Identification of a Minimal Region of the HIV-1 5′-Leader Required for RNA Dimerization, NC Binding, and Packaging

Xiao Heng1†, Siarhei Kharytonchyk2†, Eric L. Garcia2, Kun Lu1, Sai Sachin Divakaruni1, Courtney LaCotti1, Kedy Edme1, Alice Telesnitsky2⁎ and Michael F. Summers1⁎

1Howard Hughes Medical Institute and Department of Chemistry and Biochemistry, University of Maryland Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250, USA
2Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, MI 48109-0620, USA

Assembly of human immunodeficiency virus type 1 (HIV-1) particles is initiated in the cytoplasm by the formation of a ribonucleoprotein complex comprising the dimeric RNA genome and a small number of viral Gag polyproteins. Genomes are recognized by the nucleocapsid (NC) domains of Gag, which interact with packaging elements believed to be located primarily within the 5′-leader (5′-L) of the viral RNA. Recent studies revealed that the native 5′-L exists as an equilibrium of two conformers, one in which dimer-promoting residues and NC binding sites are sequestered and packaging is attenuated, and one in which these sites are exposed and packaging is promoted. To identify the elements within the dimeric 5′-L that are important for packaging, we generated HIV-1 5′-L RNAs containing mutations and deletions designed to eliminate substructures without perturbing the overall structure of the leader and examined effects of the mutations on RNA dimerization, NC binding, and packaging. Our findings identify a 159-residue RNA packaging signal that possesses dimerization and NC binding properties similar to those of the intact 5′-L and contains elements required for efficient RNA packaging.

© 2012 Elsevier Ltd. All rights reserved.

Introduction

Like all retroviruses, the human immunodeficiency virus (HIV) specifically packages two copies of its unspliced RNA genome, both of which are utilized for strand-transfer-mediated recombination during reverse transcription.1,2 Although only one DNA allele is generated by this mechanism (and retroviruses are thus considered “pseudodiploid”),3 the production of recombinant proviruses from heterozygous virions appears to serve as a primary means for promoting genetic evolution under conditions of environmental and chemotherapeutic stresses.4–7 Genome selection is mediated by interactions...
between the nucleocapsid (NC) domains of a subset of the viral structural proteins (called Gag) and RNA elements located primarily within the 5′-leader (5′-L) region of the genome. Genetic studies indicate that genomes are selected for packaging as dimers, and there is compelling evidence that dimerization and selection occur in the cytoplasm subsequent to nuclear export. HIV-1 genomes are trafficked to the plasma membrane by a small number of Gag proteins, where additional Gag proteins co-localize and virus assembly occurs. Subsequent to budding, the packaged RNA molecules exist as weakly formed, non-covalently linked dimers that form more stable dimers as the virus ages.

Agents that disrupt the structure of the NC domains of Gag and interfere with genome packaging have potent antiviral properties, and a better understanding of the structural determinants of genome selection could facilitate the development of new therapeutic approaches for the treatment of AIDS. However, despite considerable effort, current understanding of the protein–RNA interactions and mechanisms that promote and/or regulate retroviral genome selection remains limited (for reviews, see Refs. 8-17,38,39). For some retroviruses, including the Moloney murine leukemia virus and Rous sarcoma virus, relatively small regions of the 5′-L RNAs (fewer than 100 nucleotides) that are capable of directing the packaging of heterologous RNAs into virus-like particles (VLPs) have been identified, and this has facilitated high-resolution structural studies of the NC:RNA complexes responsible for RNA packaging. Although atomic-level structures have been determined for the HIV-1 NC protein bound to isolated hairpin and single-stranded RNAs, these studies did not explain how the virus discriminately packages its dimeric, unspliced genome. To date, no high-resolution structural information is available for a packaging-competent HIV-1 RNA.

Efforts to identify a minimal HIV-1 packaging signal have been complicated by a number of factors. First, the 5′-L contains elements that are important for functions that are unrelated to packaging but critical for replication (including transcriptional activation, primer binding during reverse transcription, splicing, and dimerization) (Fig. 1), and mutagenesis studies designed to test effects on packaging can adversely affect other functions. For example, mutagenesis studies originally suggested that the TAR hairpin is required for genome packaging, but more recent studies indicate that effects originally attributed to a packaging defect were actually due to dominant-negative effects on viral gene expression. A second complication is that correct Rev-mediated RNA localization is a prerequisite to lentiviral RNA packaging, so that even RNAs that possess intact packaging signals are inefficiently packaged without Rev. A third issue is that HIV-1 VLPs can efficiently assemble in the absence of their native genomes and, in the process, package similar amounts of cellular RNAs. Another complication relates to the fact that some previous packaging studies were conducted using transfection experiments in which the cellular concentration of mutant vector RNAs was high. At high cellular concentrations, the mutant RNAs can be packaged at high levels, as long as they do not have to compete with RNAs containing the wild-type 5′-L. In addition, efforts to determine the secondary structure of the HIV-1 5′-L using chemical probing, mutagenesis and biochemical approaches, and phylogenetic analyses have led to multiple structural proposals and mechanistic predictions, thus hindering the design of mutations to properly test structure/ function hypotheses.

Recently, using an NMR approach for studying relatively large RNAs, we showed that the HIV-1 5′-L exists as an equilibrium of two conformationally distinct species, one in which residues spanning the gag start codon (G328-A356; hereafter called “AUG”) form a hairpin and the GC-rich palindromic loop of the dimer promoting DIS hairpin is base paired with an upstream element (U5), and another in which AUG is base paired with U5 and the dimer-promoting loop of DIS is exposed. The AUG formation, proposed originally on the basis of phylogenetic and biochemical analyses, promotes dimerization by displacing the DIS and simultaneously exposes high-affinity NC binding sites, thereby promoting the packaging of a dimeric genome. Having identified the conformer of the 5′-L that directs packaging, we wished to explore the potential roles of specific elements within the leader on RNA packaging. Our present study identifies a 159-nucleotide segment of the 5′-L that is sufficient for RNA dimerization and NC binding in vitro and is required for efficient and selective RNA packaging.

