Identification of the initial nucleocapsid recognition element in the HIV-1 RNA packaging signal

Pengfei Ding[a,b], Siarhei Kharytonchyk[c], Alexis Waller[a], Ugonna Mbaekwe[a], Sapna Basappa[a], Nansen Kuo[a], Heather M. Frank[a], Christina Quasney[a], Aaron Kidane[a], Canessa Swanson[a], Verna Van[a], Mitali Sarkar[a], Emily Cannistraci[a], Ridhi Chaudhary[a], Hana Flores[a], Alice Telesnitsky[c,1], and Michael F. Summers[a,b,1]

[a]Department of Chemistry and Biochemistry, University of Maryland, Baltimore County, Baltimore, MD 21250; [b]Howard Hughes Medical Institute, University of Maryland, Baltimore County, Baltimore, MD 21250; and [c]Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, MI 48109

Selective packaging of the HIV-1 genome during virus assembly is mediated by interactions between the dimeric 5′-leader of the unspliced viral RNA and the nucleocapsid (NC) domains of a small number of assembling viral Gag polyproteins. Here, we show that the dimeric 5′-leader contains more than two dozen NC binding sites with affinities ranging from 40 nM to 1.4 μM, and that all high-affinity sites ($K_d \lesssim$ 400 nM) reside within a ~150-nt region of the leader sufficient to promote RNA packaging (core encapsidation signal, Ψ^{CES}). The four initial binding sites with highest affinity reside near two symmetrically equivalent three-way junction structures. Unlike the other high-affinity sites, which bind NC with exothermic energetics, binding to these sites occurs endothermically due to concomitant unwinding of a weakly base-paired [UUUU]:[GGAG] helical element. Mutations that stabilize base pairing within this element eliminate NC binding to this site and severely impair RNA packaging into virus-like particles. NMR studies reveal that a recently discovered small-molecule inhibitor of HIV-1 RNA packaging that appears to function by stabilizing the structure of the leader binds directly to the [UUUU]:[GGAG] helix. Our findings suggest a sequential NC binding mechanism for Gag-genome assembly and identify a potential RNA Achilles' heel to which HIV therapeutics may be targeted.

HIV-1 | genome | RNA | packaging | nucleocapsid

Selective incorporation of authentic viral genomes into assembling virions is essential for viral replication (1). Like all retroviruses except the spumaviruses, the HIV type 1 (HIV-1) specifically packages its positive-sense, unspliced genomic RNA (gRNA) in the form of a noncovalently linked dimer (2–6). Both strands of the dimeric gRNA are used for strand transfer-mediated recombination during reverse transcription (7–9), which enhances HIV-1 fitness by maintaining genetic integrity and promoting genetic diversity under natural or chemotherapeutic stresses (10, 11). Although the full-length gRNA only accounts for <1% of all cytosolic RNAs (including viral spliced RNAs and nonviral RNAs), the dimeric gRNA exists in >90% of all newly assembled progeny virions (12). This highly selective packaging process is mediated by specific interactions between a *cis*-acting element of the HIV-1 gRNA and the nucleocapsid (NC) domain of the major structural polyprotein Gag (13–15).

The HIV-1 packaging signal is located within the 5′-leader (5′-L) of the viral genome (2, 6, 16), which is among the most conserved regions of the viral RNA (3, 17). The leader contains discrete stretches of nucleotides with independent and sometimes overlapping functions: a *trans*-activation response region (TAR) that stimulates transcription, a primer binding site for initiation of reverse transcription (PBS), a dimer initiation site that promotes genome dimerization (DIS), the major splice donor site (SD), a polyadenylation signal (polyA), and the *gag* start codon (AUG) (3, 18–22). The major determinants for HIV-1 genome packaging were initially assigned to a relatively short segment surrounding conserved residues predicted to form a hairpin (called Ψ) (23–25), and later expanded to include a much larger region

of the leader (26–28). Combinations of nucleotide accessibility mapping and mutagenesis studies indicate that the secondary structure of the dimeric 5′-L that is selected for packaging differs from that of the monomer (6, 29–31). Although gRNA versus mRNA functions were originally thought to be modulated by a riboswitch-like mechanism, in which the function of a single RNA transcript is modulated by changes in its structure (29, 32, 33), it now appears transcript dimerization and fate are controlled at the level of transcription by heterogeneous start site usage. Thus, transcripts that begin with two or three sequential 5′-guanosines form monomers in vitro and are retained in cells for splicing and/or translation, whereas those that begin with a single 5′-guanosine form dimers and are selectively packaged into virions (34–36). A combination of mutagenesis, NMR, and RNA packaging experiments identified a region of the leader sufficient for packaging heterologous vectors into assembling virions (called the core encapsidation signal, Ψ^{CES}) (37), which has been shown by NMR to adopt a tandem three-way junction structure (30).

gRNAs are recognized by the NC domain of Gag, which contains two copies of a conserved "CCHC"-type zinc finger

Significance

Understanding the molecular determinants of retroviral genome packaging is important for drug discovery and development of vectors for gene delivery. We show that the HIV-1 leader, which contains the RNA elements necessary for genome packaging, binds approximately two dozen copies of the cognate NC protein with affinities ranging from ~40 nM to 1.4 μM. Binding to the four highest-affinity "initial" binding sites occurs with endothermic energetics attributed to NC-induced localized RNA melting. Mutations that stabilize these sites inhibit NC binding in vitro and RNA packaging in transfected cells. A small-molecule inhibitor of RNA packaging binds specifically to the initial NC binding sites and stabilizes the RNA structure. Our findings identify a potential RNA Achilles' heel for HIV therapeutic development.

See online for related content such as Commentaries.

BIOCHEMISTRY

domain that are critical for genome packaging (25, 38). Structural studies of NC bound to several 5′-L fragments all show that tight binding is mediated by a combination of electrostatic interactions and sequestration of exposed guanosine bases within hydrophobic clefts of the zinc fingers (39–42). Confocal studies indicate that genomes are sequestered at plasma membrane assembly sites by a small number of Gag proteins, probably a dozen or fewer, and that this initial complex nucleates further Gag recruitment and virus assembly (43, 44). In vitro assembly of immature virus-like particles can also be stimulated by the presence of the dimeric $[\Psi^{CES}]_2$ RNA (45).

The present studies focused on identifying the mechanism that guides assembly of Gag on the HIV-1 packaging signal. We found that the HIV-1 dimeric 5′-L contains more than two dozen NC binding sites with affinities spanning from ∼40 nM to 1.4 μM, and that binding sites with highest affinity map within the $[\Psi^{CES}]_2$ dimer. Our studies also revealed that initial binding to the highest-affinity binding sites occurs with localized and concomitant unwinding of a short and dynamic helical structure located near a tandem three-way junction structure, and that the lability of this helix is important for genome packaging and viral replication. Implications of our findings for virus assembly and antiviral development are presented.

Results

Dimeric 5′-L Binds More than Two Dozen NC Proteins with Affinities that Vary by an Order of Magnitude. The 5′-L of the HIV-1 RNA genome includes the 5′-untranslated region and the first few codons of *gag* (residues 1 to 356 for HIV-1$_{NL4-3}$) (Fig. 1*A*) (29). Dimerization is initiated by intermolecular base pairing between the palindromic "GCGCGC" sequences within the DIS regions of two 5′-L RNAs, followed by the formation of extensive intermolecular interactions (Fig. 1*B*) (31). In the dimer-promoting conformation of the full-length 5′-L, residues of the unique-5′ (U5) region base pair with residues near the *gag* start site (U5:AUG interactions) leaving residues A345–A356 as a single-stranded unstructured region (Fig. 1*A*). Previous studies have shown that truncation of these disordered 3′-residues downstream of the AUG start site shifts the equilibrium of the 5′-L from the monomer to the dimer conformation (31, 37). Thus, the leader construct we used in the current study contains residues 1 to 344 (5′-L$_{344}$ or 5′-L), which readily forms dimers when incubated under buffer conditions with physiological-like ionic strength (PI buffer, 140 mM KCl, 10 mM NaCl, 1 mM MgCl$_2$, pH 7.0) (see Fig. 4*D*). All RNA constructs used in this study are derivatives of this dimeric 5′-L with extended intermolecular base pairing (*SI Appendix*, Fig. S1).

Previous isothermal titration calorimetry (ITC) studies from our laboratory showed that the dimeric 5′-L gave rise to a bipartite NC-binding isotherm under PI buffer conditions (29, 37). Initially, high-affinity exothermic bindings were observed, followed by a subsequent endothermic event associated with weak, nonspecific NC binding and NC-induced RNA unfolding at higher NC-to-RNA ratios. Higher concentrations of Mg^{2+} (e.g., 5 mM) were later found to eliminate the nonspecific binding and unfolding (29, 30). Therefore, all of the present ITC studies were performed with a PI buffer containing 5 mM MgCl$_2$, allowing for accurate fitting of the high-affinity specific NC binding data. Interestingly, while defining the high-affinity exothermic and the Mg^{2+}-sensitive nonspecific binding modes, we noticed that the first one or two data points were not consistent with a simple exothermic high-affinity binding event. The first two data points were typically excluded in prior ITC studies of NC binding to the HIV-1 leader (29, 37). However, these unusual data points were also observed in the titrations with 5 mM Mg^{2+} (gray data point in Fig. 1*C*), and an early endothermic binding event was clearly detected using reduced NC-to-RNA ratios (Fig. 1*D*). Saturation of this endothermic binding at an NC-to-RNA ratio of ∼10

suggests the presence of approximately four NC binding sites with affinities significantly higher than those of the exothermic binding sites.

