry, PRI/DynCorp., National Cancer Institute–Frederick Cancer Research and Development Center, Frederick, MD 21702

Communicated by Bernhard Witkop, February 17, 1993 (received for review, December 2, 1992)

**ABSTRACT** The nucleocapsid (NC) protein NCp7 of human immunodeficiency virus type 1 (HIV-1) is important for encapsidation of the virus genome, RNA dimerization, and primer tRNA annealing *in vitro*. Here we present evidence from gel mobility-shift experiments indicating that NCp7 binds specifically to an RNA sequence. Two complexes were identified in native gels. The more slowly migrating complex contained two RNA molecules and one peptide, while the more rapidly migrating one is composed of one RNA and one peptide. Further, mutational analysis of the RNA shows that the predicted stem and loop structure of stem-loop 1 plays a critical role. Our results show that NCp7 binds to a unique RNA structure within the  $\psi$  region; in addition, this structure is necessary for RNA dimerization. We propose that NCp7 binds to the RNA via a direct interaction of one zinc-binding motif to stem-loop 1 followed by binding of the other zinc-binding motif to stem-loop 1, stem-loop 2, or the linker region of the second RNA molecule, forming a bridge between the two RNAs.

Type C retroviruses typically contain two copies of the single-stranded RNA genome in each virion (1). These two unspliced RNA molecules are linked together near their 5' ends by the dimer linkage structure (DLS). A specific cellular tRNA is annealed to each RNA at the primer binding site and serves as a primer during the reverse transcription of the RNA genome into DNA. Mutational analyses have shown that the assembly of the DLS is important for packaging of the viral-specific RNA. Packaging defects can be produced with genetic deletions within the  $\psi$  region, located between the lysine tRNA primer binding site and the start of the gag polyprotein coding sequences, or by mutations of codons within the *gag* gene (2, 3).

The human immunodeficiency virus type 1 (HIV-1) nucleocapsid (NC) protein, NCp7 contains two sets of a Cys<sub>3</sub>-His<sub>1</sub> sequence motif and binds zinc to form a unique ordered structure (4, 5). Deletion of one zinc finger or mutation of one cysteine to serine in either zinc finger of the HIV-1 NCp7 protein decreases RNA packaging (ref. 3; L. E. Henderson and R. Gorelick, personal communication). Also, deletions of the basic residues near the first zinc finger cause a genomic RNA packaging defect (6).

Several studies have shown that *in vitro* mature NC proteins show preferential binding to single-stranded nucleic acids with high affinity for the viral RNA (7, 8). HIV-1 NCp7 protein, as well as Rauscher and Moloney murine leukemia virus (Ra-MuLV and Mo-MuLV) NC proteins, mediates dimerization of retroviral RNA containing the encapsidation sequence  $\psi$  and annealing of tRNA to the primer binding site.

In the case of HIV-1, the interaction of NC with RNA was examined by using the crosslinking reagent *trans*-diamminedichloroplatinum(II). A 10-nucleotide (nt) fragment was identified as the NC binding site (9). Mutations in the zinc fingers or the flanking basic residues were shown to abrogate the high-affinity RNA binding *in vitro* as well as the RNA annealing activities of the NC proteins (6).

In this study, we report that a 44-nt RNA segment containing the HIV-1  $\psi$  region is folded into two stem-loop structures separated by a stretch of 11 nt. The 44-nt RNA oligomer (44-mer) contains the NCp7 binding site, which requires an intact stem and loop structure. The functional importance of this binding site for RNA dimerization is illustrated by a model for the DLS.

## MATERIALS AND METHODS

**Prediction of RNA Folding of HIV-1 Sequences.** The sequences of eight variants of HIV-1 (HIVRF, HIVHAN, HIVJRCSE, HIVSF2, HIVLA1, HIVHXB2R, HIVNL43, HIVNY5) were aligned according to the Los Alamos convention (10), and their secondary structures were predicted. Each sequence consisted of 300 bases symmetrically disposed about the  $\psi$  site and were folded independently. The method of Zuker (11) was used to minimize the free energy for the optimally and suboptimally folded structures, and free-energy values developed by Jaeger *et al.* (12) were used. For each of eight sequences, an optimal and 49 suboptimal folded structures were separated into morphologically equivalent groups by comparing the predicted secondary structures of the 50 nt that encompass the  $\psi$  region (13). The present method uses an updated version of the software and new energy rules that were not used by Harrison and Lever (14). The sequence of the linker mutant was designed by using an algorithm allowing for random mutations and clustering RNA structures with similar shape (B.A.S., unpublished data).

