RNA Secondary Structure and Binding Sites for *gag* Gene Products in the 5' Packaging Signal of Human Immunodeficiency Virus Type 1

JARED CLEVER, CHRISTOPHER SASSETTI, AND TRISTRAM G. PARSLOW*

Department of Pathology and Department of Microbiology and Immunology, University of California, San Francisco, California 94143-0506

Received 6 September 1994/Accepted 21 December 1994

The selective encapsidation of retroviral RNA requires sequences in the Gag protein, as well as a *cis*-acting RNA packaging signal (Ψ site) near the 5' end of the genomic transcript. Gag protein of human immunodeficiency virus type 1 (HIV-1) has recently been found to bind specifically to the HIV-1 Ψ element in vitro. Here we report studies aimed at mapping features within the genetically defined Ψ locus that are required for binding of HIV-1 Gag or of its processed nucleocapsid derivative. The full-length HIV-1 Gag (p55) and nucleocapsid (p15) sequences were expressed as glutathione *S*-transferase (GST) fusion proteins in *Escherichia coli*. In a gel shift assay containing excess competitor tRNA, affinity-purified GST-p15 and GST-p55 proteins bound to a 206-nucleotide Ψ RNA element spanning the major splice donor and *gag* start codons but did not bind to antisense Ψ transcripts. Quantitative filter-binding assays revealed that both GST-p55 and GST-p15 bound to this RNA sequence with identical affinities (apparent $K_d \approx 5 \times 10^{-8}$ M), indicating that all major determinants of Ψ binding affinity reside within the nucleocapsid portion of Gag. Chemical and RNase accessibility mapping, coupled with computerized sequence analysis, suggested a model for Ψ RNA structure comprising four independent stem-loops. Filter-binding studies revealed that RNAs corresponding to three of these hypothetical stem-loops can each function as an independent Gag binding site and that each is bound with approximately fourfold-lower apparent affinity than the full-length Ψ locus. Interaction of Gag with these regions is likely to play a major role in directing HIV-1 RNA encapsidation in vivo.

During retroviral capsid assembly, two copies of unspliced viral genomic RNA are selectively incorporated into each core particle as it forms, whereas most other viral and host cell RNAs tend to be excluded (reviewed in reference 31). The basis of this selective RNA encapsidation first became apparent when it was observed that mutations near the 5' end of a retroviral genome resulted in the production of virion particles which lacked genomic RNA (45). Subsequent genetic studies established that this region served as a *cis*-acting packaging signal, or Ψ site, and that similar elements were present in a wide variety of human and animal retroviruses (e.g., references 26, 31, 35, 38, 42, and 47).

The best-characterized packaging signals have been mapped to noncoding sequences near the major splice donor of a given virus, and particularly to the region between this splice junction and the *gag* start codon. Such regions often have the potential to form one or more RNA stem-loop structures, which may play a role in packaging function (3, 21, 28, 29, 53). The Ψ region is absent from most spliced forms of viral RNA, a fact which presumably accounts for the failure of subgenomic viral transcripts to be incorporated efficiently into virions (34). Additional discrete sequences situated elsewhere in the genome may also contribute to packaging of certain retroviruses (8, 42, 43, 47). Interestingly, the location of the 5' Ψ signals in many viruses coincides closely with that of the dimer linkage structure, a *cis*-acting RNA locus that mediates stable, noncovalent binding of the two genomic RNA strands of a virion to one another after the virion is formed (references 6, 15, 19, 36,

46, and 50 and references therein). The physical structure of such an RNA dimer has not yet been determined for any retrovirus, and the possible functional relationship between RNA dimerization and packaging is as yet unresolved.

Sequences from the Ψ loci of some retroviruses have been demonstrated to be not only necessary for packaging but also sufficient to direct selective encapsidation of heterologous RNAs to which they have been linked artificially (1, 4, 8, 18, 28, 34, 49, 52). In the case of human immunodeficiency virus type 1 (HIV-1), for example, functionally autonomous packaging signals have been mapped to the regions extending 30 to 40 nucleotides immediately upstream and downstream of the *gag* start codon, and additional *cis*-acting sequences in the 5' untranslated region and in the *gag* and *env* open reading frames also appear to have a role in targeting RNA into HIV-1 virions (11, 23, 30, 33, 41, 43). The relative contribution of each region to the overall efficiency and specificity of HIV-1 RNA packaging is unknown.

Retroviral encapsidation also depends upon *trans*-acting sequences in the Gag protein (2, 16, 20, 37, 40, 48). The most critical of these are provided by a region termed the cysteine-histidine box, a conserved motif with the canonical sequence CysX₂CysX₄HisX₄Cys. One or two copies of this motif are present near the C terminus in nearly all retroviral Gags, within the region which, upon cleavage of Gag by viral protease, gives rise to the mature nucleocapsid (NC) protein. Each cysteine-histidine box is thought to coordinate one zinc ion and to fold into a conformation resembling the zinc finger regions that are known to mediate nucleic acid binding by many unrelated types of proteins (37, 40, 48).

Together, these observations have led to the strong presumption that retroviral encapsidation results at least in part from direct, specific interaction between the Gag cysteine-

* Corresponding author. Mailing address: Dept. of Pathology, Box 0506, University of California, San Francisco, CA 94143. Phone: (415) 476-1015. Fax: (415) 476-9672. Electronic mail address: parslow@cgl.ucsf.edu.

FIG. 1. Fusion proteins comprising the HIV-1 Gag (p55) or NC (p15) protein linked to bacterial GST. (A) Schematic view of the GST-p55 and GST-p15 proteins. Black rectangles denote the locations of two cysteine-histidine boxes. (B) Coomassie blue-stained polyacrylamide gel analysis of the affinity-purified preparations of GST-p55 and GST-p15 used in these studies. The molecular weight standards (lane MW) are phosphorylase B, bovine serum albumin, ovalbumin, carbonic anhydrase, and soybean trypsin inhibitor.

histidine box and Ψ -site RNA. Consistent with this view, mutation or deletion of Gag cysteine-histidine boxes in the context of a virus can result in defective packaging and severely reduce viral infectivity (2, 5, 16, 17, 20). Moreover, purified Gag and NC proteins from a number of retroviruses have been shown to have significant binding affinity for RNA in vitro (25, 31, 32), though for most of the viruses studied to date this RNA binding appeared to be relatively sequence independent and so seemed inadequate to explain the specificity of packaging in vivo.