**Results**

**Construct design**

A primary goal of these studies was to systematically remove elements from 5′-L (residues 1–356) without altering the folding of other structural elements, as assessed by the NC binding, dimerization, and NMR spectroscopic properties of the RNA. Under conditions of physiological-like ionic strength (PI buffer: 140 mM KCl, 10 mM NaCl, 1 mM MgCl₂, and 10 mM Tris–HCl, pH 7.0) and at an RNA concentration of ∼1.0 μM, 5′-L exists as an equilibrium mixture of two species: a monomeric...
conformation in which the GC-rich palindrome of the DIS hairpin is sequestered by base pairing with U5 and AUG forms a hairpin, and a dimeric conformation in which AUG base pairs with U5 and the DIS is exposed (Fig. 1c) (dimer dissociation constant $K_d = 0.9 \pm 0.1 \mu M$).46 Mutations in AUG that affect base pairing with U5 influence the monomer–dimer equilibrium. Lane 1, native 5′-L exists as a mixture of monomer and dimer when incubating within PI buffer. Lane 2, substitution of 331-GAGAGUGGUG-GAGAGCGUCCGU-355 by 331-ACUACUCAUCCGAGACGUUUUGA-355, which has been reported in our previous study known as 5′-L HP-AUG, favors the AUG hairpin structure and exists mainly as a monomer. This mutation has been shown to inhibit packaging.46 Lane 3, substitution of 333-UGUGGUGG-339 by 333-CUCUAGA-339, which favors the AUG hairpin structure, mainly exists as a monomer and has been reported with packaging deficiency.47 Lane 4, substitution of 337-UGUGGUGG-353 by 337-CGGGCACAAGAAAAA-353 (5′-LU5:AUG) favors the formation of U5:AUG base pairs and promotes dimerization and packaging.46 Lane 5, 5′-L344, in which residues 345–356 were truncated from 5′-L. The AUG hairpin cannot form in 5′-L344, but the residues involved in U5:AUG base pairs are maintained. This truncation stabilizes the RNA as a dimer. (d) 5′-L344 readily forms a dimer upon incubation in PI buffer. (e) Portion of the 2D NOESY spectrum obtained for a dimeric 5′-L344 RNA sample containing protons on the non-exchangeable adenosine C-2 and ribose (A2,R), the guanosine ribose (GR), and the cytosine ribose (CR) carbons and deuterons at all other non-exchangeable aromatic and ribose carbons: A2,R,GR,CR-[5′-L344]2. NMR signal assignments were made based on comparisons with isolated oligo-RNAs corresponding to the TAR, lower-PBS, DIS, and Ψ-RNA hairpins (K.L., X.H., and M.S., unpublished results).
in which all non-exchangeable aromatic protons except the C2-H of the adenosines were substituted by deuterium. For example, $^1$H NMR signals with chemical shifts and NOE cross-peak patterns matching those observed for the isolated TAR, DIS, and $\Psi$ site oligo-RNA hairpins were readily identified in 2D NOESY spectra obtained for a dimeric 5′-L344 sample that contained protons on the non-exchangeable adenosine C-2 and ribose (A2,p), the guanosine ribose (Gp), and the cytosine ribose (Cp) carbons and deuterons at all other non-exchangeable aromatic and ribose carbons (A2,p,Gp,Cp-[5′-L344]2, Fig. 1e) (superscript p: protonated on the C-8, C-2, and ribose carbons, respectively). These outlier signals and their NOE interactions with H1' protons were monitored to assess potential effects of mutations and deletions on the structure of the RNA.

Additional leader RNAs that contained deletions of the TAR hairpin (5′-L344-ΔTAR), the upper region of the PBS loop (5′-L344-ΔPBS), and the TAR and Poly(A) hairpins and upper PBS loop (5′-L344-ΔTAR-ΔPoly(A)-ΔPBS) were prepared (Fig. 2a). RNA homogeneity was confirmed by denaturing polyacrylamide gel electrophoresis (PAGE, Fig. 2b). Base-pairing patterns of the TAR and Poly(A) hairpins and the lower stem of the PBS loop have been supported by several studies,64,72–76 including compensatory mutagenesis-based packaging experiments.77 The PBS deletion construct was designed such that the entire upper region of the loop (residues A132–A216) was substituted by a GAGA tetraloop (a member of the well-characterized GNRA family78–85). Resolved A220 C2-H signals were observed for all constructs in partially deuterated samples (prepared as described above), including those containing both the native and truncated PBS loops (Fig. 2c), indicating that base pairing in the lower stem of the PBS loop is maintained in both the native and mutant 5′-L RNAs. 2D NOESY data obtained for the 5′-L344 mutants did not exhibit line broadening or additional NMR signals (other than those expected for the GAGA tetraloop) that would be indicative of alternative folding. In particular, the outlier adenosine C2-H signals of the most highly truncated construct, 5′-L344-ΔTAR-ΔPoly(A)-ΔPBS, exhibited $^1$H NMR chemical shifts and 2D NOESY cross peaks similar to those observed for the wild-type 5′-L344 RNA (Fig. 2d). Signals assigned to the TAR, Poly(A), and upper PBS loop structures were missing in spectra obtained for the corresponding deletion mutants, supporting the $^1$H NMR chemical shift assignments for these elements.