**Synthesis of HIV-1 p7-(1–55) Fragment and Related Peptides.** p7-(1–55) and related peptides were synthesized, cleaved, and purified as described (15).

**Preparation of Wild-Type and Mutant RNAs.** Double-stranded DNAs containing a phage T7 promoter and the RNA sequence were used as templates in the transcription reaction. Transcription was carried out as described by Milligan and Uhlenbeck (16). A mixture of UTP and 5-Br-UTP was used in the transcription reaction to obtain bromouridine-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: HIV-1, human immunodeficiency virus type 1; DLS, dimer linkage structure; NC, nucleocapsid; Mo-MuLV and Ra-MuLV, Moloney and Rauscher murine leukemia virus.

¶To whom reprint requests should be addressed at: Laboratory of Cell Biology, National Cancer Institute, Building 37, Room 1B04, Bethesda, MD 20892.

containing RNA. For 5'-end-labeling of RNA, nonradioactive transcripts were dephosphorylated by calf intestine alkaline phosphatase and then labeled by phage T4 polynucleotide kinase in the presence of [ $\gamma$ - $^{32}$ P]ATP.

**RNA Secondary Structure Determination.** Approximately 10 ng of the 5'-end-labeled wild-type 44-mer ( $10^4$  cpm) was heated for 1 min at 100°C in water, chilled on ice, and incubated for 10 min at room temperature in 250  $\mu$ l of RNA-binding buffer [40 mM Tris glutamate, pH 8.0/10 mM magnesium glutamate/100 mM potassium glutamate/25 mM NaCl/1 mM dithiothreitol/5% (vol/vol) glycerol] containing 0.25  $\mu$ g of yeast tRNA (Boehringer Mannheim) per  $\mu$ l. The 20- $\mu$ l aliquots were treated for 5 min at room temperature with one of the following RNases at two concentrations: 1.5 and 3.75 units of RNase I (Promega), 0.06 and 0.12 units of RNase V1 (United States Biochemical), 0.5 and 1  $\mu$ g of RNase A (Boehringer Mannheim) per ml, or 1 and 2 units of RNase T1 (Boehringer Mannheim).

**RNA-Binding Assay.** RNA-binding experiments were performed by mixing purified peptide with  $^{32}$ P-labeled RNAs in the RNA-binding buffer. Peptides were incubated in the binding buffer for 15 min at room temperature, and then heat-denatured RNAs were added to the peptide solution and incubated for 15 min at room temperature. The peptide-RNA complexes were analyzed by 8% nondenaturing polyacrylamide (acrylamide/methylenebisacrylamide ratio of 39:1) gel electrophoresis; the running buffer was 45 mM Tris borate/1 mM EDTA. Gels were prerun for 30 min at 150 V at 4°C, and electrophoresis was allowed to proceed for 2.5 hr at 250 V at 4°C. The gels were then dried and visualized by autoradiography.

**UV Crosslinking of p7-(1-55) with Bromouridine-Containing 44-nt RNA.** The labeled bromouridine-containing 44-mer was incubated with p7-(1-55) in the RNA-binding buffer. The complexes were resolved by electrophoresis on a nondenaturing gel and were irradiated in the gel with a 254-nm lamp for 15 min at 4°C. The wet gel was visualized by autoradiography at 4°C. Bands were cut out from the gel and denatured by boiling for 2 min in 125 mM Tris-HCl (pH 6.8) containing 1% SDS and 3 mM dithiothreitol. The slices containing crosslinked complexes were analyzed by 10% Tris-tricine PAGE (17).

## RESULTS

The eight different HIV-1 sequences used in this study were folded into 400 structures, and the morphology of the stems and loops in the vicinity of the  $\psi$  site was evaluated. The overwhelming majority of the predicted structures (group 1, 224 members) produced a pair of stems and loops separated by a single-stranded linker (Fig. 1A). Stem-loop 1 contains an unpaired adenine nucleotide and the 5' major splice junction, which is found within the GGUG loop sequence. This stem-loop is followed by 9-13 nt that are palindromic: 2-5 adenine nucleotides are followed by a run of 4-6 uracil nucleotides. The second stem-loop is shorter than the first and has the loop sequence GGAG. The second loop ends 20 nt 5' to the initiator AUG for the gag polyprotein.

