

This work is on a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) license, <https://creativecommons.org/licenses/by-nc-nd/4.0/>. Access to this work was provided by the University of Maryland, Baltimore County (UMBC) ScholarWorks@UMBC digital repository on the Maryland Shared Open Access (MD-SOAR) platform.

Please provide feedback

Please support the ScholarWorks@UMBC repository by emailing [scholarworks-group@umbc.edu](mailto:scholarworks-group@umbc.edu) and telling us what having access to this work means to you and why it's important to you. Thank you.

# 5'-Cap sequestration is an essential determinant of HIV-1 genome packaging

Pengfei Ding<sup>a,b,1</sup> , Siarhei Kharytonchyk<sup>c,1</sup>, Nansen Kuo<sup>b</sup>, Emily Cannistraci<sup>b</sup>, Hana Flores<sup>b</sup> , Ridhi Chaudhary<sup>b</sup> , Mitali Sarkar<sup>b</sup>, Xinmei Dong<sup>b</sup>, Alice Telesnitsky<sup>c,2</sup> , and Michael F. Summers<sup>a,b,2</sup> 

<sup>a</sup>HHMI, University of Maryland, Baltimore County, Baltimore, MD 21250; <sup>b</sup>Department of Chemistry and Biochemistry, University of Maryland, Baltimore County, Baltimore, MD 21250; and <sup>c</sup>Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, MI 48109-5620

Contributed by Michael F. Summers, August 3, 2021 (sent for review July 6, 2021; reviewed by Reuben S. Harris and Judith G. Levin)

**HIV-1 selectively packages two copies of its 5'-capped RNA genome (gRNA) during virus assembly, a process mediated by the nucleocapsid (NC) domain of the viral Gag polypeptide and encapsidation signals located within the dimeric 5' leader of the viral RNA. Although residues within the leader that promote packaging have been identified, the determinants of authentic packaging fidelity and efficiency remain unknown. Here, we show that a previously characterized 159-nt region of the leader that possesses all elements required for RNA dimerization, high-affinity NC binding, and packaging in a noncompetitive RNA packaging assay ( $\Psi^{\text{CES}}$ ) is unexpectedly poorly packaged when assayed in competition with the intact 5' leader.  $\Psi^{\text{CES}}$  lacks a 5'-tandem hairpin element that sequesters the 5' cap, suggesting that cap sequestration may be important for packaging. Consistent with this hypothesis, mutations within the intact leader that expose the cap without disrupting RNA structure or NC binding abrogated RNA packaging, and genetic addition of a 5' ribozyme to  $\Psi^{\text{CES}}$  to enable cotranscriptional shedding of the 5' cap promoted  $\Psi^{\text{CES}}$ -mediated RNA packaging to wild-type levels. Additional mutations that either block dimerization or eliminate subsets of NC binding sites substantially attenuated competitive packaging. Our studies indicate that packaging is achieved by a bipartite mechanism that requires both sequestration of the 5' cap and exposure of NC binding sites that reside fully within the  $\Psi^{\text{CES}}$  region of the dimeric leader. We speculate that cap sequestration prevents irreversible capture by the cellular RNA processing and translation machinery, a mechanism likely employed by other viruses that package 5'-capped RNA genomes.**

HIV-1 | RNA | genome | 5' cap | packaging

**H**IV-1 replication is critically dependent on selective and efficient incorporation of two copies of the unspliced, 5'-capped RNA genome (gRNA) into assembling virus particles (1–5). Genomes are recruited to plasma membrane assembly sites from a cytoplasmic milieu that contains a substantial excess of cellular RNAs, spliced viral messenger RNAs (mRNAs) that encode for viral accessory proteins, and unspliced viral mRNAs that encode the viral Gag and Gag-Pol proteins (6). gRNA selection is mediated by the nucleocapsid (NC) domain of Gag, which interacts with *cis*-acting RNA “packaging signals” located within the 5' leader (5'-L) of the unspliced RNA (7–12). Genomes are initially anchored to the plasma membrane by a small number of Gag proteins, possibly two dozen or fewer (13, 14), forming a ribonucleoprotein nucleation site that recruits additional Gag proteins and cellular factors required for virus assembly and budding (15). Genomes are packaged as dimers, and there is considerable evidence that dimerization and packaging are mechanistically coupled (4, 5, 7, 16–18). Although gRNA versus mRNA functions were originally thought to be controlled by dimerization-dependent structural modulation of a single viral transcript (19, 20), it now appears that RNA fates are established by a heterogeneous transcriptional start site mechanism, in which two functionally distinct pools of RNAs containing one or three 5' guanosines are transcribed from the single integrated provirus (21, 22). RNAs containing a single 5' guanosine preferentially form dimers *in vitro* and

are selected for packaging as gRNA whereas those with three (and to a lesser extent, two) 5' guanosines adopt a monomeric leader structure *in vitro* and are retained in cells as mRNAs (22, 23).