### Effects of TAR, Poly(A), and PBS deletions on dimerization and NC binding

The 5′-L344, 5′-L344-ΔTAR, 5′-L344-ΔPBS, 5′-L344-ΔTAR-ΔPBS, and 5′-L344-ΔTAR-ΔPoly(A)-ΔPBS RNAs all exist predominantly as monomers at low ionic strength but form dimers upon incubation in PI buffer (Figs. 1d and 3a), indicating that the TAR, Poly(A), and upper PBS hairpin structures do not play a significant role in dimerization.

---

**Fig. 2.** (a) RNA constructs used in the present study: (1) 5′-L, (2) 5′-L344, (3) 5′-L344-ΔTAR, (4) 5′-L344-ΔPBS, (5) 5′-L344-ΔTAR-ΔPBS, and (6) 5′-L344-ΔTAR-ΔPoly(A)-ΔPBS. (b) Denaturing PAGE gels show sample purity and their relative electrophoretic migration. (c) Portions of 2D NOESY spectra showing the A220 C2-H peaks observed for (from left to right) A2,p,Gp,Cp-[5′-L344]2, for 5′-L344-ΔPBS, for 5′-L344-ΔTAR-ΔPoly(A)-ΔPBS, and for a fully protonated oligoribonucleotide corresponding to the lower PBS hairpin (5′-ggCUCUGGUgagaGCCAGAGcc; lower case, non-native), showing that the lower PBS stem structure is maintained in these 5′-L constructs. (d) Chemical shifts of the C2-H signals for the PBS lower stem, $\Psi$, and DIS hairpins are unaffected by removal of the TAR, Poly(A), and PBS upper loop [gray and black spectra correspond to RNAs labeled 2 and 6 in (a)].
Identification of a Minimal HIV-1 Packaging Signal

Two-site binding mode was utilized to characterize the high-affinity NC binding observed for 5′-L344 (and at diminished levels in the other 5′-L344 mutants) since similar ITC parameters were measured for 5′-L344 and NC (Table 1). The first set of binding parameters was used to characterize the high-affinity NC binding. The post-high-affinity interactions between 5′-L344 and NC (Table 1). (c) ITC NC titration data obtained upon titration of NC into 5′-L344, showing that these deletions have modest effects on high-affinity NC binding. (d) ITC NC titration data obtained for 5′-L344-ΔTAR-ΔPoly(A)-ΔPBS and 5′-L344-ΔTAR-ΔPoly(A). The raw data are shown in the top panel, and the integrated heat changes are shown in the bottom panel. The two exothermic components were deleted before nonlinear least-squares isotherm fitting using “Two Sets of Sites” mode (MicroCal Origin 5.0). The first binding parameters were determined as in Fig. 2 for form dimers after incubation in PI buffer (140 mM KCl, 10 mM NaCl, 1 mM MgCl2, and 10 mM Tris-HCl, pH 7.0), indicating that deletions of TAR, Poly(A), and PBS do not affect RNA dimerization.

Table 1. Thermodynamic parameters from ITC measurements

<table>
<thead>
<tr>
<th>5′-L344</th>
<th>5′-L344-ΔTAR</th>
<th>5′-L344-ΔTAR-ΔPoly(A)</th>
<th>5′-L344-ΔTAR-ΔPBS</th>
<th>5′-L344-ΔTAR-ΔPoly(A)-ΔPBS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_1$</td>
<td>$K_{d1}$ (nM)</td>
<td>$AH_1$ (kcal/mol)</td>
<td>$-T\Delta S_1$ (kcal/mol)</td>
<td>$N_2$</td>
</tr>
<tr>
<td>15±1</td>
<td>17±9</td>
<td>-12.5±0.2</td>
<td>1.7±0.5</td>
<td>31±4</td>
</tr>
<tr>
<td>13±1</td>
<td>17±15</td>
<td>-14.0±2.0</td>
<td>2.8±2.9</td>
<td>30±3</td>
</tr>
<tr>
<td>11±1</td>
<td>16±15</td>
<td>-16.7±1.8</td>
<td>5.7±1.3</td>
<td>43±9</td>
</tr>
<tr>
<td>9±1</td>
<td>21±15</td>
<td>-15.5±1.9</td>
<td>4.7±1.9</td>
<td>31±3</td>
</tr>
</tbody>
</table>

Two-site binding mode was utilized to fit the ITC isotherms from NC titrations. For those with no data (marked with an em dash), the magnitude of the endothermic component was too small for reliable calculation.
appearance to those obtained for 5′-L and 5′-L\textsuperscript{344} (Fig. 3b and c), except that these truncated RNAs bound fewer NC molecules with high affinity. Deletion of the TAR and PBS loops resulted in losses of 2±1 and 4±1 NC binding sites, respectively, and the simultaneous deletion of both the TAR and PBS loops eliminated 6±1 NC sites (Table 1), indicating that NC binds independently and non-cooperatively to these hairpins. The isolated oligo-TAR RNA was previously shown to interact with NC,\textsuperscript{86} and an oligo-RNA corresponding to the isolated TAR hairpin was found by ITC to bind a single NC molecule with high affinity ($K_d = 110±2$ nM) (X.H. and M.S., unpublished results).

Removal of the Poly(A) hairpin from 5′-L\textsuperscript{344−ΔTAR−ΔPBS} resulted in the loss of a single additional high-affinity NC site (Table 1). Thus, TAR, PBS, and Poly(A) hairpins appear to contribute approximately 2, 4, and 1 high-affinity NC binding sites, respectively. Interestingly, we also observed small but systematic reductions in the magnitude of the endothermic component of the ITC isotherms upon removal of TAR, Poly(A), and PBS elements (Fig. 3b and c), with isotherms for 5′-L\textsuperscript{344−ΔTAR−ΔPoly(A)−ΔPBS} exhibiting only a slight endothermic feature at the higher NC/RNA ratios (Fig. 3c). Thus, the NC-induced RNA unwinding that occurs after saturation of the high-affinity NC binding sites appears to diminish with decreasing size of the RNA.