The remaining predicted secondary structures could be grouped into three families. Group 2 included both stem-loop 1 and stem-loop 2 (107 members). These structures had additional base-pairing involving the run of uracils in the linker region that were not observed in group 1. Group 3 (38 members) included only stem-loop 1. Stem-loop 2 was disrupted by the base-pairing of the linker run of uracils with adenine outside the  $\psi$  region and near the 5' end of the 300-nt target. Group 4 consists of 29 members that are all variants of the structure predicted by Harrison and Lever (14). The predicted secondary structure included stem-loop 2 but indicated an alternative for stem-loop 1. In addition, the

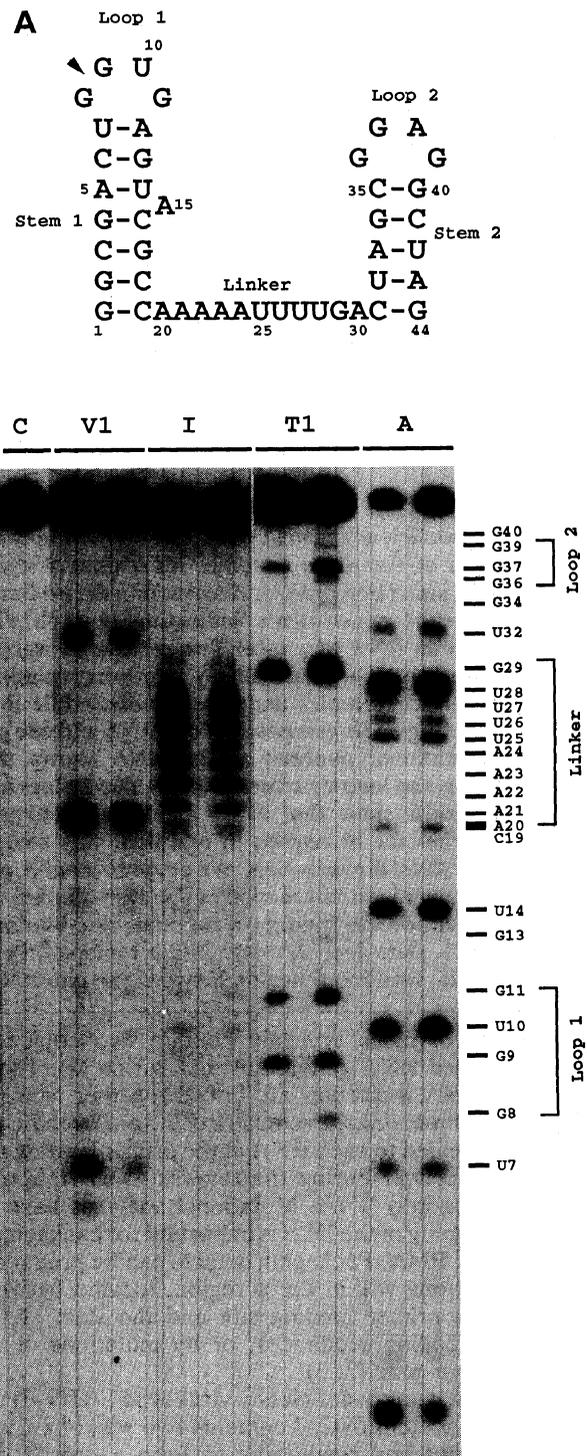


FIG. 1. Structure of 44-nt RNA from the  $\psi$  region of HIVHXB2R. (A) Computer-predicted secondary structure of the 44-mer. Nucleotides are numbered from the 5' end. The arrowhead indicates the viral splice site. (B) RNase-probing analysis of the 44-mer labeled at the 5' end. Ribonucleases V1, I, T1, and A were used for probing the 44-mer. Lanes L and C denote the alkaline hydrolysis ladder and the control RNase reaction without ribonuclease, respectively. The loop and linker regions are indicated.

linker region uracils form base pairs with nucleotides 3' to stem-loop 2 and extend this stem-loop. Stem-loop 1 was observed in all cases in groups 1-3, which account for 92% of the predicted folds. Stem-loop 2 was only unobserved in group 3 and so was present in 90% of the folds. These data provide a strong basis for our prediction shown in Fig. 1A.

To confirm the existence of the predicted structures, we synthesized a 44-nt RNA corresponding to the  $\psi$  region of the

HIVHXB2R isolate (Fig. 1A). RNase probing of the synthetic RNA was carried out with RNase I, V1, A, or T1 (Fig. 1B). The results obtained support the presence of a stem and loop structure as predicted. RNase I digestion yields a ladder that derives from the single-strand linker region, whereas RNase V1, which is specific for the double strand, shows the bands characteristic of the predicted stem structure. Both RNase A and T1 unequivocally identified the sequence of the loops.