Recently, however, several groups have reported that the Gag and NC proteins of HIV-1 exhibit a specific binding affinity for the HIV-1 Ψ site in vitro (9, 14, 15, 33, 44). In particular, recombinant HIV-1 Gag or NC proteins have been shown to bind specifically to RNA fragments representing various portions of the region extending from the major HIV-1 splice donor through the first 300 bp of the *gag* coding sequence, a region corresponding closely to the genetically defined HIV-1 Ψ site. This finding strongly suggests that the binding observed in vitro mimics an in vivo interaction that is critical for HIV-1 encapsidation. To better understand the requirements for this interaction, we have used chemical and RNase accessibility mapping and a quantitative Gag binding assay to characterize sequences in and around the genetically defined HIV-1 Ψ locus immediately upstream of *gag*. Our results shed light on the elements of secondary structure in this region of the HIV-1 genomic RNA and the role that each element plays in mediating the Gag- Ψ interaction.

MATERIALS AND METHODS

Isolation of bacterially expressed Gag fusion proteins. The Gag and NC protein open reading frames were obtained through PCR from DNA of HIV-1 strain LAI, using the primers 1 (5'-GCGGGATCCATGGGTGCGAGAGCGTCGGTA-3'), 2 (5'-CGCTTGGATCCTTATTGTGACGAGGGGTCGTTG-3'), and 3 (5'-GCGAAGGATCCGCTGAAGCAATGAGCCAAGTA-3'). Primers 1 and 2 were used to obtain full-length Gag; primers 2 and 3 were used for NC. PCR products were gel purified and digested with *Bam*HI before being subcloned into the bacterial glutathione *S*-transferase (GST) expression vector pGEX-2T (Pharmacia) cleaved with the same enzyme. Dideoxynucleotide sequencing confirmed the integrity of the products. *Escherichia coli* TOPP strain 2 (Stratagene) was used for production of the GST-Gag fusions because it gave the highest yields of intact protein among several tested strains. Overnight cultures

FIG. 2. Gel retention assay of GST-p15 binding to HIV-1 Ψ RNA. Increasing concentrations of fusion protein (0, 5, 10, 15, and 20 μ g/ml) were incubated with tracer amounts of radiolabelled sense (left) or antisense (right) form of the 206-base HIV-1 Ψ (fragment B) in the presence of 500 μ g of *E. coli* tRNA per ml as the competitor. Two protein-dependent retarded complexes containing the sense RNA are indicated by filled arrows; the open arrow indicates the presumed RNA dimer. The apparent protein-dependent loss of free antisense RNA seen in this experiment occurred reproducibly, but no discrete retarded complexes were detected; this may reflect either the formation of highly heterodisperse complexes or low-level RNase contamination of the protein.

containing the desired plasmids were diluted 1:10 into fresh media and grown at 37°C for 2 h, before being induced to express the fusions with 1 mM isopropyl- β -D-thiogalactopyranoside (IPTG) for 6 h. Bacterial pellets were collected by centrifugation, washed once with ice-cold Tris-buffered saline, and resuspended in 1/15 the original volume of NETN buffer (20 mM Tris-HCl, 0.1 M NaCl, 1 mM EDTA, 0.5% Nonidet P-40 [pH 8.0]) containing 1 mM phenylmethylsulfonyl fluoride. Cells were lysed by sonication for 30 s three times at 4°C. Insoluble material was collected by centrifugation, and GST fusions were affinity purified from the supernatants by addition of 1/80 volume of glutathione-Sepharose 4B (Pharmacia). After 30 min of shaking at 4°C, the resin was collected by brief centrifugation and washed four times with 120 times the bead volume in NETN buffer. Proteins were eluted from the matrix by addition of several bead volumes of 50 mM Tris-HCl (pH 8.0)–5 mM reduced glutathione for 30 min at 4°C. Samples were concentrated with Centricon-30 microconcentrators (Amicon) to a final concentration of 5 to 10 mg of protein per ml before being stored at –70°C in small aliquots. Protein concentrations were estimated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis and by Bradford assay.

Preparation of Ψ RNA fragments. DNA fragments encoding HIV-1 Ψ were subcloned into pBluescript II KS⁺ (Stratagene), from which RNA was then transcribed in vitro. All DNA fragments were inserted into the unique *Cl*aI site except for that encoding fragment A, which was cloned into the polylinker *Acc*I site. In all cases, sense RNA was produced from the T7 promoter whereas antisense RNA was transcribed from the T3 promoter. Plasmids were linearized with either *Xho*I or *Eco*RI prior to in vitro transcription with either T7 or T3 RNA polymerase, respectively. Radiolabelled RNAs were prepared in 20- μ l reaction mixtures containing 40 mM Tris-HCl (pH 8.0), 6 mM MgCl₂, 10 mM dithiothreitol, 2 mM spermidine, 0.5 mM ATP, 0.5 mM CTP, 0.5 mM GTP, 25 μ M UTP, 50 μ Ci of [α -³²P]UTP (3,000 Ci/mmol; Amersham), 20 U of RNase inhibitor (Boehringer Mannheim), 1 μ g of linearized plasmid DNA, and 20 U of T7 or T3 RNA polymerase (Boehringer Mannheim) for 1 h at 37°C. Template DNA was removed by treatment of the reaction mixtures with 2 U of RQ1-DNase (Promega) for 15 min at 37°C, after which the reaction mixtures were phenol-chloroform extracted and passed through NucTrap purification columns (Stratagene) to remove unincorporated nucleotides. RNAs were stored at –20°C after being checked for integrity by electrophoresis through polyacrylamide gels with subsequent autoradiography.