**Effects of TAR, Poly(A), and PBS deletions on RNA packaging**

When HIV-1 particles bud from cells in the absence of genomic RNAs with packaging signals, these particles encapsidate “random” host mRNAs roughly in proportion to their intracellular concentrations.\textsuperscript{87} In contrast, most such adventitiously packaged mRNAs are displaced when RNAs with authentic packaging signals are present, due to the specificity of HIV-1 for its own genome. Here, a packaging assay based on these observations was used to address whether or not the TAR, Poly(A), and PBS elements (Fig. 3b and c), with isotherms for 5′-L\textsuperscript{344−ΔTAR−ΔPoly(A)−ΔPBS} exhibiting only a slight endothermic feature at the higher NC/RNA ratios (Fig. 3c). Thus, the NC-induced RNA unwinding that occurs after saturation of the high-affinity NC binding sites appears to diminish with decreasing size of the RNA.

---

**Identification of a Minimal HIV-1 Packaging Signal**

![Fig. 4. TAR, Poly(A), and PBS deletions do not affect RNA packaging. RNA packaging efficiency of 5′-L variants was monitored by RPA. (a) RNA samples obtained from the co-expression of a series of packaging test vectors with CMVΔR8.2, an HIV-1 Ψ\textsuperscript{−} helper construct. Lanes 1–5, products of RNA samples harvested from media of cells transfected with the indicated constructs plus pCMVΔR8-2 helper; lane 6, from cells transfected with pCMVΔR8-2 alone; lane 7, from mock-transfected cells. Lanes 8–14 are RPA products of cellular samples from the same transfections used in lanes 1–7. P, undigested riboprobe; ΔpA = Δpoly(A). The mobilities of riboprobe fragments protected by transcripts from each of the test vectors and from the Ψ\textsuperscript{−} helper are indicated on the right. (b) RPA of viral samples using two different riboprobes specific to 7SL RNA or HIV-1 vector RNA (Ps, the two undigested riboprobes). Lanes 1–5, products of RNA samples harvested from media of cells transfected with the indicated constructs plus pCMVΔR8-2 helper; lane 6, from cells transfected with pCMVΔR8-2 alone; lane 7, from mock-transfected cells. The mobilities of riboprobe fragments protected by transcripts from each of the test vectors and from the Ψ\textsuperscript{−} helper are indicated on the right.**
activities (Fig. 4a) or viral 7SL RNA levels (Fig. 4b) as controls. 7SL RNA is a cellular transcript that is incorporated independent of genomic RNA and can be used as an internal control for quantitating packaging efficiencies.\(^1\) Relative efficiencies were determined by comparing packaging levels of test RNAs that contained 5'-L mutations with those that contained the intact, wild-type leader and gag sequences. All test vectors contained Rev-responsive elements and were expressed in the presence of Rev, and thus Rev-dependent RNA localization, which is required for efficient packaging,\(^2\) was not a variable in these experiments.

As shown in the RNase protection assay (RPA) in Fig. 4a, when the \(\Psi\) helper was transfected into 293T cells in the absence of RNAs possessing HIV-1 packaging signals, its transcript was highly expressed in cells (lane 13) and was encapsidated into virions that were released from these cells with modest efficiency (lane 6). In contrast, whereas both transcripts were detected at similar levels within cells (lane 8), when CMV\(\Delta R8.2\) was co-transfected with an HIV-1 test vector possessing the native NL4-3 5'-L (Native), RNAs with native NL4-3 packaging elements and were expressed alone (lane 6). In contrast to this \(\Psi\) transcript, 5'-L RNAs from which either the PBS loop or TAR was deleted (lanes 2–3), those that lacked both TAR and the PBS loop (lane 4), and the TAR, Poly(A), and PBS loop triple-deletion test RNA (lane 5), were all encapsidated at levels similar to those of native 5'-L RNA (lane 1). As quantified in Table 2, only small decreases in overall packaging were observed, with all of these deleted 5'-L RNAs displaying packaging >80% that observed for the native 5'-L.

**Effects of gag deletion on packaging**

To determine if elements in gag downstream of the AUG are important for packaging, we generated packaging vector variants that retained the 5'-L sequences studied by NMR above, but that lacked HIV-1 gag-derived sequences downstream of position +350 that were present in the native 5'-L vectors examined above. As shown in RPA results in Fig. 5, co-expression of either the original 5'-L or a truncated 5'-L\(^3\) vector lacking gag sequences with the CMV\(\Delta R8.2\) helper construct yielded indistinguishable RNA levels in cells (lanes 5 and 6) and identical strong packaging levels in virions (lanes 1 and 2; Table 2). These findings demonstrate that sequences in gag are not required for efficient packaging of HIV-1 5'-L containing RNAs into VLPs.