Previously, it has been shown that the NCp7 induces dimerization of HIV-1 RNA fragments (6, 9). Therefore, we have investigated the binding of a 55-amino acid fragment of NCp7, p7-(1-55) to *in vitro* transcribed RNA by gel mobility-shift analysis. As shown by the data in Fig. 2, p7-(1-55) binds specifically to the 44-nt RNA with high affinity in the presence of zinc ions, yielding two complexes, C1 and C2. Use of radioactive <sup>65</sup>Zn showed that both complexes contained metal ion, confirming the presence of peptide in both and proving that they are not just conformers of RNA (data not shown). Gel mobility shifts performed with a shorter fragment of NCp7 (positions 13-51) containing both zinc-binding motifs showed no formation of the C2 complex, while the C1 band was apparent only at a high concentration of peptide. Further, neither an N-terminal fragment of NCp7 (positions 1-14) nor a different HIV-1 protein, p6, elicited complex formation; addition of zinc ion alone did not yield any complexes. These results suggest that C1 and C2 complex formation requires the N-terminal 55 residues of NCp7 in the presence of zinc.

RNA fragments containing the  $\psi$  region can also spontaneously dimerize in the absence of NCp7 (9). Fig. 3A shows that the 44-nt RNA used in the present study is able to form a dimer in the absence of NCp7 but much less efficiently. This RNA dimer migrates more rapidly than C2 as expected, since the peptide-containing complex is of larger molecular mass. UV crosslinking of bromouridine-containing 44-nt RNA in the presence of p7-(1-55) further supported the notion that C1 and C2 are peptide-containing RNA complexes. Analysis of crosslinked product by SDS/PAGE (Fig. 3B) revealed that the single bands obtained from C1 and C2 complexes *in situ* had an apparent molecular mass of  $\approx$ 20 kDa, suggesting that for both complexes at least one peptide is crosslinked to a single RNA. The crosslinked C1 complex molecular mass derived from SDS/PAGE is close to the estimated molecular mass of this complex on native gels. This contrasts with our finding that the C2 complex from native gels is 35.5 kDa. This value for C2 has been demonstrated by analytical ultracentrifugation and is consistent with a single peptide bound to an RNA dimer (data not shown).

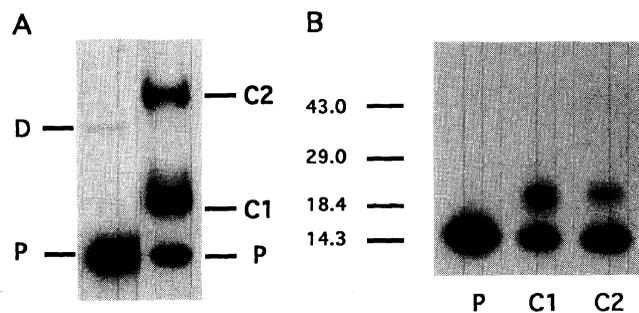


FIG. 3. Formation of complexes between the 44-nt RNA and p7-(1-55). (A) Native gels of RNA with and without p7-(1-55). P, the free RNA monomer; D, the spontaneously forming RNA dimer; C1 and C2, peptide-RNA complexes. (B) SDS/PAGE of the UV-crosslinked complexes made from bromouridine-containing 44-mer and p7-(1-55). The complexes were first resolved on native PAGE, the resulting gel was then irradiated by UV at 4°C, and the crosslinked complexes were excised. These gel slices were analyzed on 10% Tris-tricine gel, as shown.

trifugation and is consistent with a single peptide bound to an RNA dimer (data not shown).

To investigate the specific binding site on the 44-nt RNA, a series of mutants were synthesized. First, each of the four nucleotides in both loops was changed to its complementary base. These loop mutations did not alter the structure as predicted by the free-energy minimization algorithm. Gel shifts performed with p7-(1-55) and the mutant RNAs are shown in Fig. 4A. Mutation of loop 1 (Mut loop 1) elicited the formation of two complexes, neither of which correspond exactly to those obtained for wild-type RNA. Mutation of loop 2 gave results similar to wild-type RNA. Mutations of both loops 1 and 2 abolished the formation of the C2 complex; however, a small amount of a new complex, C1', was formed. This new band was more clearly observed with the loop 1 mutation and has an apparent molecular mass of 27 kDa on native gels, a value close to the value expected for a complex containing one RNA and two peptides.