Gel retention assay for Gag- Ψ interactions. Radiolabelled sense or antisense Ψ RNA was renatured by incubation in 5 mM *N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid (HEPES; pH 7.4)–0.1 M KCl at 80°C for 5 min followed by cooling to 25°C over 30 min. Binding reaction mixtures (10 μ l) contained 5 to 10 ng of renatured RNA, 5 U of RNase inhibitor, 5 μ g of *E. coli* tRNA, and 0 to 0.2 μ g of GST-p55 or GST-p15 fusion protein in buffer 1 (30 mM HEPES [pH 7.5], 50 mM KCl, 10 μ M ZnCl₂, 2 mM dithiothreitol). After incubation for 15 min at 25°C, 0.04% bromophenol blue, 0.04% xylene cyanol, and 5% glycerol were added, and the reaction mixtures were electrophoresed through nondenaturing 5% polyacrylamide gels in Tris-borate-EDTA at 15 mA for 1.5 h at 25°C. Gels were dried and autoradiographed.

Filter-binding assay for Gag- Ψ interactions. Filter-binding assays were performed in duplicate and consisted of a 30- μ l reaction mixture containing the indicated amounts of GST fusion protein, 10 μ g of tRNA, 5 U of RNase inhibitor, and 1 to 5 ng of radiolabelled RNA in buffer 1. After 15 min at 25°C, reaction mixtures were placed on prewetted nitrocellulose disks (0.45- μ m-pore-size HAWP type; Millipore) and filtered through with 3 ml of buffer 1. Radioactivity retained on dried filters was quantitated by liquid scintillation counting.

FIG. 3. Filter-binding assay for HIV-1 Ψ binding. Data shown are for binding to the sense and antisense forms of the 206-base Ψ construct (fragment B) used in the assay shown in Fig. 2. Each datum point and error bar depicts the mean $\pm$ standard error for duplicate determinations in a single representative experiment. Panels demonstrate the binding of GST-p15 in the absence of tRNA competitor (A) and of GST-p15 (B) or GST-p55 (C) in the presence of a 7,300-fold molar excess of *E. coli* tRNA (13 μ M). Unfused GST protein gave no binding above background under these conditions in the presence of excess tRNA (data not shown).

All data were corrected for background binding of probe to filters in the absence of protein (typically less than 3% of total radioactivity) and are expressed as the percentage of counts bound relative to the maximal binding achieved with a given RNA target within each experiment.

RNA secondary-structure analysis. RNA secondary structures within the HIV-1 Ψ locus were predicted by using the program MFOLD (University of Wisconsin Genetics Computer Group), which utilizes the energy minimization method of Zuker and colleagues (24, 54) and energy rules developed by Turner et al. (51). One optimal and 24 suboptimal secondary structures were displayed by using the SQUIGGLS option of PLOTFOOLD.

RNAse and chemical accessibility mapping of RNA secondary structure. Unlabelled Ψ RNAs were prepared as described above for radiolabelled RNAs except that the reaction volume was increased to 100 μ l and contained 0.5 mM each ribonucleoside triphosphate, 40 U of T7 RNA polymerase, and 2 μ g of linearized plasmid DNA template. Chemical and RNase modifications were

performed essentially as described previously (12, 27). Approximately 0.2 μ g of Ψ RNA was renatured as described above before being mixed with 4 μ g of *E. coli* tRNA in 20 or 200 μ l of buffer 1 for RNase digestion or chemical modifications, respectively. These reaction mixtures were treated with 0.02 to 0.1 U of RNase T₁ (Pharmacia) at 0°C for 20 min, 0.15 to 0.7 U of RNase CV or V1 (Pharmacia) at 0°C for 20 min, 0.5 to 1.0% dimethylsulfate (Aldrich) at 0°C for 20 min, 1.5 to 3.0% diethylpyrocarbonate (Sigma) at 0°C for 30 min, 20 mg of carbodiimide (Sigma) per ml at 25°C for 30 min, or 4 to 8 mg of kethoxal (U.S. Biochemical) per ml at 0°C for 1 h. Reactions were terminated and RNAs were precipitated essentially as described previously (12). Primer extension using ³²P-end-labelled oligonucleotides to detect RNase cleavage sites and chemically modified nucleotides was also performed as described previously (12), using 10 U of Moloney murine leukemia virus reverse transcriptase (New England Biolabs). Products were resolved on 6% polyacrylamide-8 M urea gels run in Tris-borate-EDTA buffer.

RESULTS

For this study, the full-length HIV-1 Gag protein (p55) and its processed NC derivative (p15) were expressed as C-terminal fusions with bacterial GST (Fig. 1A). The fusion proteins, designated GST-p55 and GST-p15, respectively, were obtained by ligating the appropriate viral sequences onto the 3' end of the GST coding region in a bacterial expression plasmid, then preparing lysates from *E. coli* that had been transformed with these plasmids, and affinity purifying the proteins from these extracts by a single pass over glutathione-Sepharose. The resulting preparations appeared on gel electrophoresis to be highly pure and intact (Fig. 1B).

Although *cis*-acting sequences at several locations in the HIV-1 genome may contribute to packaging (9, 30, 32, 33, 43), we elected to focus primarily on the 206-base region spanning nucleotides 632 to 837 of HIV-1 strain LAI. This region, which will be referred to in this report as the Ψ site, includes both the major splice donor and *gag* initiation codon, has been shown to be necessary for efficient HIV-1 packaging (2, 13, 16, 20, 30), and encompasses at least one short sequence element that is sufficient to direct packaging of heterologous RNAs into HIV-1 virions (23). This 206-base region also corresponds to the location of a potential region of complex RNA secondary structure first noted by Harrison and Lever (22). We isolated this sequence by PCR and subcloned it into a plasmid that permitted the plus (sense) strand to be transcribed *in vitro* by using bacteriophage T7 polymerase. The same sequence transcribed in the inverted (antisense) orientation served as a control. Biosynthetically radiolabelled RNAs transcribed from the recombinant plasmids were then used as substrates for protein binding as described below.