**Effect of TAR deletion on 5'-L structure**

The observation that deletion of TAR results in a moderate but reproducible reduction in RNA packaging, whereas deletion of the PBS loop does

---

**Table 2. Packaging efficiencies of vector RNAs containing native and mutant HIV-1 5'-L sequences**

<table>
<thead>
<tr>
<th>Test RNA</th>
<th>%Packaging/RT(^a)</th>
<th>%Packaging/(\Psi) helper(^b)</th>
<th>%Packaging/7SL(^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'-L (Native)</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>5'-L-(\Delta)TAR</td>
<td>41 ± 32</td>
<td>98 ± 4</td>
<td>86 ± 4</td>
</tr>
<tr>
<td>5'-L-(\Delta)PBS</td>
<td>133 ± 77</td>
<td>100 ± 3</td>
<td>94 ± 4</td>
</tr>
<tr>
<td>5'-L-(\Delta)TAR-(\Delta)PBS</td>
<td>37 ± 20</td>
<td>90 ± 6</td>
<td>82 ± 1</td>
</tr>
<tr>
<td>5'-L-(\Delta)TAR-(\Delta)Poly(A)-(\Delta)PBS</td>
<td>86 ± 59</td>
<td>100 ± 7</td>
<td>89 ± 11</td>
</tr>
<tr>
<td>5'-L(^4)^</td>
<td>101 ± 5</td>
<td>100 ± 1</td>
<td>ND</td>
</tr>
</tbody>
</table>

Data reported as mean ± standard deviation obtained from three independent experiments except data for 5'-L\(^4\), which are obtained from two independent experiments.

ND, not measured.

\(^a\) %Test RNA/virion based on RT activity to quantify virus production. Data normalized relative to test RNA containing the native 5'-L.

\(^b\) %Test RNA relative to the \(\Psi\) helper RNA in particles normalized relative to test RNA containing the native 5'-L.

\(^c\) %Test RNA/virion using 7SL packaging to quantify virus production. Data normalized relative to test RNA containing the native 5'-L.
not affect packaging, is difficult to reconcile with our finding that the PBS upper loop contains a greater number of high-affinity NC binding sites. 2D NOESY spectra obtained for \(5′-L_{344}\) appeared similar to spectra obtained for the parent \(5′-L_{344}\) RNA (except for the missing TAR signals; data not shown), suggesting that the TAR deletion does not affect the overall fold of the dimeric RNA. To determine if the TAR deletion affects the folding of the monomeric form of the leader, we examined the dimerization properties and 2D NOESY spectra of leader RNAs in which AUG was fully truncated \(5′-L_{ΔAUG}\). As indicated above, this RNA lacks the \(5′\)-residues of AUG necessary to displace and expose the DIS hairpin, and like full-length \(5′-L\) RNAs containing mutations that stabilize the AUG hairpin, \(5′-L_{ΔAUG}\) exists predominantly as a monomer in PI buffer (Fig. 6a). Unexpectedly, deletion of the TAR hairpin from \(5′-L_{ΔAUG}\) \(5′-L_{ΔAUG-ΔTAR}\) strongly promoted dimerization in PI buffer (Fig. 6b). In addition, whereas the outlying adenosine C2-H signals in 2D NOESY spectra obtained for \(5′-L_{ΔAUG}\) were similar in appearance to spectra obtained for \(5′-L_{344}\) (Fig. 6b), the spectrum obtained for \(5′-L_{ΔAUG-ΔTAR}\) lacked several expected signals, including the A268 C2-H signal associated with the properly folded form of the DIS hairpin (Fig. 6c). These findings indicate that removal of TAR leads to non-native-like folding and dimerization of the monomeric form of the \(5′-L\).

Discussion

The TAR, Poly(A), and PBS loops have each been implicated in at least some studies to play roles in genome dimerization, NC binding, and packaging. Isolated TAR oligo-RNA hairpins can form dimers in the presence of NC, and electron microscopy images of NC-incubated HIV-1 \(5′\)-RNA fragments (nucleotides 1–744) exhibit distinct circular structures that have been interpreted in terms of a TAR-
mediated dimer interface. Some studies suggested that deletion of TAR promotes the packaging of monomeric RNAs, to an extent similar to that observed upon deletion of the DIS hairpin. On the other hand, truncation of the TAR hairpin from a 290-nucleotide 5'-fragment of the HIV-1 leader promoted, rather than inhibited, dimerization of this RNA. The present studies show that 5'-L RNA, which spans the entire HIV-1 5'-untranslated region, forms a stable dimer even in the absence of the TAR hairpin. 2D NOESY spectra are relatively unperturbed (except for the loss of the TAR signals) upon removal of TAR, indicating that dimerization does not result from artificial misfolding of the dimeric leader. The fact that the A46 C2-H signals of TAR are relatively intense in both the monomeric and dimeric forms of 5'-L RNAs (Fig. 1e) suggests that the hairpin is relatively mobile and probably does not participate in long-range interactions of the dimeric 5'-L (although TAR could participate in intermolecular RNA–RNA interactions in the more stable dimer that is formed in virions during viral maturation).

Deletion of TAR was shown in one study to reduce packaging efficiency by 74% compared to the packaging of a parent RNA derived from the intact viral genome that encoded a stop codon within the capsid domain of gag. Although TAR plays a critical role in transcriptional activation via its interactions with the viral Tat protein, these studies concluded that reductions in packaging were too great to be attributed to the relatively modest reductions in cellular RNA levels. Mutations that destabilized the stem of TAR reduced packaging to 70% of wild-type levels (after control for reductions in cellular RNA levels) and altered the ratio of spliced/unspliced viral RNAs that were packaged. Interestingly, using a proviral expression system that is largely independent of Tat for transcriptional activation, Helga-Maria et al. found that missense mutations in the upper stem of TAR had little to no effect on packaging, but mutations at the base of the TAR stem could lead to significant reductions in packaging. Compensatory mutations in the TAR stem restored packaging to near-wild-type levels, indicating that it is the structure of TAR, and not the TAR sequence, that is important for packaging. A more recent study showed that the complete removal of TAR had little impact on replication as long as other changes were made to overcome the requirement for TAR-mediated transcriptional activation (packaging efficiencies were not reported). These studies also showed that there is evolutionary pressure to restore base pairing at the 5'-terminal end of the viral RNA and supported proposals that packaging defects induced by mutations in TAR result from misfolding of the 5'-L. Our findings are consistent with this latter hypothesis. Thus, deletion of TAR did not significantly affect the NC binding properties of the 5'-L RNA and also did not significantly impair vector RNA packaging. The modest packaging defects that we observed upon deletion of TAR are most likely due to misfolding of the monomeric form of the 5'-L, which we believe leads to a reduction in the cellular levels of the packageable, dimeric form of the RNA.