Efforts to identify the specific RNA residues and structures that promote packaging have relied primarily on biophysical studies of recombinant RNAs and *in situ* packaging experiments involving transiently transfected or infected cells (8, 24–27). The dimeric gRNA leader that promotes packaging adopts a branched multihairpin structure, in which residues important for transcriptional activation, initiation of reverse transcription, dimerization, and packaging (TAR, PBS, DIS, and  $\Psi^{\text{HP}}$ , respectively), or that contain a polyadenylation sequence (polyA), adopt hairpin structures that decorate a central tandem three-way junction structure (Fig. 1A) (8, 17, 24, 28, 29). The structure exposes approximately two dozen high-affinity NC binding sites under physiological-like conditions (27). A variant of the 5'-L that lacks the TAR and polyA hairpins and contains a GAGA tetraloop (A, adenosine; G, guanosine) substituted for the PBS loop exhibits dimerization and NC-binding properties similar to those of the intact 5'-L (24, 27), and vectors containing 5'-Ls that lack these elements can be readily packaged into virus-like particles (24). This region of the leader has therefore been called the “core encapsidation signal” ( $\Psi^{\text{CES}}$ ) (Fig. 1A and C) (24, 27, 29).

## Significance

**HIV-1 replication is critically dependent on the selective incorporation of viral RNA genomes into assembling virions. Although RNA elements that promote packaging have been identified, the determinants of authentic packaging fidelity and efficiency have until now remained unknown. The present studies show that genome selection is achieved by a bipartite mechanism that requires both dimerization-dependent exposure of packaging signals within the 5' leader of the viral RNA and structural sequestration of the 5' cap. Cap sequestration likely prevents capture by the cellular RNA processing and translation machinery, a mechanism that may help explain why some cellular RNA polymerase III transcripts that lack 5' caps can parasitize virions. Cap sequestration may be commonly employed among RNA viruses that package 5'-capped genomes.**

Author contributions: P.D., A.T., and M.F.S. designed research; P.D., S.K., N.K., E.C., H.F., R.C., M.S., and X.D. performed research; P.D., S.K., N.K., E.C., H.F., R.C., M.S., X.D., A.T., and M.F.S. analyzed data; and P.D., A.T., and M.F.S. wrote the paper.

Reviewers: R.S.H., University of Minnesota; and J.G.L., NIH.

The authors declare no competing interest.

This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

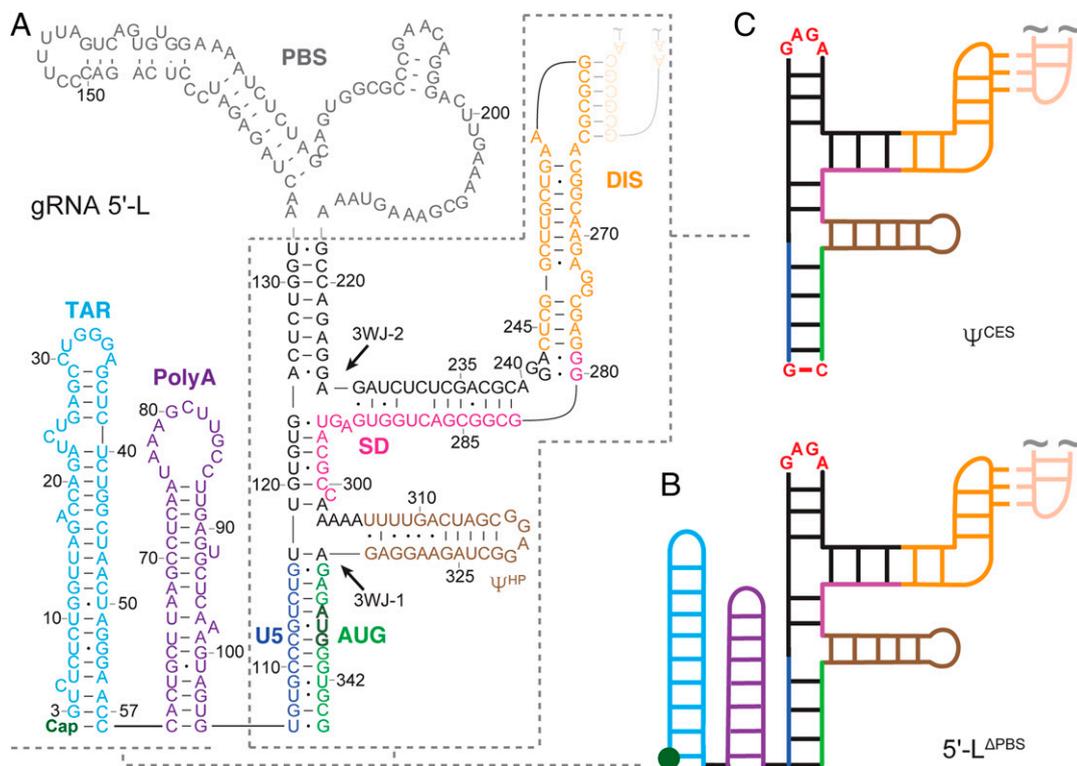
See [online](#) for related content such as Commentaries.