Further mutations involved the RNA stem structure, allowing for potentially different folding. Fig. 4B shows that the complex C2 is not formed with Mut stem 1. Instead, a second band (C1') with the same mobility as the one observed with Mut loop 1 is present. No striking changes were found with mutations introduced into stem 2. Folding calculations showed that a long single-strand sequence flanking a short stem and loop could be generated by the introduction of multiple changes in the stem sequence. Interestingly, re-

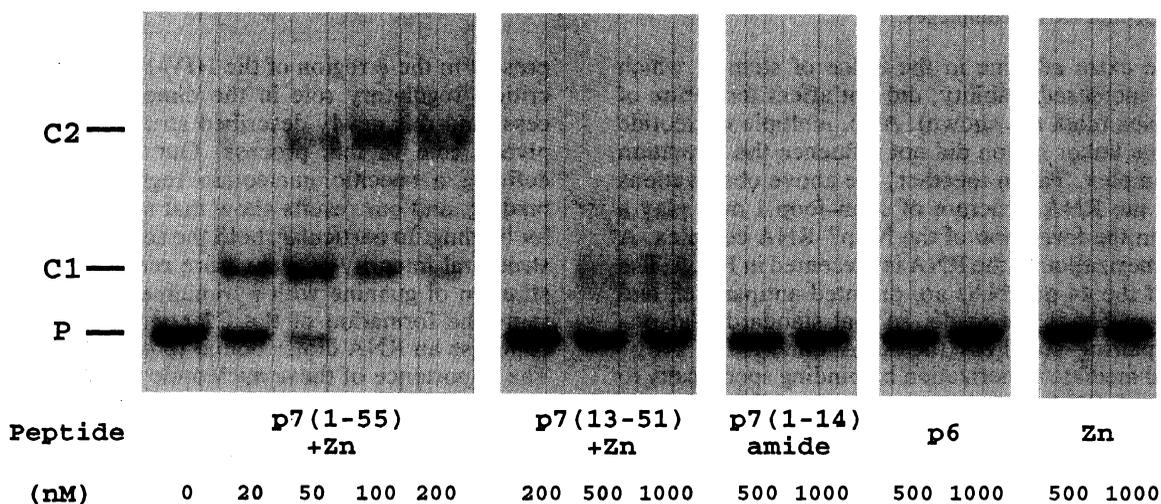


FIG. 2. Gel mobility-shift assay of the wild-type 44-nt RNA with NCp7, related peptides, and zinc. Where indicated, 2.4 molar equivalents of zinc ion were used. P, the free monomer probe; C1 and C2, peptide-RNA complexes.