The Poly(A) hairpin has also been proposed to play roles in both genome dimerization and packaging. Although early computer-modeling studies and electron microscopy imaging of the HIV-1 genome suggested that the palindromic 5'-AAGCUU-3' loop of Poly(A) might form a second dimer linkage structure, subsequent mutagenesis studies indicated that the Poly(A) palindrome does not influence RNA dimerization. Here, we showed that deletion of the TAR and Poly(A) hairpins does not affect the dimerization properties of 5'-L RNA in support of the latter conclusion. Poly(A) was originally thought to be important for RNA packaging because removal of the hairpin reduced packaging to ~10% of wild-type levels. Mutations that disrupt the Poly(A) hairpin structure were also shown to inhibit genome packaging and viral replication, whereas compensatory mutations restored packaging to near-wild-type levels, indicating that it is the structure, rather than the specific sequence of the Poly(A) hairpin that is required for packaging. However, more recent studies indicated that disruption of the Poly(A) hairpin structure can lead to reductions in cellular levels of the viral RNA due to reductions in RNA stability, and it was suggested that the Poly(A) hairpin may not actually play a significant role in genome packaging. As shown in the present studies, removal of Poly(A) from 5'-L ΔTAR does not significantly affect its structure, dimerization, or NC binding properties. Furthermore, vector RNAs in which the TAR, Poly(A), and PBS loops are deleted were packaged as well as RNAs that contained only TAR and PBS deletions, indicating that Poly(A) does not play a significant role in RNA packaging.

The PBS element contains a sequence that binds human tRNA Lys, a requirement for initiation of reverse transcription. Numerous secondary structural models of PBS have been proposed based on mutational studies, enzymatic probing, UV cross-linking, phylogenetic studies, and chemical modification assays (reviewed in Refs. 13, 39, 64, 71, 73, 74, 76, and 108). Deletions within the upper loop of the PBS generally do not significantly impair genome packaging, whereas the lower stem loop is required for RNA dimerization and efficient packaging. Our PBS deletion mutant was designed to maintain native base pairing in the lower stem of the PBS loop structure. Significantly, the NMR data presented here show that the A220 C2-H signal pattern was observed in [5'-L ΔTAR]2 and all the PBS upper loop
truncated constructs, indicating that the lower stem of the PBS element adopts the predicted secondary structure. Removal of the upper PBS loop structure eliminated ~4 high-affinity NC binding sites but did not impair RNA packaging, suggesting that these NC sites may play roles in tRNA binding and/or reverse transcription but are not critical for packaging.

In vitro biochemical studies and phylogenetic predictions led to the proposal that loop residues of PolyA base pair with residues G443–C449 in the matrix encoding region of gag. These putative long-range interactions have been supported by chemical probing and modeling studies, but their role in packaging has not been established. Early packaging studies suggested that efficient HIV-1 genome packaging requires residues within the first several hundred nucleotides of gag, and could therefore be dependent on Poly(A)–gag interactions. However, the present studies show that RNAs containing the native 5′-L (nucleotides 1–355, which include residues required for U5:AUG base pairing) but missing nearly the entire gag open reading frame are packaged with the same efficiency as RNAs containing the intact gag gene. Thus, if the Poly(A)–gag interactions exist in vivo, they likely play roles in functions unrelated to RNA packaging.

The present findings also provide insights into the number and general locations of the high-affinity NC binding sites. Early studies using UV cross-linking indicated that NC is the only major protein that binds to the genomic RNA in HIV-1 virions, and it was estimated that approximately 15 NC molecules are bound per RNA strand. In contrast, in vitro cross-linking experiments involving a fragment of the 5′-L spanning from DIS through residue 170 of gag were interpreted to indicate that only ~2–3 NC or Gag proteins bind to each RNA strand, and RNA footprinting studies performed with a 5′-L construct that included the intact 5′-untranslated region and 66 nucleotides of gag were similarly interpreted in terms of only two high-affinity NC or Gag binding sites. In addition, quantitative filter-binding and ITC assays indicated that NC can bind with high affinity (Kd ≈ 55–400 nM) to fragments of the HIV-1 5′-L between the DIS and AUG hairpins.

Our ITC data of 5′-L and the truncated mutants’ titration with NC show that they all have high-affinity NC binding sites with Kd in the same range (Table 1). Our current and recent studies are most consistent with the earlier work of Darlix et al. and indicate that the 5′-L contains approximately 16 high-affinity NC binding sites per RNA strand. The locations of the high-affinity NC binding sites were probed recently by monitoring the effects of zinc ejection on SHAPE activity. These studies suggested that the 5′-L contains 7 NC binding sites, none of which were proposed to reside within the TAR, Poly(A), or upper PBS loops. Although these results are incompatible with our present findings, we cannot rule out the possibility that structural changes associated with viral maturation, or constraints within the context of the packaged RNA, could lead to altered NC binding.