<sup>1</sup>P.D. and S.K. contributed equally to this work.

<sup>2</sup>To whom correspondence may be addressed. Email: [ateles@umich.edu](mailto:ateles@umich.edu) or [summers@umbc.edu](mailto:summers@umbc.edu).

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2112475118/-DCSupplemental>.

Published September 7, 2021.



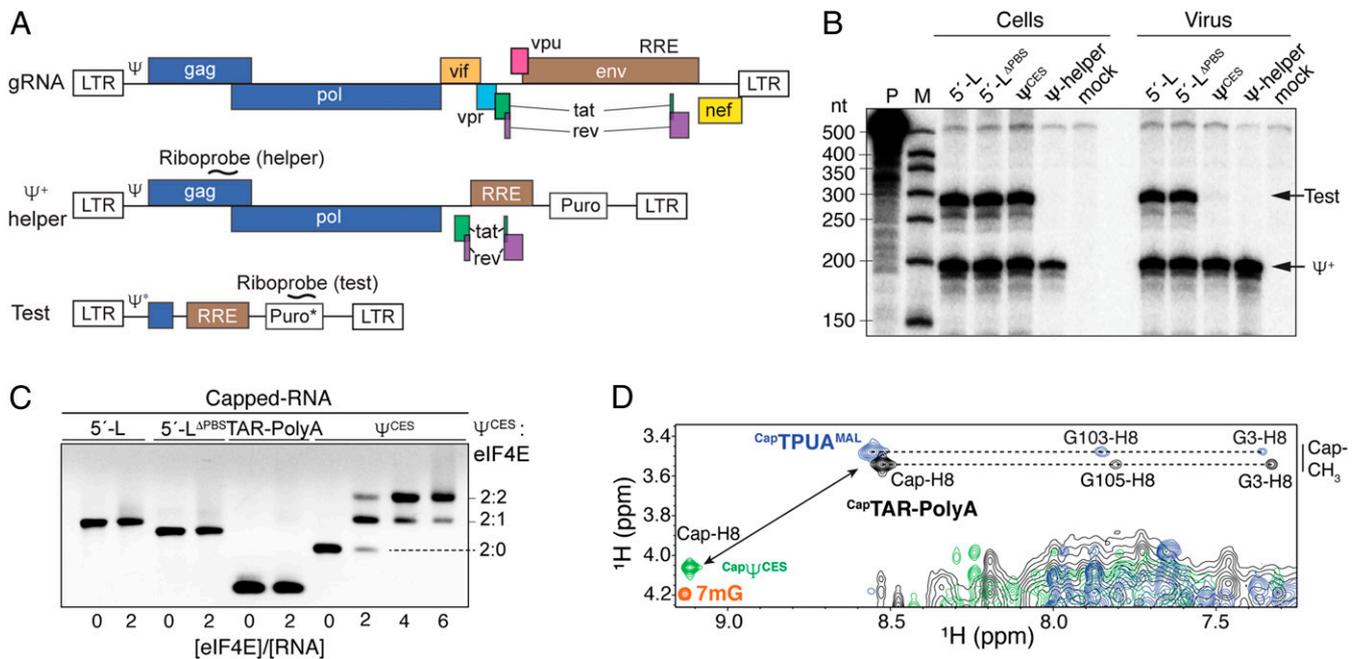
**Fig. 1.** Secondary structure of the HIV-1<sub>NL4-3</sub> gRNA 5'-L and constructs used to probe RNA packaging. (A) Native 5'-L. The PBS region is shaded in gray; nucleotides of TAR, polyA, U5, DIS, SD (splice donor site),  $\Psi^{\text{HP}}$ , and AUG are color-coded. The two three-way junctions are labeled as  $\Psi^{\text{3WJ-1}}$  and  $\Psi^{\text{3WJ-2}}$ . (B) Schematic representation of the 5'-L $\Delta^{\text{PBS}}$  RNA (PBS substituted by a GAGA tetraloop). (C) Schematic representation of  $\Psi^{\text{CES}}$  (5'-L $\Delta^{\text{TAR-}\Delta^{\text{PolyA-}\Delta^{\text{PBS}}}$ ). Native residues U106 and G345 are substituted by G and C, respectively.