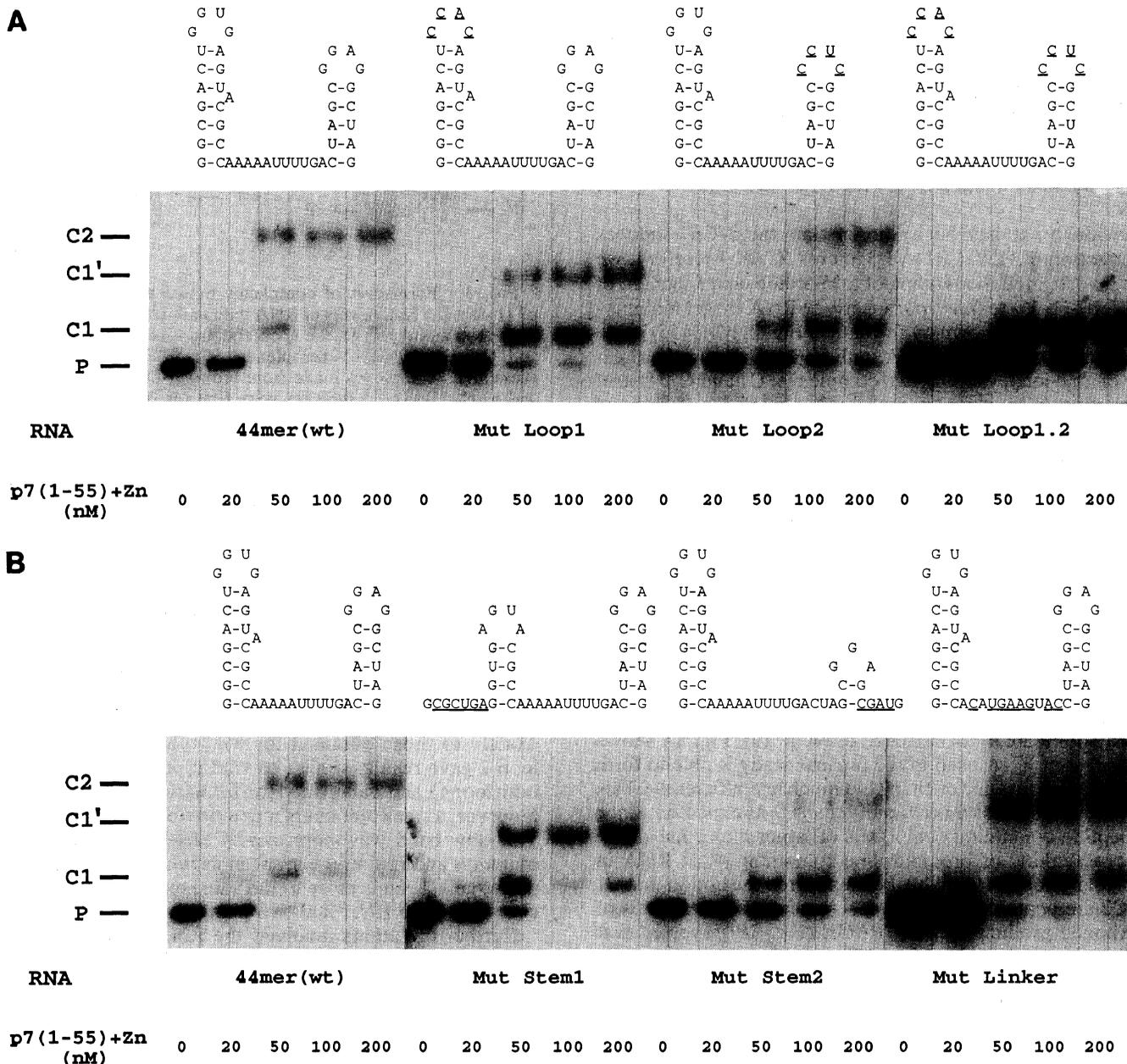


FIG. 4. Gel mobility-shift assay of p7-(1-55) in the presence of zinc with mutant 44-nt RNAs. Mutated nucleotides are indicated by underlining. Computer-predicted secondary structures of mutant RNAs are shown above the gels. All three loop mutants in *A* and the linker mutant in *B* were predicted to fold like the wild type. The stem mutants in *B* were predicted to fold differently. P, the free monomer probe; C1, C1', and C2, peptide-RNA complexes.

removal of the extra adenine in the bulge of stem 1, which should have increased stability, did not affect formation of the C2 complex (data not shown). Also, multiple nucleotide changes in the linker region did not influence the formation of the C2 complex. Taken together, the above observations suggest that the RNA structure of stem-loop 1 may play a critical role in the formation of the NCp7-RNA complex. A model for dimerization of the RNA is presented in Fig. 5. The sequences of the 44-nt RNAs are oriented antiparallel, and eight hydrogen bonds are made, so that standard Watson-Crick base pairing stabilizes the RNA dimer. The NCp7 protein could mediate dimerization by binding specifically to stem-loop 1 of RNA 1 and stem-loop 2 of RNA 2. The plausibility of NCp7 bridging the RNA loops has been demonstrated by molecular modeling (data not shown).

#### DISCUSSION

RNAse analysis supported by computational folding studies has provided evidence that a stable secondary structure is

present in the  $\psi$  region of the HIV-1 viral genome.  $\psi$  plays a critical regulatory role in the dimerization/packaging process, and this newly described structural motif may play a pivotal role in this process. Our studies were aimed at defining a specific nucleotide region required for NCp7 binding, and our results show that stem-loop 1 is important for binding. In particular, both the sequence of loop 1 and the structural integrity of stem 1 are required for binding. Substitution of guanine with pyrimidine residues in loop 1 disrupts the formation of the C2 complex, which appears to comprise an RNA dimer bound to one molecule of peptide. The importance of these nucleotide residues in the  $\psi$  region is further supported by their conservation among the various HIV-1 isolates and mammalian retroviruses (18). The mutation of stem 1 favors an optimal folding that destroys the wild-type stem-loop and does not allow formation of the C2 complex. Gel shift results point to the formation of an alternate complex (C1') in which apparently two NCp7 molecules are bound to a single RNA molecule. Our data