Initial studies using a gel retention assay confirmed the earlier observation (9, 14, 15, 33, 44) that the p55 Gag and p15 NC proteins each can bind specifically to the HIV-1 Ψ locus *in vitro*. Tracer amounts of labelled 206-base Ψ RNA were incubated with increasing quantities of each fusion protein in the presence of 20 μ M *E. coli* tRNA, which corresponded to roughly a 10,000-fold molar ratio of tRNA to Ψ RNA. The representative gel depicted in Fig. 2 illustrates that within the concentration range tested, GST-p15 bound readily to sense Ψ RNA but not to its antisense counterpart. Half-maximal binding occurred at a protein concentration of roughly 240 nM. The interaction with Ψ produced at least two retarded complexes with different mobilities (solid arrows) which may correspond to two different stoichiometries of protein-RNA binding (see below). Similar results were obtained with GST-p55, but no complexes were formed when Ψ RNA was incubated with control bacterial extracts containing unfused GST protein (data not shown). A lower-mobility Ψ RNA species (open arrow) was reproducibly observed in the absence of protein; this species was heat labile (see below) and is presumed to represent an RNA dimer (6, 15, 19, 36, 46, 50). We could not

FIG. 4. A model of RNA secondary structure in the HIV-1 Ψ locus. (A) Computerized conformational analyses of the Ψ sequence from HIV-1 strain LAI, using the MFOLD algorithm, predicted four independent stem-loop structures (SL1 to SL4) within the region examined (residues 632 to 838). Locations of the major viral splice donor site (SD) and the gag initiation codon (open lettering) are indicated. (B) Interstrain sequence variability of the HIV-1 Ψ locus. The Ψ sequence from the LAI strain is compared with those from several other HIV-1 isolates; locations of the four putative stem-loops are indicated at the top. Data are from reference 39.

discern from this analysis whether monomeric or dimeric Ψ RNA was the preferred binding substrate.

Subsequent studies were conducted using a filter-binding assay to better quantify the affinity of Gag-Ψ binding. When reactions were carried out in the absence of competitor tRNA (Fig. 3A), GST-p15 bound sense and antisense forms of the 206-base Ψ to approximately equal degrees, confirming its relatively high ($K_d \approx 50$ to 60 nM) nonspecific affinity for RNA (9, 32). In the presence of excess tRNA (Fig. 3B), binding of antisense Ψ was almost completely suppressed, but binding to the sense Ψ persisted; the apparent K_d of this specific interaction (as revealed by the protein concentration required for half-maximal binding) was approximately 50 to 60 nM. The GST-p55 fusion exhibited a similar degree of specificity and an essentially identical K_d under the same conditions (Fig. 3C), indicating that all major determinants of specificity and affinity for binding to this 206-base Ψ reside within the NC (p15) moiety of Gag.

To assess the possible secondary structure of the 206-base Ψ RNA, we analyzed its sequence theoretically by using the program MFOLD. This analysis considered standard Watson-Crick base pairing as well as potential GU base pairs and yielded several alternative folding patterns for the Ψ sequence that were each highly favorable thermodynamically, with predicted free energies of -45 to -38 kcal (1 kcal = 4.184 kJ)/mol. These patterns differed in detail, but the majority were dominated by four independent stem-loops which we designate

SL1 to SL4, as shown in Fig. 4A. Database searches further indicated that these four stem-loops could form at the same locations in sequences from nearly all published HIV-1 isolates (Fig. 4B). The sequences required for stem formation were particularly well conserved among isolates, and in a few instances there was a suggestion of covariance between hypothetically paired bases that preserved their complementary (e.g., preservation of UG or UA pairing at position 719 in the stem of SL1).

This hypothetical folding pattern was confirmed experimentally by the results of nucleolytic accessibility mapping. As shown by the representative data in Fig. 5A, we exposed protein-free 206- or 335-base Ψ RNA to various RNases and chemical agents whose ability to cleave or modify RNA is dictated by sequence and secondary structure, and we then mapped the resulting cleavage sites by primer extension. The patterns of cleavage that we observed (summarized in Fig. 5B) provide strong, direct evidence that SL1, SL2, and SL3 form as predicted. The evidence regarding SL4 was somewhat less compelling: we reproducibly observed preferential dimethyl-sulfate accessibility in the hypothesized loop of SL4 but also detected partial reactivity with some of the single-strand-selective agents within the predicted stem. We interpret these findings as evidence that SL4 exists but is less thermodynamically stable than the other three stem-loop owing to its relatively high content of GU base pairs.

Guided by this model for the Ψ, we then asked which sub-

FIG. 5. Accessibility mapping of secondary structures in HIV-1 Ψ RNA in vitro. (A) Portions of two representative autoradiograms showing nuclease and chemical accessibility data for portions of the Ψ locus. Each shows the products obtained from primer extension analyses in which the template consisted of either untreated Ψ RNA (RNA only) or the same RNA following exposure to the chemical or enzyme listed above each lane (DMS, dimethylsulfate; DEP, diethylpyrocarbonate; CMCT, 1-cyclohexyl-3-[2-morpholinoethyl]carbodiimide-*p*-toluene sulfonate). General modification or cleavage specificities of each reagent are shown in parentheses; with the exception of RNase CV, each acts preferentially on unpaired nucleotides. Additional details can be found in reference 12. RNA sequence, deduced by sequencing plasmid DNA with the same primer, appears at the right of each panel, with the approximate locations of the hypothetical stem-loops indicated. (B) Summary of accessibility data. Symbols indicate the sites of cleavage or modification by each of five selected reagents within the Ψ locus, with the four hypothetical stem-loops indicated. The *gag* initiation codon is shown in open lettering. Numbering is from reference 39.