Packaging efficiencies have generally been assessed by quantifying extracellular viral RNA levels (virion-associated RNA) relative to extracellular Gag protein levels or copackaged host RNA levels (24, 30–32) and, in prior studies, the packaging efficiency of  $\Psi^{\text{CES}}$  was measured indirectly by comparisons with the noncoding host 7SL RNA that is enriched in virions (24). However, in the absence of authentic genomes, Gag proteins can assemble into virus-like particles that incorporate nongenomic mRNAs, many in proportions that roughly match their intracellular concentrations (33–36), and this promiscuity can confound assessment of packaging efficiencies relative to that of the authentic gRNA. To better understand the contributions of specific leader elements to packaging, we have now employed a competitive RNA packaging approach that allows quantitative comparison of packaging efficiencies from cells cotransfected with vectors containing native and modified 5'-Ls. Unexpectedly, the ability of  $\Psi^{\text{CES}}$  to direct packaging is severely attenuated compared with that of the native leader, despite having similar structural, dimerization, and NC-binding properties. Our studies further show that the PBS loop does not contribute to packaging but that the TAR and polyA helices are indispensable for competitive packaging even though these elements do not contribute to high-affinity NC binding or dimerization. Recent NMR studies revealed that the TAR and polyA helices adopt an end-to-end stacking arrangement in which the 5' cap (5'-5'-triphosphate-linked 7-methylguanosine) is base-paired with, and sandwiched between, the TAR and polyA helices (23). We now show that cap sequestration by the TAR–polyA cassette is an essential determinant of packaging and that all other elements required for authentic packaging selectivity and efficiency reside fully within the  $\Psi^{\text{CES}}$  region of the leader. Mechanistic implications of cap sequestration for genome packaging by HIV-1 and other viruses that encapsidate 5'-capped RNAs are discussed.

## Results

**RNA Elements Required for Competitive Packaging.** To identify leader elements that contribute to wild-type packaging efficiency, competitive packaging experiments were performed using human 293T cells cotransfected with plasmid DNAs encoding native ( $\Psi^+$ ) and modified (test) HIV-1 5'-L sequences (27) (Fig. 2A). The  $\Psi^+$  vector contains the HIV-1<sub>NL4-3</sub> genome with deletions of the *vpu*, *vif*, and part of *env* genes and with *nef* replaced by a puromycin-resistance expression cassette (20). Test plasmids contain native or modified forms of the initial 688 nt of the HIV-1<sub>NL4-3</sub> genome followed by the Rev responsive element (RRE) and a puromycin-resistance gene (24) (Fig. 2A). Only cells expressing similar levels of  $\Psi^+$  and test RNAs, as assessed by an RNase protection assay (RPA) (37), were used for comparative packaging analyses. Experiments were conducted with test constructs containing the native 5'-L as a control, and with leaders that lack the primer binding site loop (5'-L $\Delta^{\text{PBS}}$ ) (Fig. 1B) or the PBS, TAR, and polyA hairpins (5'-L $\Delta^{\text{TAR-}\Delta^{\text{PolyA-}\Delta^{\text{PBS}}}$ ;  $\Psi^{\text{CES}}$ ) (Fig. 1C). Packaging data are reported throughout as the mean  $\pm$  SD from three individual packaging experiments.

Test RNAs containing the authentic 5'-L were packaged as efficiently as  $\Psi^+$ , indicating that differences in residues downstream of the 5'-L do not affect packaging (Fig. 2B) (38). Test RNAs containing the 5'-L $\Delta^{\text{PBS}}$  were also efficiently packaged relative to  $\Psi^+$  ( $104 \pm 11\%$ ) (Fig. 2B), as expected since 5'-L $\Delta^{\text{PBS}}$  exhibits dimerization and NC-binding properties similar to those of the intact leader (27). However, test RNAs containing the  $\Psi^{\text{CES}}$  leader were poorly packaged relative to  $\Psi^+$  (Fig. 2B) even though  $\Psi^{\text{CES}}$  efficiently promotes packaging in noncompetitive packaging assays (24).  $\Psi^{\text{CES}}$  differs from 5'-L $\Delta^{\text{PBS}}$  only by the absence of the TAR and polyA hairpins (Fig. 1), revealing that this 5'-tandem hairpin element is essential for competitive packaging. Neither the TAR nor polyA hairpins exhibit significant affinity for the