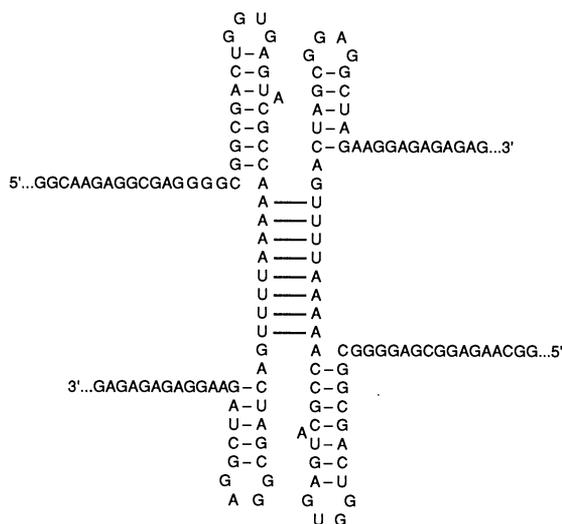


FIG. 5. Proposed dimer structure of the RNA. Antiparallel base pairing between the linker regions results in eight hydrogen bonds. A single zinc-binding motif could bind the sequence of stem-loop 1, leaving the other motif free to interact with the stem-loop 2 of RNA 2.

show that the sequence of stem-loop 1 plays a role in the binding of NCp7 and the ensuing dimerization of the RNA molecules. Thus, we believe that both the sequence and the structure in this region are important for RNA binding. This would be similar to the situation found for bacteriophage MS2 RNA, which contains the binding sites for the coat protein and the replicase enzyme. Both of these sites reside in stem-loop regions that control their accessibility (19).

In contrast to the findings regarding stem-loop 1, stem-loop 2 is not required for dimerization *in vitro*. As shown for MoMuLV RNA (20), spontaneous RNA dimerization is a slow and inefficient process, and the NC protein is needed either to trigger a change in the conformation of the RNA that favors the formation of the dimer or to drive the monomer-dimer equilibrium towards dimer formation by preferential binding to the dimer structure. The dimeric structure may be formed by more than one intermolecular contact, including the establishment of a complex conformation through unwinding of adjacent RNA structures. This process of dimerization of the retroviral RNA provided by the NC protein and/or gag precursor molecules is important to repress the 35S RNA translation and facilitate formation of a chromatin-like structure essential for infectivity (21).

Recent data show that synthetic NCp7 of HIV-1 LAV strain, in which both zinc-binding motifs have been deleted, can provide *in vitro* viral RNA dimerization and replication primer tRNA annealing (6). This finding is in contrast to our data, which indicate that a peptide of 55 amino acids containing both zinc-binding motifs is the minimal amino acid sequence required for the dimerization. Furthermore, neither an N-terminal peptide (positions 1–14) nor a C-terminal 39 amino acid-containing peptide (positions 13–51) (nor mixtures of both) appears to be functional. The NCp10 of MoMuLV and that of RaMuLV also have been synthesized (22, 23). Both of these proteins require zinc for their biological activities; in the absence of zinc and in the presence of EDTA, they are completely inactive. These data suggest that both the zinc-binding motifs and the flanking basic residues are important for the recognition of the  $\psi$  RNA region. This interaction may represent the first step toward packaging of the viral genome.

Several models for NCp7-dependent dimerization of HIV RNA are compatible with the data presented here. *In vitro*

dimerization probably proceeds by one zinc motif of NC protein binding to stem-loop 1 followed by binding of the other zinc motif to stem-loop 1 or stem-loop 2. Thus, the NC protein forms a bridge between two RNAs. If NCp7 links stem-loop 1 with stem-loop 2 of a second RNA, then a stable base pairing of the palindromic linker region and a twofold symmetric structure in a direction perpendicular to the base pairs could result. A similar model can be developed for the paired stems and loops of MoMuLV in which six G-C base pairs are made in the linker region (18).

The involvement of a stem-loop RNA binding site for NC protein within the  $\psi$  region allows us to explain some interesting experimental observations. First, there is a report that dimerization of the HIV-1 RNA inhibits the translation of gag proteins (9). We have shown that two stems and loops are formed by the sequence immediately preceding the AUG initiation codon. It seems plausible that such secondary structures, and perhaps an even more condensed dimer linkage structure, could interfere with ribosome attachment in this region. Second, it has been observed that multispliced RNAs of MoMuLV and cellular RNAs are not specifically packaged into mature virions (24). All spliced HIV RNAs have been spliced within the loop sequence of stem-loop 1. In an extensive computational study of multispliced RNAs (not shown), no stem-loops reminiscent of stem-loop 1 are formed upon splicing. Thus, it may be that the stem-loop 1 structure constitutes a specific targeting signal that mediates RNA selectivity *in vivo*.