Fitting of the ITC data using a single-set-of-sites model with the initial "outlier" data points deleted (gray data point due to endothermic binding in Fig. 1*C*) indicates that the dimeric 5′-L binds 28 ± 1 NC molecules with an average K_d of 880 ± 30 nM (Table 1). This average NC-binding affinity of the dimeric 5′-L is much lower than that of the well-studied Ψ-hairpin apical loop, which contains a single high-affinity site (K_d = 320 ± 30 nM) (Table 1), indicating that NC binding to the intact, dimeric leader occurs with heterogeneous binding affinities. Consistent with this hypothesis, an RNA construct comprising the TAR-PolyA tandem hairpin was found to have two weak NC binding sites with an average K_d of 1,320 ± 140 nM (*SI Appendix*, Fig. S2B). A leader construct without the PBS region (replaced by a "GAGA" tetraloop, 5′-L$^{\Delta PBS}$) binds 23 ± 1 NC with decreased average K_d (760 ± 50 nM; *SI Appendix*, Fig. S2A), suggesting that the PBS region only has two to three weak NC binding sites. Thus, the more than two dozen specific NC binding sites of the dimeric 5′-L are notably heterogeneous in terms of both binding energetics and binding affinities.

High-Affinity NC Binding Sites Reside within the Core Encapsidation Signal. The finding that the TAR-PolyA and PBS regions have very weak NC binding sites is consistent with the previous reports that Ψ^{CES}, which lacks the TAR-PolyA and PBS elements, is the minimal region of the 5′-L required for packaging (30, 37). NC titrations with $[\Psi^{CES}]_2$ gave rise to a two-component ITC binding profile with quality sufficient for two-sets-of-sites model fitting (Fig. 2*A* and Table 1). The initial endothermic component of the isotherm corresponds to the binding of four NC proteins with average K_d values of 45 ± 19 nM, consistent with results obtained for the intact leader. Fits for the exothermic portion of the isotherm are indicative of ∼16 NC binding sites with moderate average affinities (K_d = 330 ± 20 nM), similar to the isotherms and affinities observed for the apical loop of the Ψ-hairpin (Table 1 and *SI Appendix*, Fig. S2C). These results indicate that the high-affinity (endothermic) binding sites reside within the Ψ^{CES} region of the 5′-L.

The $[\Psi^{CES}]_2$ structure comprises a central A-helical duplex that connects two tandem three-way junctions (30). To identify the locations of the NC binding sites within $[\Psi^{CES}]_2$, particularly the initial endothermic sites, we prepared three additional RNA constructs corresponding to the DIS duplex (residue 237 to 281 plus a nonnative G–C pair at each end of the duplex; *SI Appendix*, Fig. S1F) (46), the tandem three-way junction (all residues of Ψ^{CES} except the DIS helix, which was substituted by a "GAGA" tetraloop) (Ψ^{T-3WJ}; *SI Appendix*, Fig. S1E), and the U5:AUG-H1-Ψ three-way junction (Ψ^{3WJ-1}; *SI Appendix*, Fig. S1G). The NC binding stoichiometry and affinities of these $[\Psi^{CES}]_2$ subfragments were characterized by ITC (Fig. 2 *B–D*). The DIS dimer binds six NC molecules with exothermic energetics and affinity (average K_d = 340 ± 30 nM) similar to that observed for the apical loop of the Ψ-hairpin (Table 1). In contrast, the high-affinity endothermic phase of the NC binding energetics was retained by both the Ψ^{T-3WJ} and Ψ^{3WJ-1} constructs, indicating that the initial binding site resides within Ψ^{3WJ-1}. Two-sets-of-sites fitting of the ITC data for Ψ^{3WJ-1} revealed that this fragment contains 2.2 ± 0.2 high-affinity endothermic NC binding sites (K_d = 28 ± 15 nM) and 2.7 ± 0.3 additional exothermic NC binding sites (K_d = 320 ± 180 nM). The exothermic binding sites across different subdomains of $[\Psi^{CES}]_2$ exhibit similar affinities with K_d in the range of 300 to 400 nM (Table 1). Thus, all of the high-affinity NC binding sites of the dimeric 5′-L are clustered within $[\Psi^{CES}]_2$, with the initial endothermic binding located within the Ψ^{3WJ-1} region of the RNA.

NC Binds Initially to a Conserved [UUUU]:[GGAG] Element. To identify the specific nucleotides that contribute to high-affinity endothermic
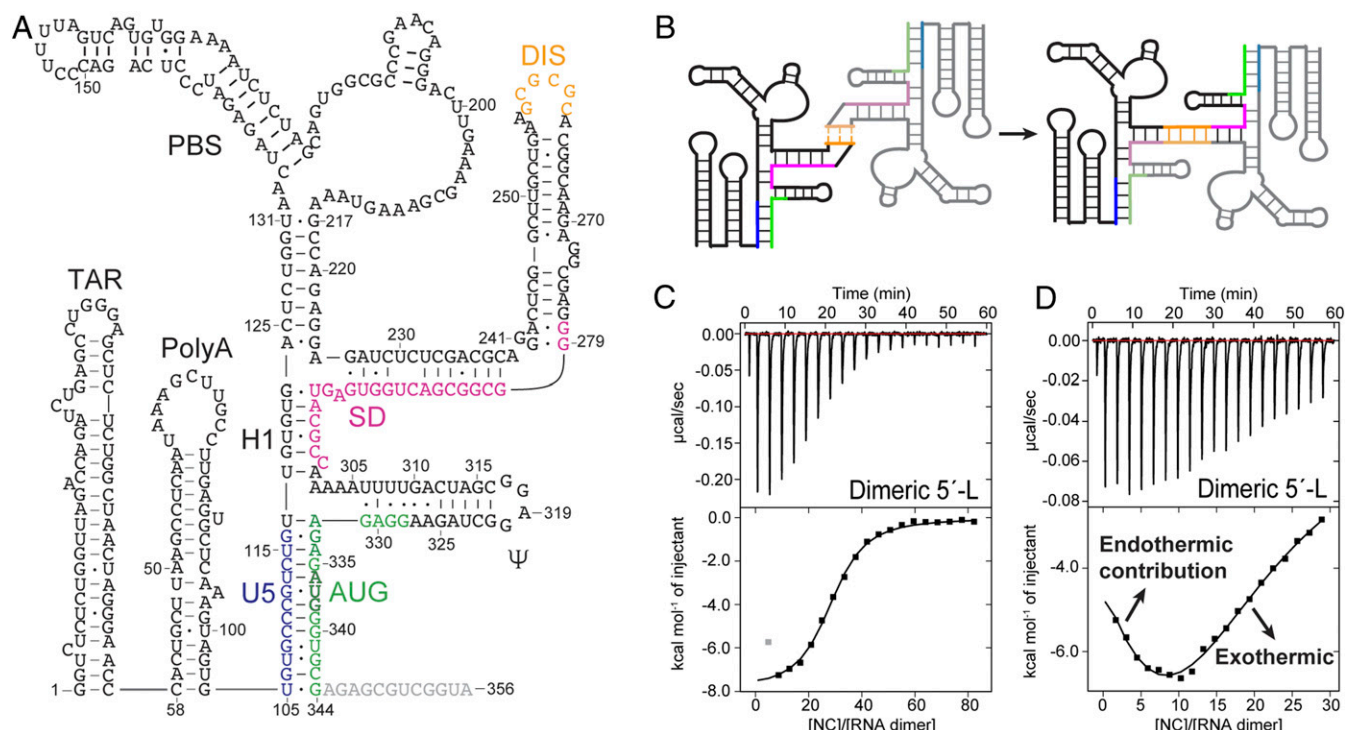
Fig. 1. The dimeric 5′-L contains more than two dozen NC binding sites. (*A*) Secondary structure of the HIV-1 5′-L in the dimer promoting conformation. The U5, DIS, SD, and AUG sequences are colored in blue, yellow, magenta, and green, respectively. (*B*) Dimerization of 5′-L initiated by DIS "kissing" interaction and stabilized by extensive intermolecular interactions. ITC isotherms for the dimeric 5′-L titrated by NC with high (*C*) or low (*D*) protein-to-RNA ratios. The gray calorimetric data point was not included in data fitting.

binding, NMR-detected NC binding data were obtained for Ψ^{3WJ-1}. The structure of the junction was previously characterized in the context of a larger RNA (155 nt) comprising half of the symmetric $[\Psi^{CES}]_2$ dimer (called Ψ^{CESm}) (30). The improved NMR spectral quality obtainable for the 73-nt Ψ^{3WJ-1} construct enabled unambiguous assignment of nearly all of the nonexchangeable aromatic and ribose H1′ protons, facilitating structural and NC binding analyses.