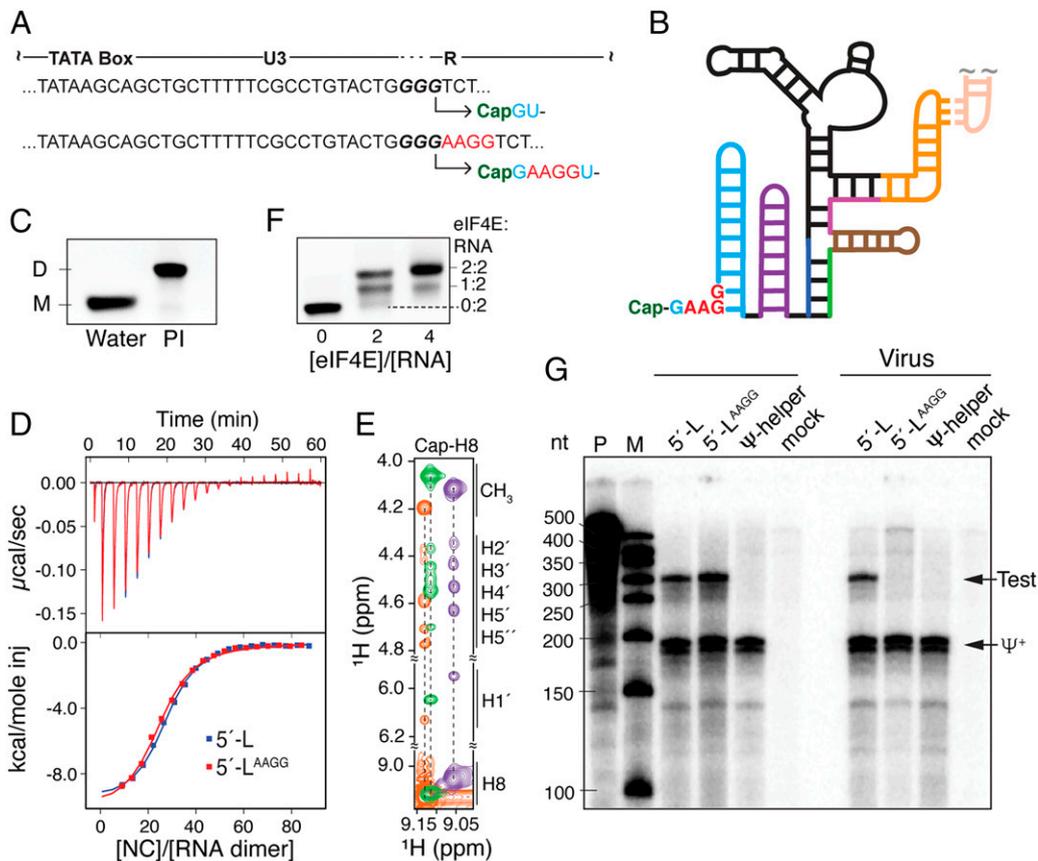
**Fig. 2.** RNA packaging signal in the HIV-1 5'-L. (*A, Top*) The proviral DNA of HIV-1<sub>NL4-3</sub>. (*A, Middle*) The  $\Psi^+$ -helper vector is a replication-defective version of the HIV-1<sub>NL4-3</sub> proviral DNA. (*A, Bottom*) The test vector's 5'-L includes the 5' untranslated region plus 353 nt from the gag-coding region. In the test vectors, the puromycin cassette (Puro\*) has an extra CMV promoter region, for which the riboprobe is targeted. LTR, long terminal repeat. (*B*) Packaging of test RNAs containing 5'-L, 5'-L<sup>APBS</sup>, and  $\Psi^{CES}$  in competition with  $\Psi^+$ , assayed by ribonuclease protection. M, RNA size marker; P, undigested probe. (*C*) Native gel electrophoresis showing that eIF4E binds the dimeric  $\text{Cap}\Psi^{CES}$  but not capped RNAs containing TAR-polyA ( $\text{Cap}5'-\text{L}$ ,  $\text{Cap}5'-\text{L}^{\text{APBS}}$ , and  $\text{CapTAR-polyA}$ ). (*D*) Portion of 2D NOESY spectra overlay showing cap chemical shifts and NOEs indicative of a sequestered 5' cap in  $\text{CapTAR-polyA}$  (black) and exposed 5' cap in  $\text{Cap}\Psi^{CES}$  (green); ppm, parts per million. Corresponding spectra for 7mG (orange) and  $\text{CapTPUA}^{\text{MAL}}$  [TAR, polyA, and U5:AUG regions of the HIV-1<sub>MAL</sub> leader (23)] (blue) are also shown. The chemical shifts and NOEs observed for the cap-CH<sub>3</sub> and -H8 protons of  $\text{CapTPUA}^{\text{MAL}}$  were similar to those of the HIV-1<sub>MAL</sub> 5'-L (23). Residue G105 of HIV-1<sub>NL4-3</sub> gRNA corresponds to G103 in the HIV-1<sub>MAL</sub> leader (*SI Appendix, Fig. S1*).