This work was supported by the Intramural AIDS Targeted Antiviral Program of the Office of the Director of the National Institutes of Health (to E.A., G.M.C., and A.M.G.) and by Federal funds from the U.S. Department of Health and Human Services under Contract N01-CO-74102 with PRI/DynCorp (to E.T.B. and J.W.E.).

- Varmus, H. & Swanstrom, R. (1985) in *RNA Tumor Viruses*, eds. Weiss, R., Teich, N., Varmus, H. & Coffin, J. M. (Cold Spring Harbor Lab., Plainview, NY), Vol. 2, pp. 75–134.
- Aldovini, A. & Young, R. A. (1990) *J. Virol.* **64**, 450–459.
- Gorelick, R. J., Nigida, S. M., Bess, J. W., Arthur, L. O., Henderson, L. E. & Rein, A. (1990) *J. Virol.* **64**, 3207–3211.
- Henderson, L. E., Copeland, T. D., Sowder, R. C., Symthers, G. W. & Oroszlan, S. (1981) *J. Biol. Chem.* **256**, 8400–8406.
- Summers, M. F., South, T. L., Kim, B. & Hare, D. R. (1990) *Biochemistry* **29**, 329–340.
- De Rocquigny, H., Gabus, C., Vincent, A., Fournié-Zulski, M. C., Roques, B. & Darlix, J. L. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 6472–6476.
- Prats, A. C., Sarih, L., Gabus, C., Litvak, S., Keith, G. & Darlix, J. L. (1988) *EMBO J.* **7**, 1777–1783.
- Prats, A. C., Roy, C., Wang, P., Edward, M., Housset, V., Gabus, C., Paoletti, C. & Darlix, J. L. (1990) *J. Virol.* **64**, 774–783.
- Darlix, J. L., Cagus, C., Nugeyre, M. T., Clavel, F. & Barré-Sinoussi, F. (1990) *J. Mol. Biol.* **216**, 689–699.
- Myers, G., Berzofsky, J. A., Korber, B. & Smith, R. F., eds. (1992) *Human Retroviruses and AIDS, 1992* (Theor. Biol. and Biophys. Group, Los Alamos Natl. Lab., NM).
- Zuker, M. (1989) *Science* **244**, 48–52.
- Jaeger, J. A., Turner, D. H. & Zuker, M. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 7706–7710.
- Shapiro, B. A. & Zhang, K. (1990) *Comput. Appl. Biosci.* **6**, 309–318.
- Harrison, G. P. & Lever, A. M. L. (1992) *J. Virol.* **66**, 4144–4153.
- Sakaguchi, K., Appella, E., Omichinski, J. G., Clore, G. M. & Gronenborn, A. M. (1991) *J. Biol. Chem.* **266**, 7306–7311.
- Milligan, J. F. & Uhlenbeck, O. C. (1989) *Methods Enzymol.* **180**, 51–62.
- Schägger, H. & von Jagow, G. (1987) *Anal. Biochem.* **166**, 368–379.
- Konings, D. A. M., Nash, M. A., Maizel, J. V. & Arlinghaus, R. B. (1992) *J. Virol.* **66**, 632–640.
- Skripkin, E. A., Adhin, M. R., de Smit, M. H. & van Duin, J. (1990) *J. Mol. Biol.* **211**, 447–463.
- Roy, C., Tounekti, N., Mougél, M., Darlix, J. L., Paoletti, C., Ehresmann, C., Ehresmann, B. & Paoletti, J. (1990) *Nucleic Acids Res.* **18**, 7287–7292.
- Boone, L. R. & Skalka, A. M. (1981) *J. Virol.* **37**, 109–116.
- Cornille, F., Mely, Y., Fichoux, D., Salvignil, I., Genard, D., Darlix, J. L., Fournié-Zulski, M. C. & Roques, B. (1990) *Int. J. Pept. Protein Res.* **36**, 551–558.
- Roberts, W. J., Elliot, J. I., McMurry, W. J. & Williams, K. R. (1988) *Peptide Res.* **1**, 74–80.
- Mann, R. & Baltimore, D. (1985) *J. Virol.* **54**, 401–407.