All NMR signals from the fingerprint region of ^{2}H-edited two-dimensional (2D) nuclear Overhauser effect spectroscopy (NOESY) spectra for A^H-, G^H-, C^H-, or U^H-labeled Ψ^{3WJ-1} RNAs (corresponding to RNAs with fully protonated adenosines, guanosines, cytosines, and uridines, respectively, with all other nucleotides in the sample perdeuterated) were assigned using standard methods (*SI Appendix*, Fig. S3) (47). ^{2}H-edited 2D NOESY spectra were

also obtained for $A^{2r}G^r–$ and $A^{2r}G^rU^r–\Psi^{3WJ-1}$ samples (superscripts denote sites of protonation; e.g., $A^{2r}G^r$ corresponds to a sample with protonated adenosine-C2 and ribose carbons, and protonated guanosine ribose carbons, with all other nucleotides perdeuterated) (Fig. 3*B*). A311-H2 gave rise to a well-resolved NMR signal and exhibits cross-strand NOEs with the H2 and H1′ protons of both A326 and A327. Similar results were observed for A327-H2, which exhibits cross-strand NOEs with the H1′ protons of G310 and A311, providing evidence for the formation of non-canonical G310:A327 and A311:A326 base pairs. The cross-strand NOEs between A330-H2 and U308-H1′, as well as A330-H2 and U309-H1′, indicate that the "UUUU" and "GGAG" sequences are close in 3D structure (the two H2–H1′ proton pairs separated by <5.5 Å) and likely form a [UUUU]:[GGAG] duplex with multiple G:U wobble base pairs (Fig. 3*A*). However, the weaker

Table 1. NC binding stoichiometry and affinity of HIV-1 5′-L RNA constructs

Constructs	Endothermic		Exothermic	
	N	K_d, nM	*N*	K_d, nM
5′-L	N/D	N/D	28.0 ± 1.1	880 ± 30
5′-L$^{\Delta PBS}$	N/D	N/D	22.6 ± 1.1	760 ± 50
Ψ^{CES}	4.0 ± 0.9	45 ± 19	15.9 ± 0.9	330 ± 20
Ψ^{T-3WJ}	2.8 ± 0.2	43 ± 18	5.2 ± 0.2	430 ± 50
Ψ^{3WJ-1}	2.2 ± 0.2	28 ± 15	2.7 ± 0.3	320 ± 180
DIS	N/A	N/A	6.6 ± 0.2	340 ± 30
Ψ^{HP}	2.3 ± 0.2	26 ± 21	2.8 ± 0.3	1,540 ± 240
$\Psi^{HP-stable}$	N/A	N/A	1.2 ± 0.3	300 ± 150
Ψ-fragment	N/A	N/A	1.8 ± 0.1	450 ± 30
Ψ-apical loop	N/A	N/A	1.2 ± 0.1	320 ± 30
TAR-PolyA	N/A	N/A	1.9 ± 0.1	1,320 ± 140

N/A, not applicable; N/D, not determined. Data are represented as mean ± SEM (*n* = 3).
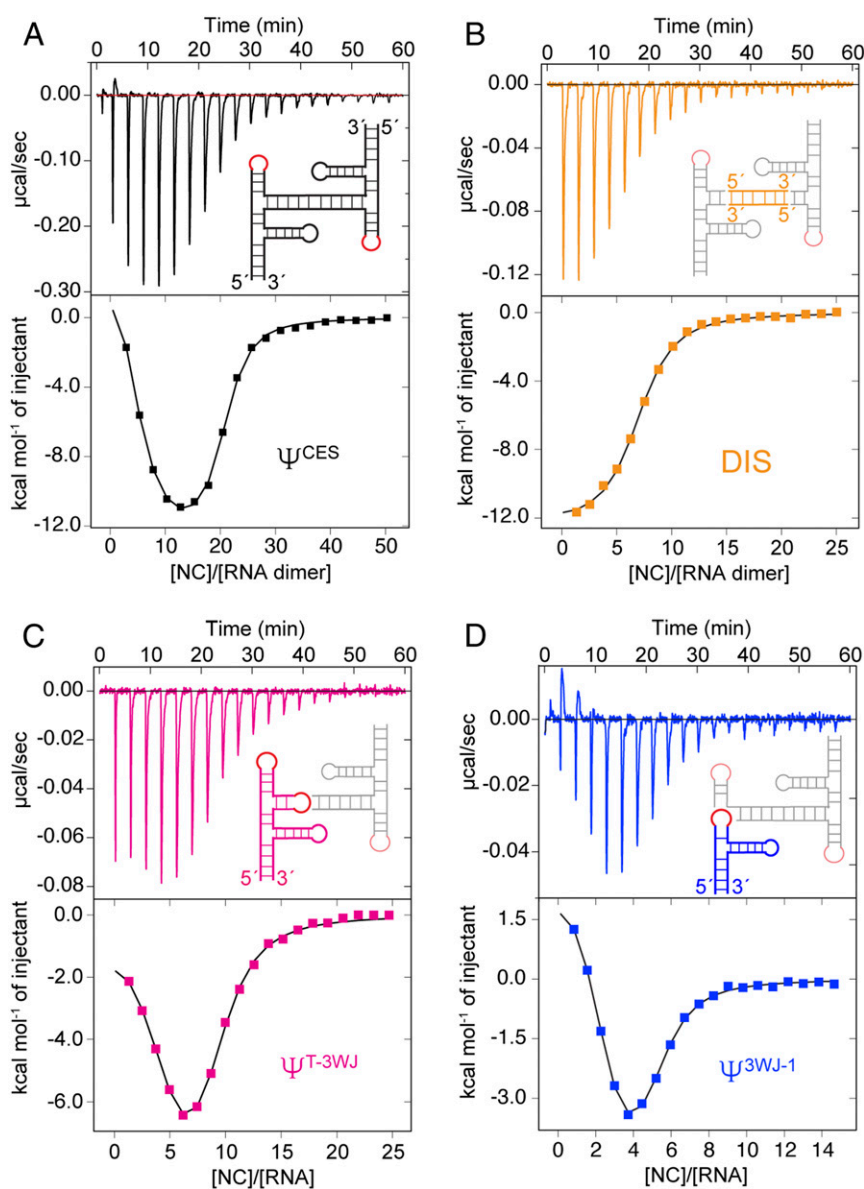
Fig. 2. Localization of NC binding sites in [Ψ^{CES}]$_2$. Secondary structure models for [Ψ^{CES}]$_2$ (*A*) and its subdomains (*B*–*D*) are shown with ITC isotherms. Nonnative "GAGA" tetraloop is colored in red, the DIS region in yellow (*B*), Ψ$^{T–3WJ}$ region in pink (*C*), and the Ψ$^{3WJ–1}$ in blue (*D*).

cross-strand NOE between A330-H2 and U308-H1′, in comparison with the sequential A330-H2 to G331-H1′ NOE, suggests a dynamic nature of the interstrand interactions of the [UUUU]:[GGAG] region (Fig. 3*B*). The structure differs from one proposed in a previous NMR study in which the residues of the [UUUU]:[GGAG] stem were proposed to be nonbase paired and partially disordered (48), possibly due to differences in the constructs or experimental conditions employed, but is consistent with NMR findings for a larger 5′-L fragment (30).

NMR titration experiments were subsequently performed to identify the nucleotides directly involved in NC binding. Surprisingly, adenosine-H2 signals of residues A304, A305, A311, A326, A327, and A330, which reside within or near the [UUUU]:[GGAG] helix, were all substantially perturbed when a twofold excess of NC was added, whereas signals of A314, A319, and A324 from the well-known NC binding Ψ-apical loop were barely affected (Fig. 3 *A* and *C*). Since HIV-1 NC binds with high affinity to unpaired or weakly paired guanosines (39), we also

performed NC titration experiment with a G^H-labeled Ψ$^{3WJ–1}$ sample (Fig. 3 *D* and *E*). Consistent with findings from the adenosine-H2 detected titrations, signals for G328 and G329 of the [UUUU]:[GGAG] duplex were significantly broadened, whereas the guanosines of the Ψ-apical loop were relatively unaffected (Fig. 3 *D* and *E*). Thus, these data provide direct evidence that binding of NC to the guanosines of the dynamic [UUUU]:[GGAG] region are responsible for the initial high-affinity endothermic binding.