regions of the locus are necessary to bind NC protein. Selected portions of Ψ were subcloned by PCR, then transcribed in vitro, and used as substrates for the filter-binding assay with GST-p15. We also extended our analysis upstream to include the TAR locus at the 5' end of the HIV-1 transcript and downstream to nucleotides 829 to 958, which lie within the *gag* coding region, since these nucleotides have recently been proposed to form a separate high-affinity site for Gag and NC binding (9). The various Ψ subregions that we tested (designated A to P) are depicted schematically in Fig. 6A, and binding data for a few representative subregions are presented in Fig. 6B. In general, we found that the subregions could be grouped in three categories with respect to NC binding affinity. The strongest binding ($K_d \approx 55$ nM) was that achieved by the 206-base Ψ sequence (fragment B) as well as by a somewhat larger fragment that extended an additional 130 bases downstream into the *gag* coding region (fragment A). Much shorter fragments (D to F) corresponding to SL1, SL3, or SL4 each bound with an affinity roughly fourfold lower ($K_d \approx 200$ nM) than that of the full-length Ψ. Interestingly, a fragment containing only those Ψ sequences located upstream of SL1 (frag-

ment M) also showed this intermediate level of affinity, though we had detected no significant stem-loop elements in this region either by computerized structure prediction or by accessibility mapping (data not shown). We tested these intermediate-affinity substrates (fragments D, E, F, and M) in various combinations with one another and with SL2, as well as with the remaining 3' sequences of the 206-base Ψ, but none of the resulting constructs (fragments G to L) achieved the full affinity exhibited by the 206-base Ψ itself.

Of the plus-strand substrates that we tested, those which bound most weakly were SL2 alone (fragment N), the TAR-containing region upstream of Ψ (fragment P), and a fragment encompassing residues 829 to 958 (fragment O), each of which bound with a K_d of ≈ 400 nM. The comparatively weak binding of the fragment comprising TAR indicates that NC does not bind indiscriminately to RNA stem-loop structures. Nevertheless, even these relatively poor plus-strand subregions bound with at least 12-fold-higher affinity than the 206-base antisense control ($K_d > 5,000$ nM).

These data indicated that SL1, SL3, and SL4 can each serve individually as an NC-binding site but that maximal binding

FIG. 6. Deletion mapping of specific GST-p15 binding sites within the HIV-1 Ψ locus. (A) Various sequences in and adjacent to Ψ (fragments A to P) were assayed for the ability to be bound by GST-p15 in the filter-binding assay and were found to have the approximate K_d values shown at the right. The subfragments are shown schematically to indicate their putative conformations. Subfragments corresponding to the four individual stem-loops were residues 697 to 731 (fragment D), 731 to 759 (fragment N), 751 to 796 (fragment E), and 783 to 816 (fragment F); other constructs either consisted of fusions of these smaller subfragments or had the boundaries shown. (B) Representative filter-binding data for the interactions of selected Ψ subfragments with GST-p15 in the presence of excess tRNA competitor. B- denotes the antisense form of fragment B (the 206-nucleotide Ψ RNA). Data shown for subfragments B and B- are from the experiment depicted in Fig. 3B.

affinity requires the full 206-base Ψ . One hypothetical interpretation of this finding was that the fourfold-higher affinity of full-length Ψ might result from its ability to form RNA dimers, which in turn might constitute a preferred substrate for NC binding. This hypothesis was consistent with the finding that certain Ψ -related RNA oligonucleotides dimerize spontaneously in solution (Fig. 2 and references 6, 15, 36, 46, and 50), and with reports variously indicating that a purine-rich tract at residues 817 to 840 (6, 50) and an AT-rich sequence at residues 755 to 763 (44) must be present for such dimerization to occur. Our own results also implied that dimerization was sequence dependent, since we had detected presumptive dimers of the sense Ψ fragment but not of the antisense form (Fig. 2). We therefore chose four Ψ subregions from our series (Fig. 7a) that either contained or lacked these putative dimerization sequences and compared their electrophoretic mobilities (Fig. 7b) under native conditions and after heat denaturation. We found that all four subregions efficiently formed heat-labile presumptive dimers, regardless of the presence or absence of the purine-rich or AT-rich tracts. Our data therefore failed to support the proposal (6, 44, 50) that these tracts are required for Ψ dimerization. We further conclude that although the fragments differ in NC binding affinity (Fig. 6), these differences cannot be attributed to any obvious disparities in the tendency to form RNA dimers in vitro.

FIG. 7. In vitro dimerization of RNA oligonucleotides from the HIV-1 Ψ locus. (a) Schematic depiction of the four Ψ subfragments tested, with the locations of an AU-rich tract (open rectangle) and a purine-rich tract (filled rectangle) indicated. (b) Autoradiogram of a polyacrylamide electrophoretic analysis of the four RNA species, each tested in native conformation in buffer 1 (Nat) and immediately after heat denaturation at 80°C for 5 min in 10 mM Tris (pH 7.5)–1 mM EDTA (Den).

DISCUSSION

Recent reports from several laboratories (9, 14, 15, 33, 44) have established that *gag* gene products from HIV-1 bind specifically to HIV-1 Ψ RNA in vitro, an interaction that almost certainly mimics a critical step in HIV-1 RNA encapsidation in vivo. Our results of gel retention and filter-binding assays are in full accordance with those reports in demonstrating that HIV-1 Gag fusion proteins bind Ψ with an affinity of up to approximately 55 nM. These data also confirm that all major determinants of Ψ binding affinity and specificity reside within the NC portion of the HIV-1 Gag polypeptide, though this may not be the case for all retroviruses (5, 17, 25). In the present study, we have extended the earlier findings by probing the secondary structure of HIV-1 Ψ region RNA and searching for features within this structure that correspond to the minimal in vitro binding sites for Gag or NC.

The results of our chemical and RNase accessibility mapping of the 5' Ψ region suggest that it is composed of four independent stem-loops, SL1 to SL4. These together occupy a 110-nucleotide region spanning residues 697 to 806 of the RNA. This proposed structure differs significantly from earlier models for HIV-1 Ψ but incorporates several features that have been observed or hypothesized by other workers. For example, formation of SL1 and of SL2 and SL3 was demonstrated previously by Skripkin et al. (46) and by Sakaguchi et al. (44), respectively, using RNase mapping, whereas Luban and Goff have predicted the existence of SL4 on theoretical grounds (33). The more complex and extensive fold proposed by Harrison and Lever (22) includes the elements that we designate as SL1 and SL3 but postulates entirely different conformations for the regions corresponding to SL2 and SL4, whereas the structure deduced by Baudin and coworkers (10) includes SL3 and SL4 but not SL1 or SL2.