Sakuragi et al. have used a novel approach to identify the minimal determinants of HIV-1 genome dimerization and packaging that involves duplication of the 5′-L in downstream (ectopic) regions of the RNA, which can lead to the unusual packaging of monomeric RNAs (in addition to dimeric RNAs). In this system, RNAs with ectopic sequences lacking the Poly(A) and SD hairpins (to prevent unwanted splicing and polyadenylation) were efficiently packaged as monomers. This group subsequently showed that RNAs lacking the native 5′-L, but containing two downstream copies of a sequence that includes TAR through the first half of gag, but that lacks the Poly(A) and PBS hairpins (a 19-nucleotide deletion in the upper loop to eliminate ectopic initiation of reverse transcription), exclusively package monomeric RNAs, although virion production and packaging efficiency were reduced to ~20% and ~63% of wild-type levels, respectively. Using a similar packaging system that included a much smaller portion of gag, these investigators showed that deletions in the upper loop of the PBS do not significantly affect dimerization or packaging, but deletions in U5 and/or the lower PBS stem lead to significant reductions in packaging. Surprisingly, severe modifications and deletions within the DIS hairpin led to relatively modest reductions in packaging levels (to 18–50% of WT levels). A systematic deletion mutagenesis study revealed that relatively efficient packaging of monomeric RNAs can be achieved by RNAs that contain a fully native 5′-L and an ectopic element that lacks the TAR, Poly(A), upper PBS loop, and SD sequences. Since the RNAs employed contained a native 5′-L sequence, these studies do not rule out the possibility that packaging could depend on the presence of at least one fully native leader RNA per dimer. However, the findings are nevertheless fully compatible with our observation that, under the experimental conditions used here, relatively efficient packaging can be achieved by RNAs that lack the TAR, Poly(A), and upper PBS loops.

In summary, our findings identify 5′-L Δ344–ΔTAR-ΔPoly(A)–ΔPBS as an essential HIV-1 core encapsidation element. This fragment of the 5′-L exhibits 1H NMR spectral features similar to those of the intact 5′-L, possesses native-like in vitro dimerization and NC binding properties, and is required for promoting RNA packaging into VLPs. Packaging deficiencies observed here and by others upon deletion of TAR are likely due to misfolding of the monomeric form of the 5′-L. The clustering of high-affinity NC binding sites, coupled with the known self-association propensities of the CA domain of Gag, could function synergistically to promote cooperative assembly of the Gag:RNA complex that is trafficked
to the plasma membrane and nucleates virus assembly. Since the minimal packaging signal comprises only 159 nucleotides, high-resolution structural studies by NMR or combined NMR/cryo-electron tomographic approaches\textsuperscript{117} should now be feasible.

**Materials and Methods**

**Plasmid construction**

The 5′-L, 5′-L\textsuperscript{ULP-AUG}, and 5′-L\textsuperscript{HP-AUG} mutants were cloned into PUC19 as previously described.\textsuperscript{46} 5′-L\textsuperscript{344-Δ TAR} and 5′-L\textsuperscript{344-Δ TAR} were subcloned from the 5′-L plasmid using 5′-CGG TCG AAT TCT AAT ACG ACT CAC TAT AGG TCT TCC TGC TTA GTA CAC ATC TGA TGG TCG GAG CTC TCT GCC G3′- and 5′-CGG TCG AAT TCT AAT ACG ACT CAC TAT AGG CAC TGT TTA AGC CTC AAT AAA GCT TGC CTT GAT GTC-3′ as forward primer, respectively, and 5′-CCC TAG GAT CCC CCG GGC GCA CCC ATC TCT CTC TCT CTA-3′ as their mutual reverse primer. 5′-L\textsuperscript{344-Δ PBS} was made by MegaPrimer mutagenesis method using 5′-L\textsuperscript{344} as template. The four primers used are 5′-CGG TCG AAT TCT AAT ACG ACT CAC TAT AGG TCT TCC TGG TTA GTA CAC ATC TGA TGG TCG GAG CTC TCT GCC GAC-3′, 5′-CTT CCG TCG AGA GAT CTC CTC TGG TCT CCC AGA GTC ACA CAA CAG ACG GCC ACA CAC TAC TTT GAG CA-3′, 5′-GAG ACC AGA GAT CTC CTC TGG TCG AGC GAC TCG GCT TGC TGG AGA CGG CAA GAG GCG AGG GGC GCC GAC-3′, and 5′-CCC TAG CTA GCT TCT GCC GCA CCC ATC TCT CTC TCT CTA-3′. All of the mutations were sent to the DNA sequencing facility (University of Maryland, Baltimore County) to ensure that the desired mutations were obtained.

The packaging test vector 5′-L\textsuperscript{ΔL344-Δ TAR-L344} was then cloned from 5′-L\textsuperscript{ΔL344-Δ PBS-L344} plasmid using the forward primer 5′-CGG TCG AAT TCT AAT ACG ACT CAC TAT AGG CAC TGC TTA GTA CAC ATC AAT AAA GCT TGC CTT GAT GTC CTT GAG ATC-3′ and 5′-L\textsuperscript{344-Δ TAR-Δ Poly(A)-Δ PBS} were then cloned from 5′-L\textsuperscript{ΔL344-Δ PBS-L344} as forward primer, respectively, and the mutual reverse primer 5′-CCC TAG GAT CCC CCG GCC GCA CCC ATC CTC TCT CTC CTA-3′. All of the mutations were sent to the DNA sequencing facility (University of Maryland, Baltimore County) to ensure that the desired mutations were obtained.