[UUUU]:[GGAG] Lability Is Required for High-Affinity NC Binding. We next asked whether the endothermic nature of the binding is due to the disruption of the G:U wobble base pairs in the [UUUU]:[GGAG] helix. To stabilize the stem, two of the uridines within this helix (U308 and U309) were substituted by cytosines, thereby converting two G:U base pairs in Ψ$^{3WJ–1}$ to G–C pairs. Interestingly, these substitutions completely eliminated high-affinity endothermic NC binding to Ψ$^{3WJ–1}$ (Fig. 4*A*). These findings suggest that
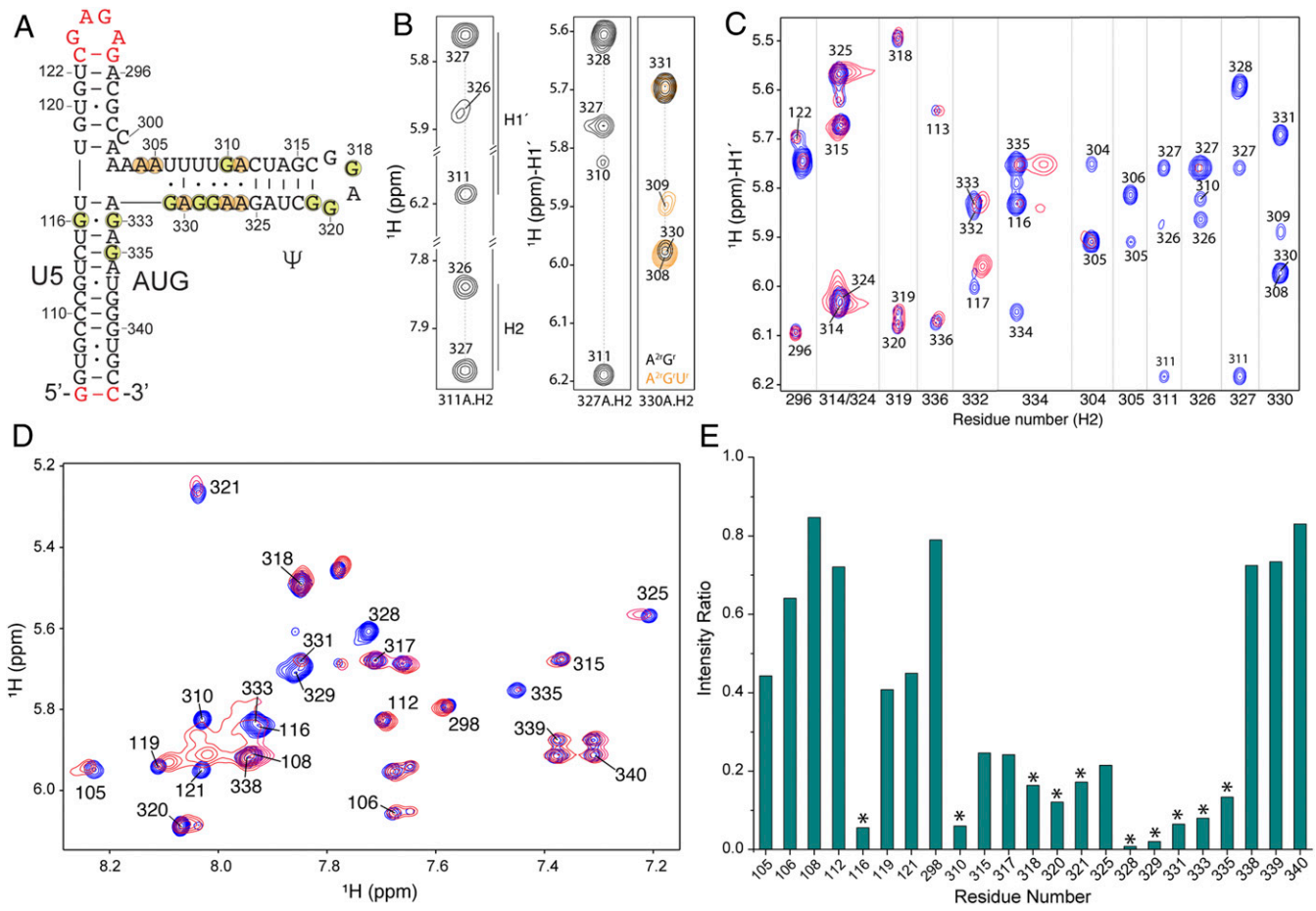
Fig. 3. Structure and NC binding characterization of Ψ^{3WJ-1}. (*A*) Secondary structure of Ψ^{3WJ-1}. Nonnative residues are shown in red. Adenosines with chemical shifts significantly perturbed by NC binding are highlighted in orange, and guanosines in yellow. (*B*) Selected region of the 2D NOESY spectra of A^{2r}G^r– and A^{2r}G^{r}U^r–Ψ^{3WJ-1}. Cross-strand NOEs confirm the structure of the [UUUUAG]:[AAGGAG] region. (*C*) Two-dimensional NOESY spectra overlay of A^{2r}G^{r}U^r–Ψ^{3WJ-1} with (red) and without (blue) twofold excess of NC. The H2–H1′ cross-peak region is shown. (*D*) Two-dimensional NOESY spectra overlay of G^H–Ψ^{3WJ-1} with (red) and without (blue) twofold excess of NC. Residue numbers are shown for the H8–H1′ cross-peaks. (*E*) Intensity ratios of the protein-bound (red) and the free- (blue) RNA NMR signals in *D* are plotted against the residue number. Residues with severe signal intensity reduction are marked with asterisks.

structural lability of the [UUUU]:[GGAG] helix may be important for high-affinity endothermic NC binding.

To further investigate structural and NC binding features of the [UUUU]:[GGAG] helix, two Ψ-hairpin RNAs that extend through the [UUUU]:[GGAG] element were prepared, one with 5′- and 3′-terminating residues that cannot form Watson–Crick base pairs (5′-GA and AG-3′, respectively; Ψ^{HP}) (*SI Appendix,* Fig. S1*H*), and the other with helix-stabilizing 5′-GG and CC-3′ terminating residues at the 5′- and 3′-ends (Ψ$^{HP-stable}$) (Fig. 4*C* and *SI Appendix,* Fig. S1*I*). NMR chemical shifts and cross-peak patterns observed for the nonstabilized Ψ^{HP} construct were similar to those of corresponding residues in Ψ^{3WJ-1} (*SI Appendix,* Figs. S4 and S5), indicating that the structures are similar. Spectra for both constructs exhibited sequential H8/ H6-to-H1′ and cross-strand adenosine-H2 to H1′ NOEs for residues 326-AAGGAG-331 and 306-UUUUGA-311 (spectrum for Ψ^{HP} shown in *SI Appendix,* Fig. S4) consistent with a helical structure. However, Watson–Crick imino signals for the [UUUU]:[GGAG] element were broadened beyond detection in ^{1}H NMR data obtained for Ψ^{HP} but were readily observable in spectra obtained for Ψ$^{HP-stable}$ (Fig. 4*C*), indicating that the [UUUU]:[GGAG] element of Ψ^{HP} is structurally labile. Importantly, as observed for Ψ^{3WJ-1}, NC titrations with Ψ^{HP} afforded

two-component ITC isotherms with initial high-affinity endothermic binding (*n* = 2.3 ± 0.2; K_d = 26 ± 21 nM) followed by exothermic binding (*n* = 2.8 ± 0.3; K_d = 1.5 ± 0.2 μM), whereas the RNA containing helix-stabilizing G–C closing base pairs (Ψ$^{HP-stable}$) exhibited only exothermic NC binding behavior (*n* = 1.2 ± 0.3; K_d = 300 ± 150 nM) (Fig. 4*B* and Table 1). These findings further indicate that high-affinity endothermic NC binding to [UUUU]:[GGAG] occurs with concomitant unfolding and/or structural rearrangement of this element. Notably, a modified Ψ^{HP} construct that lacks the "UUUU" nucleotides and leaves the GGAG-3′ element unpaired did not bind NC with the same high affinity as observed for the intact RNA (*SI Appendix,* Fig. S2*D*). Moreover, the Ψ-apical loop and the [UUUU]:[GGAG] element both contain an identical "GGAG" sequence but exhibit distinct binding energetics and affinities (Table 1). Thus, it appears that the lability and the structure of the [UUUU]:[GGAG] element are both required for high-affinity NC binding.

[UUUU]:[GGAG] Element Promotes Efficient HIV-1 RNA Packaging. To investigate whether the [UUUU]:[GGAG] element is important for HIV-1 genome packaging, we compared the packaging efficiencies of vector RNAs containing the wild-type and U308/309C mutated leaders. The vector with the wild-type leader (L688-RRE-Puro-LTR)
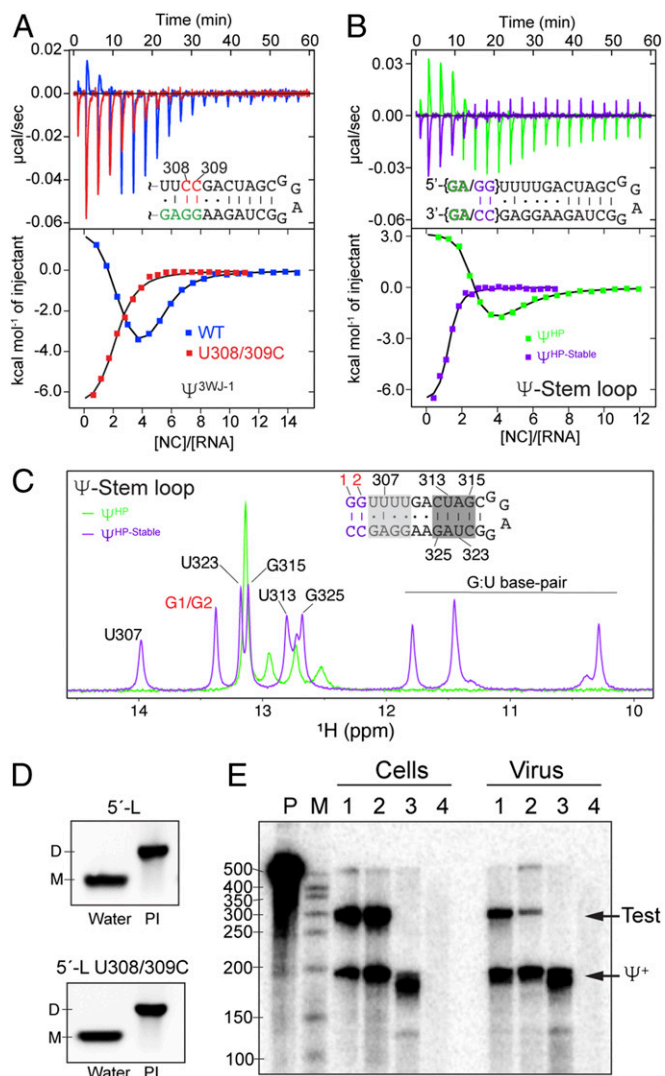
Fig. 4. Stabilizing the stem region of Ψ-hairpin eliminates high-affinity endothermic binding and impairs in vivo genome packaging. (*A*) ITC data for wild-type Ψ^{3WJ-1} (blue) and the U308/309C mutant (red). (*B*) ITC data for $\Psi^{HP\text{-stable}}$ (purple) and Ψ^{HP} (green). (*C*) Overlay of 1D imino proton spectra for $\Psi^{HP\text{-stable}}$ (purple) and Ψ^{HP} (green). Imino proton chemical shifts assignments are shown for $\Psi^{HP\text{-stable}}$. (*D*) Dimerization gel shift for wild-type 5′-L and the U308/309C mutant. (*E*) The G:U to G–C mutations significantly reduced the RNA packaging in competition with the wild-type RNA. Lanes 1 and 2 are native HIV-1 5′-L versus test vectors containing wild-type 5′-L and G:U to G–C mutant, respectively. Lane 3 is HIV-1 5′-L helper expressed without test RNA. Lane 4 is mock transfected cells.