Of these models, the four-stem conformation that we propose here is the most compatible with the accessibility data that we obtained. Comparisons among viral strains further indicate that this folding pattern would be thermodynamically favorable for all known sequence variants of the HIV-1 Ψ region (Fig. 4B), though it should be noted that this 5' region shows relatively little interstrain variability overall. The region is also quite rich in purines (67% purines within residues 632 to 838), and three of the four structural elements in our model are tetraloops with purine-rich loops. This is reminiscent of the two similarly situated tetraloops that have been identified as conserved features in the packaging signals of murine leukemia virus and related type C retroviruses (29). In our model, the 5' viral splice donor site is located within such a loop, which might conceivably increase its accessibility for cleavage.

Quantitative filter-binding studies indicate that three of the four putative stem-loop-forming regions in Ψ can serve as independent binding sites for Gag or NC protein. In particular, RNAs coinciding with SL1, SL3, or SL4 were each independently bound by our NC fusion protein with a K_d of approximately 200 nM. We are currently conducting studies aimed at determining whether stem-loop formation is a requirement for NC binding to these Ψ RNA fragments and also at identifying the nature of the additional NC-binding site that we detected within the 50-base region immediately upstream of SL1 (fragment M in Fig. 6A). Significantly, however, two of the NC-binding stem-loops in our model have previously been implicated functionally in packaging activity: linkage to a short HIV-1 RNA containing the SL3 sequence is sufficient to target heterologous RNAs into HIV-1 capsid-like particles (23), whereas the isolated SL4 sequence can interfere with such targeting in *trans* (33). Because SL1 is located upstream of the

major splice donor and hence is present in many subgenomic HIV-1 RNAs, the possibility that it, too, might function as a packaging signal is unexpected and will require further investigation. We note, however, that subgenomic HIV-1 RNAs are packaged to a significant degree, at least in artificial systems (33), and that evidence for *cis*-acting packaging signals located upstream of the major splice donor has previously been reported for Rous sarcoma virus (47).

The presence of multiple retarded complexes in the gel retention assay (Fig. 2) have been noted previously (9, 44) and is consistent with the view that two or more Gag or NC proteins can bind each copy of the 206-base Ψ . Although SL1, SL3, and SL4 each bound GST-p15 with an affinity only four-fold lower than that of the 206-base Ψ , we were unable to identify any combination of these minimal elements that bound as strongly as the complete locus. This may reflect the limitations of the assay that we used: when only a single bound protein is sufficient to retain labelled RNA on the filter, the K_d values obtained in a filter-binding assay by the approach we that used approximate those of the individual binding sites unless there is significant cooperativity among sites (7). Our data did not reveal convincing evidence of cooperative binding, but neither are they sufficient to exclude it. An alternative interpretation, however, is that the four fold-higher binding affinity of the 206-base Ψ reflects its ability to undergo higher-order folding as a single looped domain, which might then enhance accessibility and/or binding affinity of the individual stem-loops. We also observed that GST-p15 bound to each of the various plus-strand Ψ fragments that we tested with at least 12-fold-higher affinity than to any fragment derived from the minus strand; this finding may suggest that Gag and NC bind with some degree of preference to purine-rich RNA per se.

Our data also address, to a limited extent, the sequence requirements for HIV-1 RNA dimerization *in vitro*. Recent studies have provided evidence that RNA oligonucleotides derived from the HIV-1 Ψ region will dimerize spontaneously *in vitro* and that this dimerization is facilitated kinetically by Gag protein (6, 15, 36, 46, 50). On the basis of analyses of this phenomenon, several groups have identified specific short sequence tracts within the Ψ locus that appear necessary for dimerization, and these have served as the bases for structural models of the dimer interface. For example, one group (44) has hypothesized that NC cross-linking of SL2 and SL3 on a pair of HIV-1 RNA strands might promote Watson-Crick base pairing of the strands through the AU-rich palindrome located between SL2 and SL3. Weighing against that model, however, are occasional HIV-1 strains in which the palindromic nature of the AU-rich tract is not fully conserved (Fig. 4B), as well as our finding that NC binds comparatively weakly to SL2 (Fig. 6). An alternative model holds that viral RNA dimerization may occur through formation of purine quartets, a type of noncanonical base interaction that occurs exclusively in purine-rich (and, especially, in G-rich) strands (6, 36, 46, 50); in support of this model, some groups have reported that one specific purine-rich tract in the HIV-1 Ψ region is essential for its *in vitro* RNA dimerization (6, 50). The purine quartet model has considerable elegance and appeal, though the occurrence of purine quartets in native retrovirus particles (19) has been called into question. Our study, however, failed to confirm the earlier reports: all four of the Ψ subfragments that we tested dimerized efficiently *in vitro* irrespective of the presence of the AU-rich or GA-rich tracts (Fig. 7). Hence, the differences in Gag binding affinity among these fragments do not correlate in any meaningful way with differences in *in vitro* dimerization, and the precise relationship of Ψ RNA dimerization to Gag binding remains obscure.

ACKNOWLEDGMENTS

We thank N. Defranoux for insightful comments on the manuscript, A. Davies for advice on the RNA-folding software, and M. R. Madlansacay, B. Bond, and Z. Mosquera for technical help.

This work was supported by NIH grants AI29313 and GM37036. J.C. and C.S. were supported in part by NIH training grants AI07395 and AI07334, respectively.