NC in vitro (27, 39) and removal of these two hairpins does not affect 5'-L dimerization (24), suggesting that factors other than NC binding or dimerization are responsible for TAR-polyA-dependent packaging. Recent NMR studies revealed that the TAR and polyA helices of the MAL strain of the HIV-1 (subtype A; HIV-1<sub>MAL</sub>) gRNA leader are coaxially stacked in a manner that structurally sequesters the 5' cap and prevents binding by eukaryotic translation initiation factor 4E (eIF4E) in vitro (23). Cap sequestration has been proposed to inhibit cap-dependent degradation of the gRNA and is likely to interfere with translation and splicing (23), but has not been previously considered important for packaging. Experiments were therefore conducted to test this possibility.

**State of the 5' Cap in Native and Mutant HIV-1<sub>NL4-3</sub> Leader RNAs.** NMR and cap accessibility titration experiments were performed to determine if native and mutant HIV-1<sub>NL4-3</sub> gRNA leader constructs adopt a cap-sequestered structure similar to that observed for the HIV-1<sub>MAL</sub> leader (23). Titrations of HIV-1<sub>NL4-3</sub> 5'-L and 5'-L<sup>APBS</sup>-capped RNAs with recombinant cap-binding protein eIF4E did not give rise to detectable band shifts in native electrophoretic mobility-shift assays (Fig. 2C), indicating that these RNAs do not bind eIF4E. An RNA construct comprising only the capped TAR and polyA hairpins ( $\text{CapTAR-polyA}$ ) was also unable to bind eIF4E (Fig. 2C), indicating that the  $\text{CapTAR-polyA}$  tandem hairpin functions as an independent cassette for cap sequestration. Two-dimensional (2D) nuclear Overhauser effect spectroscopy (NOESY) data obtained for  $\text{CapTAR-polyA}$  exhibited NMR chemical shifts and NOE cross-peaks similar to those observed for the HIV-1<sub>MAL</sub> 5'-L (Fig. 2D) (23), indicating that the 5'-capped HIV-1<sub>NL4-3</sub> gRNA leader also adopts a cap-sequestered structure with end-to-end stacking of the TAR (G3) and polyA (G105) hairpins. In contrast,  $^1\text{H}$  NMR chemical shifts

and NOEs observed for the cap-CH<sub>3</sub> and -H8 protons of 5'-capped  $\Psi^{CES}$  ( $\text{Cap}\Psi^{CES}$ ) were similar to those of 7-methylguanosine triphosphate (7mG; Fig. 2D), indicating that the 5' cap of  $\text{Cap}\Psi^{CES}$  exists in an exposed conformation. eIF4E titration experiments confirmed that the dimeric  $\text{Cap}\Psi^{CES}$  RNA readily binds to eIF4E molecules (Fig. 2C). These findings suggest a possible link between 5'-cap exposure and the inability of  $\text{Cap}\Psi^{CES}$  to promote competitive packaging (Fig. 2), and also that the TAR and polyA helices of the native leader may promote packaging by sequestering the 5' cap.

**Cap Exposure in the Full-Length Leader Abrogates RNA Packaging.** To test the hypothesis that cap exposure inhibits RNA packaging, we prepared a test vector with a stretch of AAGG residues inserted immediately downstream of the transcriptional start site (Fig. 3A). This insertion affords RNA transcripts with a leader sequence containing a nonnative A(1\*)-A(2\*)-G(3\*)-G(4\*) segment inserted immediately after the 5'-capped G3 residue ( $\text{Cap}5'-\text{L}^{\text{AAGG}}$ ) (Fig. 3A). The nonnative residues were engineered to substitute the native Cap:C58 and G3:C57 base pairs by G(3\*):C58 and G(4\*):C57 pairs and extend the cap away from the TAR hairpin (Fig. 3B). Under physiological-like ionic strength conditions (PI buffer: 10 mM Tris-HCl, 140 mM KCl, 10 mM NaCl, and 1 mM MgCl<sub>2</sub>, pH 7.5),  $\text{Cap}5'-\text{L}^{\text{AAGG}}$  forms dimers and retains NC-binding properties of the native leader (Fig. 3C and D). However, the 5' cap of  $\text{Cap}5'-\text{L}^{\text{AAGG}}$  exhibits NMR chemical shifts and NOEs consistent with a nonstacked and disordered conformation (Fig. 3E). Titration experiments confirmed that the 5' cap of the dimeric  $\text{Cap}5'-\text{L}^{\text{AAGG}}$  is accessible for eIF4E binding (Fig. 3F). Test RNAs containing the 5'-AAGG-modified leader were unable to detectably promote packaging when competing with  $\Psi^+$  (Fig. 3G). These findings collectively indicate that cap exposure is detrimental to