was previously shown to display competitive RNA packaging levels indistinguishable from those of HIV-1 gRNAs (37). The U308/309C mutations did not affect the dimerization property of the 5′-L in vitro (Fig. 4*D*), as expected since they also did not significantly perturb the structures of well-characterized 5′-L fragments (see above and *SI Appendix*, Fig. S6). However, the packaging efficiency of the U308/309C mutant was substantially diminished compared to that of the wild-type vector (13 ± 2%) (Fig. 4*E*). Thus, the U308/309C double mutation that stabilizes the [UUUU]:[GGAG] element and prevents high-affinity endothermic NC binding in vitro profoundly impairs HIV-1 RNA packaging in transfected cells.

An HIV-1 RNA Packaging Inhibitor Targets the [UUUU]:[GGAG] Element. The requirement of the [UUUU]:[GGAG] element for high-affinity NC binding and RNA packaging suggested that this region of the

leader might serve as a potent HIV-1 therapeutic target. Recently, a small-molecule quinolinium derivative (NSC 260594; hereafter referred to as NSC) was shown to inhibit HIV-1 RNA packaging, apparently by binding to the Ψ-hairpin region of the HIV-1 5′-L and exerting a global stabilizing effect on the RNA structure (49). NSC was originally identified in a screen for small molecules that inhibit Gag binding to the apical "GGAG" loop of the Ψ-hairpin (50). We extended those studies to characterization of NSC interactions with the larger Ψ^{3WJ-1} RNA using ^{2}H-edited NMR. Many ^{1}H NMR signals of NSC exhibited substantial chemical shifts (and some line broadening) when bound to a $A^{2r}G^{r}U^{r}$-labeled Ψ^{3WJ-1} sample (compare Fig. 5 *A*, *Top*, and Fig. 5 *A*, *Bottom*), suggesting specific interactions between the compound and RNA. Intermolecular NOEs between the RNA ribose protons and the compound were also distinguishable (Fig. 5*B*). Several adenosine-H2 signals of Ψ^{3WJ-1} were significantly perturbed and exhibited reduced intensities (Fig. 5*A*). Two-dimensional NOESY data obtained for $A^{2r}G^{r}U^{r}$-labeled Ψ^{3WJ-1} revealed that H2 signals of adenosines near the apical loop of the Ψ-hairpin (A314, A319, and A324) were barely perturbed upon NSC binding, whereas the H2 signals of adenosines near stem region of the Ψ-hairpin exhibited substantial broadening (Fig. 5 *C* and *D*). These findings indicate that the NSC RNA packaging inhibitor targets residues of the [UUUU]:[G-GAG] element of the HIV-1 packaging signal and not the apical loop of the Ψ-hairpin.

Conclusions

Early efforts to identify HIV-1 RNA packaging determinants identified stretches of nucleotides surrounding (and including) the Ψ-hairpin that, if deleted, resulted in severe HIV-1 packaging defects (23–25), and this led to proposals that the Ψ-hairpin is a major RNA packaging determinant. Studies suggested that the apical tetraloop of the Ψ-hairpin was the critical Gag/NC recognition element (48, 51), and its complex structure with NC revealed that two exposed guanosines from the "GGAG" loop are inserted into hydrophobic clefts on the surfaces of the two zinc finger domains of NC (39). Subsequent studies showed that substitution of the GGAG loop by GCUA abolished NC binding but did not significantly affect RNA packaging (16, 52), and it soon became clear that regions other than the apical loop of the Ψ-hairpin, including sequences in the stem of that hairpin (52, 53), are important for competitive RNA packaging (6, 26, 37, 54).

The present findings reveal that the conserved [UUUU]:[G-GAG] element located at the base of the Ψ-hairpin and adjacent to a tandem three-way junction structure is critical both for high-affinity NC binding and competitive RNA packaging. Each of these two elements in the dimeric 5′-L binds two NC molecules with affinities (~40 nM) approximately an order of magnitude greater than those of other NC binding sites within the encapsidation-promoting region of the viral leader, and more than 30 times tighter than NC binding sites in other regions of the leader (Table 1). The absence of this [UUUU]:[GGAG] element from spliced viral transcripts may contribute to the selective packaging of the full-length gRNA. The [UUUU]:[GGAG] element adopts a labile helical structure in the absence of NC and undergoes structural changes, possibly helix melting, upon endothermic binding by NC. These findings appear consistent with chemical probing studies showing that residues of [UUUU]:[GGAG] are resistant to RNase V1 cleavage in the absence of NC but exhibit enhanced chemical accessibility after Gag/NC binding (48, 55).

High-affinity NC binding sites identified in the RNA packaging signals of other retroviruses also appear to be structurally dynamic (56–60). For example, the Moloney murine leukemia virus (MoMuLV) NC binds specifically to a "UCUG" single-stranded region that links two stem loops (DIS-2 and SL-C of the MoMuLV packaging signal) with K_{d} ~100 nM (56, 57). The Rous sarcoma virus (RSV) packaging signal is a comparatively small four-way junction (82-nt) with three stem loops (SLA,
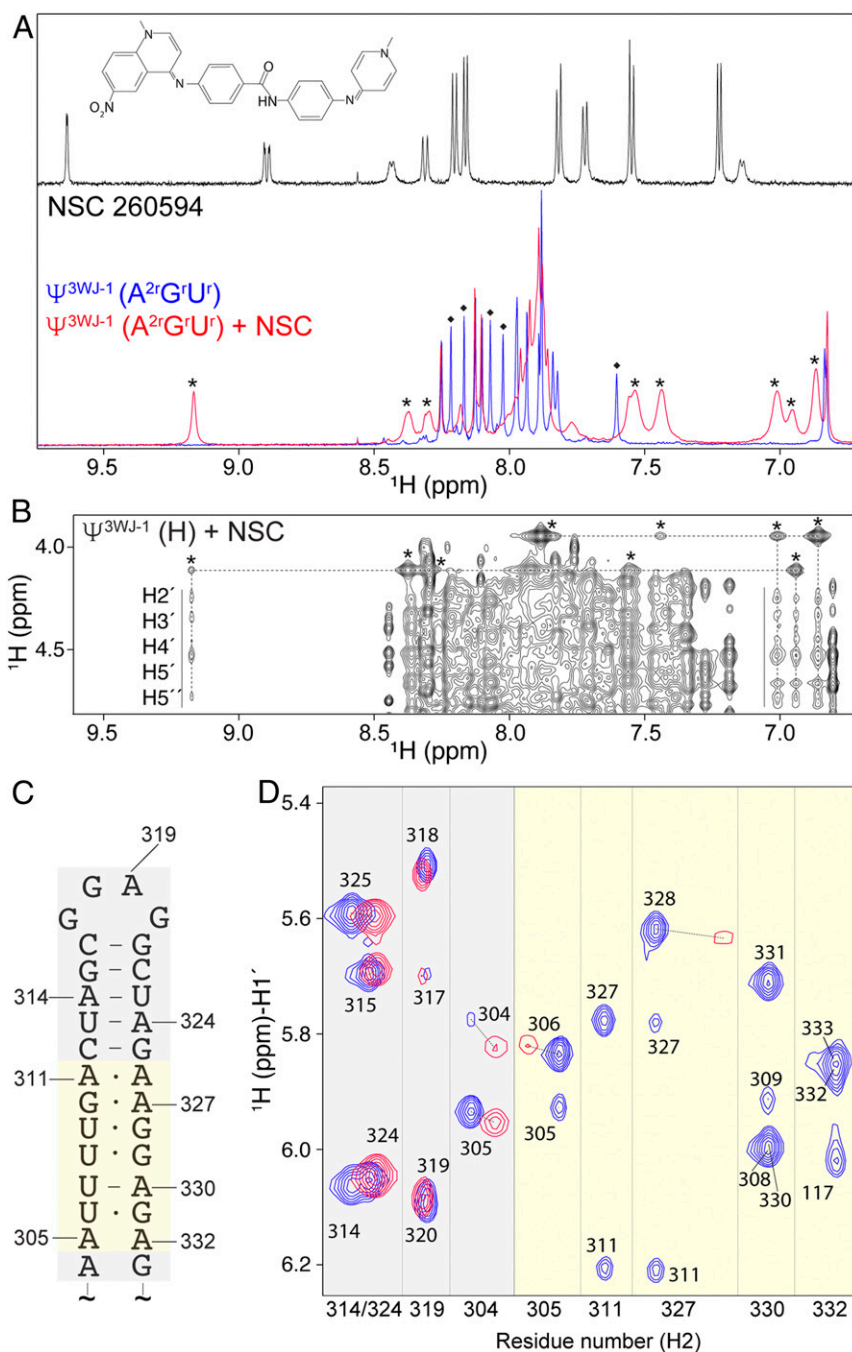
Fig. 5. NSC 260594 binds to the stem region of Ψ-stem loop. (*A*) One-dimensional ^{1}H spectrum for NSC (*Top*). Overly of 1D ^{1}H spectra of $A^{2r}G^rU^r$–$Ψ^{3WJ-1}$ with (red) or without (blue) twofold excess of NSC (*Bottom*). The ^{1}H signals of the NSC compound in the RNA-bound state are indicated by asterisks. The ^{1}H signals of the $Ψ^{3WJ-1}$ with severe line broadening after NSC binding are indicated by diamonds. (*B*) NOESY spectrum of fully protonated $Ψ^{3WJ-1}$ with twofold excess of NSC. Asterisks denote intramolecular NOEs between aromatic and methyl protons of NSC. Intermolecular NOEs between NSC aromatic protons and RNA ribose protons are marked by vertical lines. (*C*) Secondary structure of the Ψ-stem region. (*D*) Two-dimensional NOESY overlay of $A^{2r}G^rU^r$–$Ψ^{3WJ-1}$ with (red) or without (blue) twofold excess of NSC. Light yellow denotes the signals that are severely perturbed by NSC binding. Light gray shows the residues that are mildly affected by NSC binding.