REFERENCES

- Adam, M. A., and A. D. Miller. 1988. Identification of a signal in a murine retrovirus that is sufficient for packaging of nonretroviral RNA into virions. *J. Virol.* **62**:3802–3806.
- Aldovini, A., and R. A. Young. 1990. Mutations of RNA and protein sequences involved in human immunodeficiency virus type 1 packaging result in production of noninfectious virus. *J. Virol.* **64**:1920–1926.
- Alford, R. L., S. Honda, C. B. Lawrence, and J. W. Belmont. 1991. RNA secondary structure analysis of the packaging signal for Moloney murine leukemia virus. *Virology* **183**:611–619.
- Armentano, D., S. F. Yu, P. W. Katoff, T. von Ruden, W. F. Anderson, and E. Gilboa. 1987. Effect of internal viral sequences on the utility of retroviral vectors. *J. Virol.* **61**:1647–1650.
- Aronoff, R., A. M. Hajjar, and M. L. Linial. 1993. Avian retroviral RNA encapsidation: reexamination of functional 5' RNA sequences and the role of nucleocapsid Cys-His motifs. *J. Virol.* **67**:178–188.
- Awang, G., and D. Sen. 1993. Mode of dimerization of HIV-1 genomic RNA. *Biochemistry* **32**:11453–11457.
- Bailey, J. M. 1979. Interpretation of nitrocellulose filter assays of protein-nucleic acid binding. *Anal. Biochem.* **93**:204–206.
- Bender, M. A., T. D. Palmer, R. E. Geinas, and A. D. Miller. 1987. Evidence that the packaging signal of Moloney murine leukemia virus extends into the gag region. *J. Virol.* **61**:1639–1646.
- Berkowitz, R. D., J. Luban, and S. P. Goff. 1993. Specific binding of human immunodeficiency virus type 1 gag polyprotein and nucleocapsid protein to viral RNAs detected by RNA mobility shift assays. *J. Virol.* **67**:7190–7200.
- Baudin, F., R. Marquet, C. Isel, J.-L. Darlix, B. Ehresmann, and C. Ehresmann. 1993. Functional sites in the 5' region of human immunodeficiency virus type 1 RNA from defined structural domains. *J. Mol. Biol.* **229**:382–397.
- Buchsacher, G. L., and A. T. Panganiban. 1992. Human immunodeficiency virus vectors for inducible expression of foreign genes. *J. Virol.* **66**:2731–2739.
- Christiansen, J., J. Egebjerg, N. Larson, and R. A. Garrett. 1990. Analysis of rRNA structure: experimental and theoretical considerations, p. 229–252. *In* G. Spedding (ed.), *Ribosomes and Protein Synthesis*. IRL and Oxford University Press, Oxford.
- Clavel, F., and J. M. Orenstein. 1990. A mutant of human immunodeficiency virus with reduced RNA packaging and abnormal particle morphology. *J. Virol.* **64**:5230–5234.
- Dannull, J., A. Surovov, G. Jung, and K. Moelling. 1994. Specific binding of HIV-1 nucleocapsid protein to PSI RNA *in vitro* requires N-terminal zinc finger and flanking basic amino acid residues. *EMBO J.* **13**:1525–1533.
- Darlix, J.-L., C. Gabus, M.-T. Nugeyre, F. Clavel, and F. Barre-Sinoussi. 1990. *Cis* elements and *trans*-acting factors involved in the RNA dimerization of the human immunodeficiency virus HIV-1. *J. Mol. Biol.* **216**:689–699.
- Dorfman, T., J. Luban, S. P. Goff, W. A. Haseltine, and H. G. Göttinger. 1993. Mapping of functionally important residues of a cysteine-histidine box in the human immunodeficiency virus type 1 nucleocapsid protein. *J. Virol.* **67**:6159–6169.
- Dupraz, P., and P. Spahr. 1992. Specificity of Rous sarcoma virus nucleocapsid protein in genomic RNA packaging. *J. Virol.* **66**:4662–4670.
- Embertson, J. E., and H. M. Temin. 1987. Lack of competition results in efficient packaging of heterologous murine retroviral RNAs and reticuloendotheliosis virus encapsidation-minus RNAs by the reticuloendotheliosis virus helper cell line. *J. Virol.* **61**:2675–2683.
- Fu, W., R. J. Gorelick, and A. Rein. 1994. Characterization of human immunodeficiency virus type 1 dimeric RNA from wild-type and protease-defective virions. *J. Virol.* **68**:5013–5018.
- Gorelick, R. J., S. M. Nigida, Jr., J. W. Bess, Jr., L. O. Arthur, L. E. Henderson, and A. Rein. 1990. Noninfectious human immunodeficiency virus type 1 mutants deficient in genomic RNA. *J. Virol.* **64**:3207–3211.
- Hackett, P. B., M. W. Dalton, D. P. Johnson, and R. B. Petersen. 1991. Phylogenetic and physical analysis of the 5' leader RNA sequences of avian retroviruses. *Nucleic Acids Res.* **19**:6929–6934.
- Harrison, G. P., and A. M. L. Lever. 1992. The human immunodeficiency virus type 1 packaging signal and major splice donor region have a conserved stable secondary structure. *J. Virol.* **66**:4144–4153.
- Hayashi, T., T. Shioda, Y. Iwakura, and H. Shibuta. 1992. RNA packaging signal of human immunodeficiency virus type 1. *Virology* **188**:590–599.
- Jaeger, J. A., D. H. Turner, and M. Zuker. 1989. Predicting optimal and suboptimal secondary structure for RNA. *Methods Enzymol.* **183**:281–306.
- Katoh, I., H. Kyushiki, Y. Sakamoto, Y. Ikawa, and Y. Yoshinaka. 1991.