The packaging test vector 5′-L\textsuperscript{ΔL344-Δ TAR-L344} was then cloned from 5′-L\textsuperscript{ΔL344-Δ PBS-L344} plasmid using the forward primer 5′-CGG TCG AAT TCT AAT ACG ACT CAC TAT AGG CAC TGC TTA GTA CAC ATC AAT AAA GCT TGC CTT GAT GTC CTT GAG ATC-3′ and 5′-L\textsuperscript{344-Δ TAR-Δ Poly(A)-Δ PBS} were then cloned from 5′-L\textsuperscript{ΔL344-Δ PBS-L344} as forward primer, respectively, and the mutual reverse primer 5′-CCC TAG GAT CCC CCG GCC GCA CCC ATC CTC TCT CTC CTA-3′. All of the mutations were sent to the DNA sequencing facility (University of Maryland, Baltimore County) to ensure that the desired mutations were obtained.

**RNA packaging experiments**

Human 293T cells were cultured and maintained in Dulbecco’s modified Eagle’s medium (Invitrogen) supplemented with 10% fetal bovine serum (Gemini). Cells were grown at 37 °C under 5% CO\textsubscript{2}. To generate virions for RNA packaging analysis, we co-transfected 293T cells with a total of 20 μg of plasmid DNA that included one test vector plasmid and pCMVΔR8.2, a ψ'-HIV-1 helper vector driven by the CMV promoter 87, at an approximate 1:3 molar ratio on 100-mm plates. Modest, empirically determined variation in co-transfection ratios was introduced for non-native vectors, to achieve similar test/helper expression ratios for all samples. For transfection by polyethyleneimine, plasmid DNA was mixed with 4 μg of polyethyleneimine (Polysciences) per microgram of DNA in 1 ml of 150 mM NaCl by vortexing at medium speed for 10 s. After room temperature incubation for 15 min, the mixture was added dropwise to medium on cells. The transfection mixture was replaced with fresh Dulbecco’s modified Eagle’s medium 24 h after transfection.

Tissue culture medium was harvested at 48 h post-transfection, pooled, and filtered through 0.2-μm-pore size filters. Virus-containing medium was concentrated by centrifugation through a 2 ml 20% sucrose cushion in phosphate-buffered saline for 2 h at 4 °C and 25,000 rpm using a Sorvall Surespin 630 rotor in a Sorvall Discovery 90 ultracentrifuge. Viral pellets were suspended in 0.5 ml of Trizol reagent (Invitrogen), and RNA was isolated according to the manufacturer’s instructions. Cellular RNA was obtained by harvesting cells 48 h after transfection by scraping cells into 2 ml of Trizol reagent per 100-mm plate, and RNA was isolated according to the manufacturer’s instructions.

Riboprobe templates were used for phosphorimager as previously described.\textsuperscript{48} Riboprobes were then isolated from RNA and quantified. A riboprobe designed to contain the HIV-1 gag open reading frame was then prepared by transcription by T7 polymerase\textsuperscript{176} in solution with a mixture of transcription buffer, MgCl\textsubscript{2}, NTPs, and DNA template.
Identification of a Minimal HIV-1 Packaging Signal

The amount of MgCl₂ and NTPs was optimized for each RNA sample in order to guarantee best RNA yield. The RNA was then purified using 6% denaturing PAGE gels and recovered via Elutrap Electroelution System (Whatman).

NC purification

The 55-residue NC protein (from HIV-1 strain NL4–3) was expressed from the bacteria vector pRD2, which was transformed into BL21 (DE3) pLysE. The protein was purified under non-denaturing conditions as previously described.¹²⁰

Dimerization assay

RNA was prepared in 10 mM Tris–HCl first, heat denatured for 3 min, and cooled on ice; 10× ITC salt was then added to the RNA sample to reach a final salt concentration of 140 mM KCl, 10 mM NaCl, 1 mM MgCl₂. After incubating at 37 °C for 1, 2, 4, 24, and 48 h, the RNA samples were loaded to 2% agarose gel prestained with ethidium bromide and run at 110 V in 1× TB buffer (44.5 mM Tris–HCl, pH 7.0, 140 mM NaCl, 10 mM NaCl, 1 mM MgCl₂ and 100 μM β-mercaptoethanol) was loaded into the injection syringe. The calorimetry cell was loaded with RNA (0.8 μM). The RNA sample (in 10 mM Tris–HCl) was heat denatured and then cooled on ice; 10× ITC salt was then added to reach a final salt concentration of 140 mM KCl, 10 mM NaCl, and 1 mM MgCl₂. The RNA sample was then incubated overnight at 37 °C before loading into the calorimetry cell. After thermal equilibration at 30 °C and the initial 60-s delay, 28 serial injections of NC protein (10 μl each) were made into the calorimetry cell. The thermodinamic parameters of the NC–RNA interaction were determined by a nonlinear squares fit of the data using two-site binding mode.

Isothermal titration calorimetry

ITC experiments were carried out using a VP-ITC MicroCalorimeter (MicroCal Corp., Northampton, MA; 1989). NC (180 μM) in ITC buffer (10 mM Tris–HCl, pH 7.1), 140 mM NaCl, 10 mM NaCl, 1 mM MgCl₂, and 100 μM β-mercaptoethanol) was loaded into the injection syringe. The calorimetry cell was loaded with RNA (0.8 μM). The RNA sample (in 10 mM Tris–HCl) was heat denatured and then cooled on ice; 10× ITC salt was then added to reach a final salt concentration of 140 mM KCl, 10 mM NaCl, and 1 mM MgCl₂. The RNA sample was then incubated overnight at 37 °C before loading into the calorimetry cell. After thermal equilibration at 30 °C and the initial 60-s delay, 28 serial injections of NC protein (10 μl each) were made into the calorimetry cell. The thermodinamic parameters of the NC–RNA interaction were determined by a nonlinear squares fit of the data using two-site binding mode.

References

Identification of a Minimal HIV-1 Packaging Signal


Identification of a Minimal HIV-1 Packaging Signal