SLB, and SLC) and the O3 stem. The high-affinity NC recognition element of RSV (K_d ~2 nM at low salt condition) is a two-component RNA structure, with the "UGCG" tetraloop of SLC bound by the N-terminal zinc finger of NC and the "AUG" residues linking SLA and SLB recognized by the C-terminal zinc finger (58, 59). The present studies show that structurally labile [UUUU]:[GGAG] elements serve as the initial sites for Gag binding to the HIV-1 leader. Thus, a general mechanism may be

employed by retroviruses in which genome packaging is initiated by the binding of a small number of Gag molecules to a cluster of exposed and/or labile high-affinity NC binding sites, which stimulates recruitment of additional Gag molecules and nucleates virus assembly.

Our findings also indicate that the previously characterized HIV-1 RNA packaging inhibitor NSC binds to the [UUUU]:[GGAG] element of the $Ψ^{3WJ-1}$ RNA. The fact that both NSC and NC bind

tightly to the apical loop of Ψ-hairpin constructs that lack the [UUUU]:[GGAG] element and shift their binding preferences to the [UUUU]:[GGAG] helix in larger RNA constructs (50) suggests that the "GGAG" sequence (also present in the apical loop of the Ψ-hairpin) is important for binding but insufficient for high-affinity binding. The potent inhibitory effect of NSC on RNA packaging (49), and the sensitivity of helix stabilizing (and NC-blocking) G:U to G–C mutations, are mutually consistent with the proposal that high-affinity NC binding to the [UUUU]:[GGAG] site is essential for genome packaging. Our findings are also consistent with chemical probing analyses, suggesting that the inhibitory function of NSC is due to its stabilization of the Ψ-hairpin region of the leader (49). These collective findings suggest that the HIV-1 packaging signal, and particularly the [UUUU]:[GGAG] helix that bridges the Ψ-hairpin and a three-way junction structure, could serve as a potent RNA target for HIV therapeutic development.

Materials and Methods

Preparation of DNA Templates. The DNA templates for the transcription of DIS and Ψ-hairpin derivative RNAs were prepared by annealing a 17-nt T7 promoter sequence (Top17, 5′-TAA TAC GAC TCA CTA TA-3′) to the corresponding reverse oligonucleotides (IDT): DIS, 5′-mGmGC CCC TCG CCT CTT GCC GTG CGC GCT TCA GCA AGC CGA GTC CTG CCT ATA GTG AGT CGT ATT A-3′; Ψ^{HP}, 5′-mCTC TCC TTC TAG CCT CCG CTA GTC AAA ATC TAT AGT GAG TCG TAT TA-3′; Ψ$^{HP-stable}$, 5′-mGmGC TCC TTC TAG CCT CCG CTA GTC AAA ACC TAT AGT GAG TCG TAT TA-3′; Ψ-apical loop, 5′-mGmGA CTA GCC TCC GCT AGT CCT ATA GTG AGT CGT ATT A-3′; Ψ-fragment, 5′-mCTC CTT CTA GCC TCC GCT AGT CTA TAG TGA GTC GTA TTA (m denotes 2′-O-methyl modification to reduce nontemplated nucleotide addition by T7 RNA polymerase) (61). Briefly, Top 17 (20 μL, 600 μM) was mixed with reverse DNA oligo (40 μL, 200 μM). The mixture was then boiled for 3 min and slow cooled overnight. H$_2$O (940 μL) was added to yield 1-mL DNA template, which was directly used in the transcription reaction.

The DNA templates for other larger RNA constructs were obtained by standard PCR amplification (EconoTaq PLUS 2× Master Mix; Lucigen) using pUC19 plasmids containing the Top17 sequence and the RNA encoding sequences. A common forward amplification primer 80 nt upstream to the T7 promoter sequence was used for all constructs (5′-GGG ATG TGC TGC AAG GCG ATT AAG TTG GG-3′). The reverse amplification primers are 5′-mCmGC ACC CAT CTC TCT CCT TCT AGC CTC C-3′ for 5′-L, 5′-L U308/309C, and 5′-L^{ΔPBS}; 5′-mGmGC ACC CAT CTC TCT CCT TCT AGC CTC C-3′ for Ψ^{CES}, Ψ^{T-3WJ}, Ψ^{3WJ-1}, and Ψ^{3WJ-1} U308/309C; 5′-mCmAC TAC TTT GAG CAC TCA AGG CAA GC-3′ for TAR-PolyA.

The construction of plasmids for 5′-L, 5′-L^{ΔPBS}, and Ψ^{CES} were described previously (30, 37). The DNA fragment for Ψ^{T-3WJ} was generated by overlapping PCR using the Ψ^{CES} plasmid as the template with the primers 5′-CCT AGT AAG GAA TTC TAA TAC GAC TCA CTA TAG GTG CCC GTC TG-3′ and 5′-GTA CTC ACC AGT CGT CTC CGA GAG ATC TCC TC-3′ for the first fragment, and primers 5′-GAG GAG ATC TCT CGG AGA CGA CTG GTG AGT AC-3′ and 5′-CTT ACT AGG GGA TCC CCC GGG CAC-3′ for the second fragment. The amplified full-length DNA fragment for Ψ^{T-3WJ} was ligated into the pUC19 plasmid after digestion with EcoRI and BamHI (New England Biolabs). The plasmid for Ψ^{3WJ-1} was generated using the same overlapping PCR method with 5′-CTG TTG TGT CGA GAG ACG CCA AAA ATT TTG AC-3′ and 5′-TTT TGG CGT CTC TCG ACA CAA CAG ACG-3′ as the two overlapping primers. Plasmid for 5′-L U308/309C was prepared by site-directed mutagenesis (QuikChange Lightning; Agilent) on 5′-L plasmid using forward primer 5′-GGT GAG TAC GCC AAA AAT TCC GAC TAG CGG AGG CTA GAA GG-3′ and reverse primer 5′-CCT TCT AGC CTC CGC TAG TCG GAA TTT TTG GCG TAC TCA CC-3′. Plasmid for Ψ^{3WJ-1} U308/309C was prepared similarly using Ψ^{3WJ-1} plasmid and forward primer 5′-GTC GAG AGA CGC CAA AAA TTC CGA CTA GCG GAG GCT AGA AGG-3′ and reverse primer 5′-CCT TCT AGC CTC CGC TAG TCG GAA TTT TTG GCG TCT CTC GAC-3′.

RNA In Vitro Transcription. RNAs were prepared by in vitro transcription using T7 RNA polymerase (purified in-house) in 7.5- to 30-mL reactions. A 7.5-mL reaction contained ∼0.25 mg of PCR-amplified DNA template or ∼4 nmol of annealed DNA template, 20 mM MgCl$_2$, 3 to 6 mM NTPs, 2 mM spermidine, 2 mM DTT, 20% (vol/vol) DMSO, 80 mM Tris·HCl (pH 9.0), and 0.15 mg of T7 RNA polymerase. The reaction was quenched after a 5-h incubation at 37 °C by addition of an EDTA-urea mixture (250 mM EDTA, 7 M urea, pH 8.0). The reaction mixture was boiled for 5 min and snap cooled on ice for 5 min prior

to addition of glycerol (final concentration, 6% [vol/vol]). RNAs were purified by electrophoresis on urea-containing polyacrylamide denaturing gels (SequaGel; National Diagnostics) at 20 W overnight, visualized by UV shadowing, and eluted using the Elutrap electroelution system (Whatman) at 150 V overnight. The eluted RNAs were concentrated and washed twice with 2 M high-purity NaCl followed by extensive desalting using Amicon Ultra Centrifugal Filter Device (Millipore).