- Bovine leukemia virus matrix-associated protein MA(p15): further processing and formation of a specific complex with the dimer of the 5'-terminal genomic RNA fragment. *J. Virol.* **65**:6845-6855.
26. Katz, R. A., R. W. Terry, and A. M. Skalka. 1986. A conserved *cis*-acting sequence in the 5' leader of avian sarcoma virus RNA is required for packaging. *J. Virol.* **59**:163-167.
 27. Kjems, J., M. Brown, D. D. Chang, and P. A. Sharp. 1991. Structural analysis of the interaction between the human immunodeficiency virus Rev protein and the Rev response element. *Proc. Natl. Acad. Sci. USA* **88**:683-687.
 28. Knight, J. B., Z. H. Si, and C. M. Stoltzfus. 1994. A base-paired structure in the avian sarcoma virus 5' leader is required for efficient encapsidation of RNA. *J. Virol.* **68**:4493-4502.
 29. Konings, D. A. M., M. A. Nash, J. V. Maizel, and R. B. Arlinghaus. 1992. Novel GACG-hairpin pair motif in the 5' untranslated region of type C retroviruses related to murine leukemia virus. *J. Virol.* **66**:632-640.
 30. Lever, A., H. Gottling, W. Haseltine, and J. Sodroski. 1989. Identification of a sequence required for efficient packaging of human immunodeficiency virus type 1 RNA into virions. *J. Virol.* **63**:4085-4087.
 31. Linial, M., and A. D. Miller. 1990. Retroviral RNA packaging: sequence requirements and implications. *Curr. Top. Microbiol. Immunol.* **157**:125-152.
 32. Luban, J., and S. P. Goff. 1991. Binding of human immunodeficiency virus type 1 (HIV-1) RNA to recombinant HIV-1 gag polyprotein. *J. Virol.* **65**:3203-3212.
 33. Luban, J., and S. P. Goff. 1994. Mutational analysis of *cis*-acting packaging signals in human immunodeficiency virus type 1 RNA. *J. Virol.* **68**:3784-3793.
 34. Mann, R., and D. Baltimore. 1985. Varying the position of a retrovirus packaging signal results in the encapsidation of both unspliced and spliced RNAs. *J. Virol.* **54**:401-407.
 35. Mann, R., R. Mulligan, and D. Baltimore. 1983. Construction of a retrovirus packaging mutant and its use to produce helper-free defective retrovirus. *Cell* **33**:153-159.
 36. Marquet, R., F. Baudin, C. Gabus, J.-L. Darlix, M. Mougé, C. Ehresmann, and B. Ehresmann. 1991. Dimerization of human immunodeficiency virus (type 1) RNA: stimulation by cations and possible mechanism. *Nucleic Acids Res.* **19**:2349-2357.
 37. Morellet, N., N. Jullian, H. De Rocquigny, B. Maigret, J.-L. Darlix, and B. P. Roques. 1992. Determination of the structure of the nucleocapsid protein NCp7 from the human immunodeficiency virus type 1 by ¹H NMR. *EMBO J.* **11**:3059-3065.
 38. Murphy, J. E., and S. P. Goff. 1989. Construction and analysis of deletion mutations in the U5 region of Moloney murine leukemia virus: effects on RNA packaging and reverse transcription. *J. Virol.* **63**:319-327.
 39. Myers, G., S. Wain-Hobson, B. Korber, R. F. Smith, and G. N. Pavlakis. 1993. Human retroviruses and AIDS. Los Alamos National Laboratory, Los Alamos, N. Mex.
 40. Omichinski, J. G., G. M. Clore, K. Sakaguchi, E. Appella, and A. M. Gronenborn. 1991. Structural characterization of a 39-residue synthetic peptide containing the two zinc binding domains from the HIV-1 p7 nucleocapsid protein by CD and NMR spectroscopy. *FEBS Lett.* **292**:25-30.
 41. Poznansky, M., A. Lever, L. Bergeron, W. Haseltine, and J. Sodroski. 1991. Gene transfer into human lymphocytes by a defective human immunodeficiency virus type 1 vector. *J. Virol.* **65**:532-536.
 42. Pugatsch, T., and D. Stacey. 1983. Identification of a sequence likely to be required for avian retroviral packaging. *Virology* **128**:505-511.
 43. Richardson, J. H., L. A. Child, and A. M. Lever. 1993. Packaging of human immunodeficiency virus type 1 RNA requires *cis*-acting sequences outside the 5' leader region. *J. Virol.* **67**:3997-4005.
 44. Sakaguchi, K., N. Zambrano, E. T. Baldwin, B. A. Shapiro, J. W. Erickson, J. G. Omichinski, G. M. Clore, A. M. Gronenborn, and E. Appella. 1993. Identification of a binding site for the human immunodeficiency virus type 1 nucleocapsid protein. *Proc. Natl. Acad. Sci. USA* **90**:5219-5223.
 45. Shank, P. R., and M. Linial. 1980. Avian oncovirus mutant (SE21Q1b) deficient in genomic RNA: characterization of a deletion in the provirus. *J. Virol.* **36**:450-456.
 46. Skripkin, E., J.-C. Paillart, R. Marquet, B. Ehresmann, and C. Ehresmann. 1994. Identification of the primary site of the human immunodeficiency virus type 1 RNA dimerization *in vitro*. *Proc. Natl. Acad. Sci. USA* **91**:4945-4949.
 47. Sorge, J., W. Ricci, and S. Hughes. 1983. *cis*-acting RNA packaging locus in the 115-nucleotide direct repeat of Rous sarcoma virus. *J. Virol.* **48**:667-675.
 48. South, T. L., and M. F. Summers. 1993. Zinc- and sequence-dependent binding to nucleic acids by the N-terminal zinc finger of the HIV-1 nucleocapsid protein: NMR structure of the complex with the psi-site analog, dACGCC. *Protein Sci.* **2**:3-19.
 49. Sullenger, B. A., and T. R. Cech. 1993. Tethering ribozymes to a retroviral packaging signal for destruction of viral RNA. *Science* **262**:1566-1569.
 50. Sundquist, W. I., and S. Heaphy. 1993. Evidence for interstrand quadruplex formation in the dimerization of human immunodeficiency virus 1 genomic RNA. *Proc. Natl. Acad. Sci. USA* **90**:3393-3397.
 51. Turner, D. H., N. Sugimoto, and S. M. Freier. 1988. RNA structure prediction. *Annu. Rev. Biophys. Chem.* **17**:167-192.
 52. Watanabe, S., and H. Temin. 1982. Encapsidation sequences for spleen necrosis virus, an avian retrovirus, are between the 5' long terminal repeat and the start of the gag gene. *Proc. Natl. Acad. Sci. USA* **79**:5986-5990.
 53. Yang, S., and H. M. Temin. 1994. A double hairpin structure is necessary for the efficient encapsidation of spleen necrosis virus retroviral RNA. *EMBO J.* **13**:713-726.
 54. Zuker, M. 1989. On finding all suboptimal foldings of an RNA molecule. *Science* **244**:48-52.