**Fig. 3.** Cap sequestration is required for competitive gRNA packaging. (A) The AAGG sequence was inserted after the transcription start site. (B) The nonnative AAGG residues extend the cap away from the TAR hairpin. (C and D) The inserted AAGG residues do not affect dimerization (C) and NC binding (D) (D, dimer; M, monomer). (E) Portion of 2D NOESY spectra overlay showing cap chemical shifts and NOEs indicative of the exposed 5' cap for  $^{cap}\Psi^{CES}$  (green) and  $^{cap}5'-L^{AAGG}$  (purple). NOEs for 7mG are shown in orange. ppm, parts per million. (F) eIF4E binds the  $^{cap}5'-L^{AAGG}$  dimer. (G) Packaging of test leader RNA with an AAGG insertion under competitive conditions as measured by ribonuclease protection.

packaging, even though the RNA exhibits dimerization and NC-binding properties similar to those of the native leader.

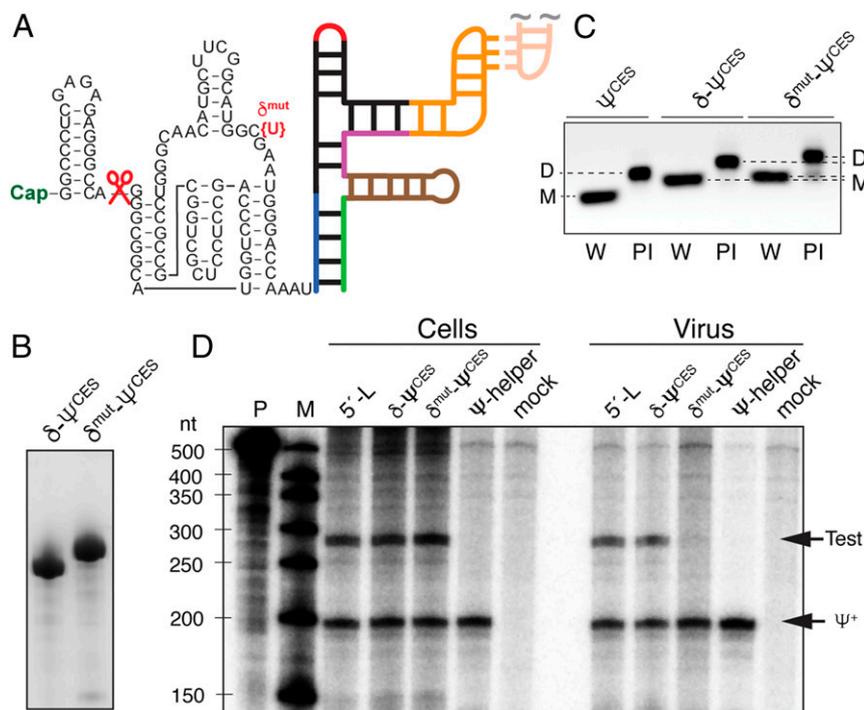
**Cotranscriptional Shedding of the 5' Cap Enables Authentic Packaging by  $\Psi^{CES}$ .** To determine if cap exposure accounts for the poor competitive packaging of  $\Psi^{CES}$  test RNAs, we generated chimeric RNA constructs in which active and inactive forms of the self-cleaving hepatitis delta virus (HDV) ribozyme were appended to the 5' end of  $\Psi^{CES}$  ( $\delta-\Psi^{CES}$  and  $\delta^{mut}-\Psi^{CES}$ , respectively) (Fig. 4A). The HDV element undergoes rapid self-cleavage to liberate its capped 5' hairpin both in vitro (Fig. 4B) and within the nucleus in vivo (40), whereas the HDV ribozyme with a single cytidine-to-uridine substitution remains intact. Both RNAs with 5' HDV sequences exhibit in vitro dimerization properties similar to those of the parent  $\Psi^{CES}$  RNA (Fig. 4C). However, whereas test RNAs containing the cap-shedding  $\delta-\Psi^{CES}$  leader exhibited packaging efficiencies similar to those of  $\Psi^{+}$  ( $106 \pm 11\%$ ),  $\delta^{mut}-\Psi^{CES}$  RNAs that retain the 5' cap were poorly packaged ( $9 \pm 6\%$  compared with the authentic 5'-L) (Fig. 4D). These findings further confirm that cap sequestration is required for competitive packaging and indicate that all other elements required for native-like RNA encapsidation reside within the  $\Psi^{CES}$  region of the leader.