Preparation of NTPs. Perdeuterated (A^D, G^D, C^D, and U^D) and partially deuterated (U^r) NTP reagents used for in vitro transcription were purchased from Cambridge Isotope Laboratories (CIL). A^{2r} and G^r were obtained by deuteration of the C8 position of fully protonated ATP and GTP. A volume of 140 μL of triethylamine (TEA) was added to 0.2 g of ATP or GTP dissolved in 8 mL of 99.8% D$_2$O (CIL). One-milliliter aliquots of ATP and GTP were incubated at 60 °C for 5 d and 24 h, respectively. TEA was subsequently removed by lyophilization.

NC Purification. The pET-3a plasmid containing the DNA fragment encoding HIV-1$_{NL4-3}$ NC protein (55 amino acids with the sequence of MQKGN FRNQR KTVKC FNCGK EGHIA KNCRA PRKKG CWKCG KEGHQ MKDCT ERQAN) was transformed into BL21 (DE3) pLysE. The protein was overexpressed and purified as described previously (37). Briefly, the cells were lysed by freeze-thawing and microfluidization. After centrifugation, the supernatant containing NC protein was applied to ion exchange chromatography using tandem Q- and SP- columns (the Q-column was removed before elution). Protein was further purified with size exclusion chromatography on a Superdex 30 column in PI buffer containing 5 mM TCEP.

ITC. ITC experiments were carried out using an ITC200 (MicroCal Corporation). A volume of 40 μL of NC (120 to 240 μM) in ITC buffer (10 mM Tris·HCl, pH 7.0, 140 mM KCl, 10 mM NaCl, 5 mM MgCl$_2$, and 5 mM TCEP) was loaded into the injection syringe. The calorimetry cell was loaded with 200 μL of RNA (0.6 to 4 μM) in the same buffer as NC. After thermal equilibration at 30 °C and the initial 60-s delay, a single injection of 0.2 μL followed by 19 serial injections of 2 μL were made into the calorimetry cell.

Native Gel Electrophoresis. RNA (1 μM) was prepared in PI buffer with 10 mM Tris·HCl (pH 7.5), 140 mM KCl, 10 mM NaCl, and 1 mM MgCl$_2$, heat denatured for 3 min, and slowly cooled down to room temperature. The 500-ng RNA samples were loaded onto 1% agarose gels prestained with ethidium bromide and run at 110 V in 1× TB buffer (44.5 mM Tris–boric acid, pH 7.5) on ice to prevent thermal denaturation during electrophoresis.

NMR Spectroscopy. All NMR data were collected with 600- or 800-MHz Bruker AVANCE spectrometer equipped with cryoprobe. NMR samples were prepared in 10 mM Tris-d11 buffer with an RNA concentration of ∼200 μM in a standard 5-mm NMR tube (500 μL). Nonexchangeable ^{1}H assignments were obtained from 2D NOESY experiments (NOE mixing time, 300 ms; relaxation delay, 3.2 s; T = 35 °C) using samples in D$_2$O (99.8%; CIL). Imino ^{1}H assignments were obtained from 2D NOESY experiments with jump-return water suppression (T = 10 °C) using samples in 90% H$_2$O/10% D$_2$O. The NSC compound (provided by National Cancer Institute–Developmental Therapeutics Program; https://dtp.cancer.gov/) used for NMR titration experiments was dissolved in DMSO-d6 (CIL) with a final concentration of 20 mM and stored at −20 °C. All NMR data were processed with NMRFx (62) and analyzed with NMRViewJ (63).

In Vivo Packaging Experiments. The Ψ$^+$ helper NL4-3 GPP contains HIV-1 NL4-3 sequences with *env* replaced by a puromycin-resistance expression cassette and has been described previously (29). The test vector L688-RRE-Puro-LTR contains 688 base pairs of the wild-type HIV-1 NL4-3 leader and has been described previously (37). U308/309C mutation was introduced into L688-RRE-Puro-LTR by replacement of the corresponding region in the test vector with a fragment generated using overlap extension PCR.

Human 293T cells were cultured at 37 °C under 5% CO$_2$ in DMEM media, containing 10% FBS and 50 μg/mL gentamicin. To generate virions for RNA packaging analysis, 293T cells at 70% confluence in 10-cm plates were cotransfected with a total of 6 μg of plasmid DNA that included one test vector plasmid and NL4-3 GPP at an ∼1:1 molar ratio using polyethylenimine (64). Cells and media were harvested 48 h after transfection. Media was filtered through 0.22-μm filters and viral particles concentrated by ultracentrifugation (25,000 rpm) through a 20% sucrose cushion. RNA from virus pellet and cells was isolated using TRIzol according to the manufacturer's protocol (Ambion) and DNase treated.

RNA content of transfected cells and viral particles was analyzed by RNase protection assay (RPA) as previously described (65). The chimeric riboprobe HIVgag/CMV used in this study was designed to protect a 201-nt fragment of HIV-1 gag sequence (presented only in NL4-3 GPP helper) and 289 nt of CMV promoter sequence (presented only in test vectors). Riboprobe was transcribed from linearized plasmid templates using T7 RNA polymerase (Promega) and [α-^{32}P]-rCTP (Perkin-Elmer). Products of RPA were quantified by phosphorimaging using standard methods.

Data Availability. All necessary data for substantiating the conclusions in the manuscript are included in the manuscript and *SI Appendix*. Additional data are available upon request from the authors.

1. S. Sun, V. B. Rao, M. G. Rossmann, Genome packaging in viruses. *Curr. Opin. Struct. Biol.* **20**, 114–120 (2010).
2. R. Berkowitz, J. Fisher, S. P. Goff, RNA packaging. *Curr. Top. Microbiol. Immunol.* **214**, 177–218 (1996).
3. J. M. Coffin, S. H. Hughes, H. E. Varmus, *Retroviruses*, (Cold Spring Harbor Laboratory Press, Plainview, NY, 1997).
4. J.-C. Paillart, M. Shehu-Xhilaga, R. Marquet, J. Mak, Dimerization of retroviral RNA genomes: An inseparable pair. *Nat. Rev. Microbiol.* **2**, 461–472 (2004).
5. S. F. Johnson, A. Telesnitsky, Retroviral RNA dimerization and packaging: The what, how, when, where, and why. *PLoS Pathog.* **6**, e1001007 (2010).
6. V. D'Souza, M. F. Summers, How retroviruses select their genomes. *Nat. Rev. Microbiol.* **3**, 643–655 (2005).
7. W.-S. Hu, H. M. Temin, Genetic consequences of packaging two RNA genomes in one retroviral particle: Pseudodiploidy and high rate of genetic recombination. *Proc. Natl. Acad. Sci. U.S.A.* **87**, 1556–1560 (1990).
8. W. S. Hu, H. M. Temin, Retroviral recombination and reverse transcription. *Science* **250**, 1227–1233 (1990).
9. S. R. King, N. K. Duggal, C. B. Ndongmo, C. Pacut, A. Telesnitsky, Pseudodiploid genome organization AIDS full-length human immunodeficiency virus type 1 DNA synthesis. *J. Virol.* **82**, 2376–2384 (2008).
10. T. Nora *et al.*, Contribution of recombination to the evolution of human immunodeficiency viruses expressing resistance to antiretroviral treatment. *J. Virol.* **81**, 7620–7628 (2007).
11. A. Onafuwa-Nuga, A. Telesnitsky, The remarkable frequency of human immunodeficiency virus type 1 genetic recombination. *Microbiol. Mol. Biol. Rev.* **73**, 451–480 (2009).
12. J. Chen *et al.*, High efficiency of HIV-1 genomic RNA packaging and heterozygote formation revealed by single virion analysis. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 13535–13540 (2009).
13. E. Mailler *et al.*, The life-cycle of the HIV-1 Gag-RNA complex. *Viruses* **8**, 248 (2016).
14. P. Bieniasz, A. Telesnitsky, Multiple, switchable protein:RNA interactions regulate human immunodeficiency virus type 1 assembly. *Annu. Rev. Virol.* **5**, 165–183 (2018).
15. E. D. Olson, K. Musier-Forsyth, Retroviral Gag protein-RNA interactions: Implications for specific genomic RNA packaging and virion assembly. *Semin. Cell Dev. Biol.* **86**, 129–139 (2019).
16. K. Lu, X. Heng, M. F. Summers, Structural determinants and mechanism of HIV-1 genome packaging. *J. Mol. Biol.* **410**, 609–633 (2011).
17. A. M. Lever, HIV-1 RNA packaging. *Adv. Pharmacol.* **55**, 1–32 (2007).
18. K. M. Weeks, C. Ampe, S. C. Schultz, T. A. Steitz, D. M. Crothers, Fragments of the HIV-1 Tat protein specifically bind TAR RNA. *Science* **249**, 1281–1285 (1990).
19. A. T. Das, B. Klaver, B. Berkhout, A hairpin structure in the R region of the human immunodeficiency virus type 1 RNA genome is instrumental in polyadenylation site selection. *J. Virol.* **73**, 81–91 (1999).
20. V. Goldschmidt, C. Ehresmann, B. Ehresmann, R. Marquet, Does the HIV-1 primer activation signal interact with tRNA3(Lys) during the initiation of reverse transcription? *Nucleic Acids Res.* **31**, 850–859 (2003).
21. E. Skripkin, J. C. Paillart, R. Marquet, B. Ehresmann, C. Ehresmann, Identification of the primary site of the human immunodeficiency virus type 1 RNA dimerization in vitro. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 4945–4949 (1994).
22. C. M. Stoltzfus, J. M. Madsen, Role of viral splicing elements and cellular RNA binding proteins in regulation of HIV-1 alternative RNA splicing. *Curr. HIV Res.* **4**, 43–55 (2006).
23. A. Lever, H. Göttlinger, W. Haseltine, J. Sodroski, Identification of a sequence required for efficient packaging of human immunodeficiency virus type 1 RNA into virions. *J. Virol.* **63**, 4085–4087 (1989).
24. A. Aldovini, R. A. Young, 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 (1990).
25. F. Clavel, J. M. Orenstein, A mutant of human immunodeficiency virus with reduced RNA packaging and abnormal particle morphology. *J. Virol.* **64**, 5230–5234 (1990).
26. M. S. McBride, A. T. Panganiban, The human immunodeficiency virus type 1 encapsidation site is a multipartite RNA element composed of functional hairpin structures. *J. Virol.* **70**, 2963–2973 (1996).
27. J. L. Clever, T. G. Parslow, Mutant human immunodeficiency virus type 1 genomes with defects in RNA dimerization or encapsidation. *J. Virol.* **71**, 3407–3414 (1997).
28. B. Berkhout, Structure and function of the human immunodeficiency virus leader RNA. *Prog. Nucleic Acid Res. Mol. Biol.* **54**, 1–34 (1996).
29. K. Lu *et al.*, NMR detection of structures in the HIV-1 5′-leader RNA that regulate genome packaging. *Science* **334**, 242–245 (2011).
30. S. C. Keane *et al.*, RNA structure. Structure of the HIV-1 RNA packaging signal. *Science* **348**, 917–921 (2015).
31. S. C. Keane *et al.*, NMR detection of intermolecular interaction sites in the dimeric 5′-leader of the HIV-1 genome. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 13033–13038 (2016).
32. J.-L. Darlix, C. Gabus, M.-T. Nugeyre, F. Clavel, F. Barré-Sinoussi, *Cis* elements and *trans*-acting factors involved in the RNA dimerization of the human immunodeficiency virus HIV-1. *J. Mol. Biol.* **216**, 689–699 (1990).
33. M. Ooms, H. Huthoff, R. Russell, C. Liang, B. Berkhout, A riboswitch regulates RNA dimerization and packaging in human immunodeficiency virus type 1 virions. *J. Virol.* **78**, 10814–10819 (2004).
34. T. Masuda *et al.*, Fate of HIV-1 cDNA intermediates during reverse transcription is dictated by transcription initiation site of virus genomic RNA. *Sci. Rep.* **5**, 17680 (2015).
35. S. Kharytonchyk *et al.*, Transcriptional start site heterogeneity modulates the structure and function of the HIV-1 genome. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 13378–13383 (2016).
36. J. D. Brown *et al.*, Structural basis for transcriptional start site control of HIV-1 RNA fate. *Science* **368**, 413–417 (2020).
37. X. Heng *et al.*, Identification of a minimal region of the HIV-1 5′-leader required for RNA dimerization, NC binding, and packaging. *J. Mol. Biol.* **417**, 224–239 (2012).
38. R. J. Gorelick *et al.*, Noninfectious human immunodeficiency virus type 1 mutants deficient in genomic RNA. *J. Virol.* **64**, 3207–3211 (1990).
39. R. N. De Guzman *et al.*, Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element. *Science* **279**, 384–388 (1998).
40. G. K. Amarasinghe *et al.*, NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the psi-RNA packaging signal. Implications for genome recognition. *J. Mol. Biol.* **301**, 491–511 (2000).
41. S. Spriggs, L. Garyu, R. Connor, M. F. Summers, Potential intra- and intermolecular interactions involving the unique-5′ region of the HIV-1 5′-UTR. *Biochemistry* **47**, 13064–13073 (2008).
42. W. I. Sundquist, H. G. Kräusslich, HIV-1 assembly, budding, and maturation. *Cold Spring Harb. Perspect. Med.* **2**, a006924 (2012).
43. N. Jouvenet, S. M. Simon, P. D. Bieniasz, Visualizing HIV-1 assembly. *J. Mol. Biol.* **410**, 501–511 (2011).
44. N. Jouvenet, S. M. Simon, P. D. Bieniasz, Imaging the interaction of HIV-1 genomes and Gag during assembly of individual viral particles. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 19114–19119 (2009).
45. I. Kucharska *et al.*, Biochemical reconstitution of HIV-1 assembly and maturation. *J. Virol.* **94**, e01844-19 (2020).
46. K. Zhang *et al.*, Structure of the 30 kDa HIV-1 RNA dimerization signal by a hybrid Cryo-EM, NMR, and molecular dynamics approach. *Structure* **26**, 490–498.e3 (2018).
47. K. Wüthrich, *NMR of Proteins and Nucleic Acids*, (Wiley, New York, 1986).
48. A. Zeffman, S. Hassard, G. Varani, A. Lever, The major HIV-1 packaging signal is an extended bulged stem loop whose structure is altered on interaction with the Gag polyprotein. *J. Mol. Biol.* **297**, 877–893 (2000).
49. C. K. Ingemarsdotter, J. Zeng, Z. Long, A. M. L. Lever, J. C. Kenyon, An RNA-binding compound that stabilizes the HIV-1 gRNA packaging signal structure and specifically blocks HIV-1 RNA encapsidation. *Retrovirology* **15**, 25 (2018).
50. N. M. Bell *et al.*, Targeting RNA-protein interactions within the human immunodeficiency virus type 1 lifecycle. *Biochemistry* **52**, 9269–9274 (2013).
51. G. P. Harrison, G. Miele, E. Hunter, A. M. L. Lever, Functional analysis of the core human immunodeficiency virus type 1 packaging signal in a permissive cell line. *J. Virol.* **72**, 5886–5896 (1998).
52. R. S. Russell *et al.*, Sequences downstream of the 5′ splice donor site are required for both packaging and dimerization of human immunodeficiency virus type 1 RNA. *J. Virol.* **77**, 84–96 (2003).
53. G. Singh *et al.*, Identification of conserved, primary sequence motifs that direct retrovirus RNA fate. *Nucleic Acids Res.* **46**, 7366–7378 (2018).
54. J. H. Richardson, L. A. Child, A. M. L. Lever, Packaging of human immunodeficiency virus type 1 RNA requires *cis*-acting sequences outside the 5′ leader region. *J. Virol.* **67**, 3997–4005 (1993).
55. E. W. Abd El-Wahab *et al.*, Specific recognition of the HIV-1 genomic RNA by the Gag precursor. *Nat. Commun.* **5**, 4304 (2014).
56. V. D'Souza *et al.*, Identification of a high affinity nucleocapsid protein binding element within the Moloney murine leukemia virus Psi-RNA packaging signal: Implications for genome recognition. *J. Mol. Biol.* **314**, 217–232 (2001).

BIOCHEMISTRY

57. V. D'Souza, M. F. Summers, Structural basis for packaging the dimeric genome of Moloney murine leukaemia virus. *Nature* **431**, 586–590 (2004).

58. J. Zhou, J. K. McAllen, Y. Tailor, M. F. Summers, High affinity nucleocapsid protein binding to the muPsi RNA packaging signal of Rous sarcoma virus. *J. Mol. Biol.* **349**, 976–988 (2005).

59. J. Zhou, R. L. Bean, V. M. Vogt, M. Summers, Solution structure of the Rous sarcoma virus nucleocapsid protein: muPsi RNA packaging signal complex. *J. Mol. Biol.* **365**, 453–467 (2007).

60. F. Z. Yildiz, K. Babalola, M. F. Summers, Identification of a high affinity nucleocapsid protein binding element from the bovine leukemia virus genome. *Virus Res.* **171**, 278–286 (2013).

61. C. Kao, M. Zheng, S. Rüdisser, A simple and efficient method to reduce nontemplated nucleotide addition at the 3 terminus of RNAs transcribed by T7 RNA polymerase. *RNA* **5**, 1268–1272 (1999).

62. M. Norris, B. Fetler, J. Marchant, B. A. Johnson, NMRFx processor: A cross-platform NMR data processing program. *J. Biomol. NMR* **65**, 205–216 (2016).

63. B. A. Johnson, Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol. Biol.* **278**, 313–352 (2004).

64. O. Boussif *et al*., A versatile vector for gene and oligonucleotide transfer into cells in culture and in vivo: Polyethylenimine. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 7297–7301 (1995).

65. A. A. Onafuwa-Nuga, S. R. King, A. Telesnitsky, Nonrandom packaging of host RNAs in Moloney murine leukemia virus. *J. Virol.* **79**, 13528–13537 (2005).