**$\Psi^{CES}$  Is the Minimal Gag Recognition Signal.** We also asked if efficient packaging could be directed by cap-sequestered RNAs containing smaller regions of  $\Psi^{CES}$ . The highest-affinity NC binding sites ( $K_d \sim 30$  nM) known to be critical for packaging are located within the base of the  $\Psi$ -hairpin ( $\Psi^{HP}$ ), where it joins the proximal

three-way junction (3WJ-1; Fig. 1A) (27). Test RNAs containing the cap-sequestering TAR-polyA cassette and either the proximal three-way junction or an extended  $\Psi$ -hairpin (TP- $\Psi^{3WJ-1}$  and TP- $\Psi^{HP}$ , respectively; Fig. 5A) were not detectably packaged (Fig. 5B). In addition, test RNAs containing GAGA tetraloops substituted for the apical loop of the DIS hairpin, which prevents dimerization without altering the structure or NC-binding properties of the RNA (29), were also poorly packaged ( $36 \pm 6\%$  and  $38 \pm 8\%$  for 5'-L<sup>DISm</sup> and 5'-L<sup>APBS-DISm</sup>, respectively; Fig. 5B-E). These findings indicate that authentic packaging efficiency requires the entire dimeric  $\Psi^{CES}$  element.

## Discussion

Understanding the molecular mechanism of HIV-1 genome packaging has been challenging for several reasons. Residues important for packaging sometimes have multiple functions that can complicate interpretation of mutagenesis outcomes. In addition, Gag can efficiently assemble into RNA-containing virus-like particles from cells lacking viral gRNAs, and mutant RNAs with packaging defects can therefore be packaged at substantial levels if overexpressed in transfected cells in the absence of a competing authentic genome. As a result, nearly every portion of the HIV-1 5'-L, and even stretches of the gag-coding region (31, 41, 42), have been controversially proposed to play roles in genome packaging (38, 43). Prior studies of the TAR element illustrate these challenges. Early studies showed that base pairing near the 5' end of the TAR hairpin is required for optimal viral replication (44), and studies implicating TAR in translation, dimerization, packaging, and



**Fig. 4.** Cap-dependent modulation of  $\Psi^{\text{CES}}$  packaging. (A) Construct design of  $\delta\text{-}\Psi^{\text{CES}}$  and  $\delta^{\text{mut}}\text{-}\Psi^{\text{CES}}$ . (B) The wild-type HDV ribozyme cotranscriptionally cleaves the capped 5' fragment in vitro but the C-to-U mutant does not. (C)  $\delta\text{-}\Psi^{\text{CES}}$  and  $\delta^{\text{mut}}\text{-}\Psi^{\text{CES}}$  exhibit the same dimerization properties as the parent  $\Psi^{\text{CES}}$  RNA. M, monomer; D, dimer; W, water; PI, PI buffer. (D) Packaging of test RNAs containing  $\delta\text{-}\Psi^{\text{CES}}$  and  $\delta^{\text{mut}}\text{-}\Psi^{\text{CES}}$  in competition with  $\Psi^+$  as measured by ribonuclease protection.

reverse transcription were complicated by the dominant-negative effects of TAR mutations on transcription (45). Berkhout and coworkers conducted a series of studies with HIV-1 variants that do not require TAR for transcription. These studies revealed that, although transcriptionally competent variants lacking the TAR hairpin can replicate (albeit less efficiently than the wild-type virus), long-term replication leads to restoration of a 5'-stem-loop structure (45). These and other studies showed that a stable 5' hairpin, but not necessarily TAR, is required for efficient genome packaging, and that disruption of base pairing can alter the folding and dimerization properties of the 5'-L (46). Similar studies revealed that a stable hairpin at the polyA position is required for efficient replication and RNA packaging (47). These results can be explained by our findings that cap sequestration, mediated by end-to-end stacking of the TAR and polyA helices, is required for efficient genome packaging. Mutations that disrupt base pairing near the bases of the TAR and polyA helices are likely to disrupt coaxial helical stacking and consequently expose the 5' cap.

Our finding that 5'-L<sup>AAGG</sup> and  $\Psi^{\text{CES}}$  test RNAs are not packaged, despite possessing wild-type NC binding and dimerization behaviors, indicates that cap exposure is a dominant-negative determinant of packaging. How does cap exposure prevent packaging? The inability of HIV-1 Gag proteins to package cap-exposed viral RNAs is not likely due to altered intracellular trafficking since live-cell imaging studies have shown that both translating and nontranslating pools of HIV RNAs actively transit sites of virus assembly (48, 49). Retroviral transcripts involved in translation and packaging exist in separate, noninterconverting pools (22, 48–51), and splicing of HIV-1 RNAs transcribed with three 5' guanines, which adopt a cap-exposed conformation, occurs at significantly greater levels than those transcribed with a single 5' guanine that are known to adopt a cap-sequestered conformation (52). These findings are collectively consistent with a mechanism in which HIV-1 transcripts containing an exposed 5' cap are irreversibly sequestered by the cellular processing and translation machinery prior to becoming accessible for Gag-dependent

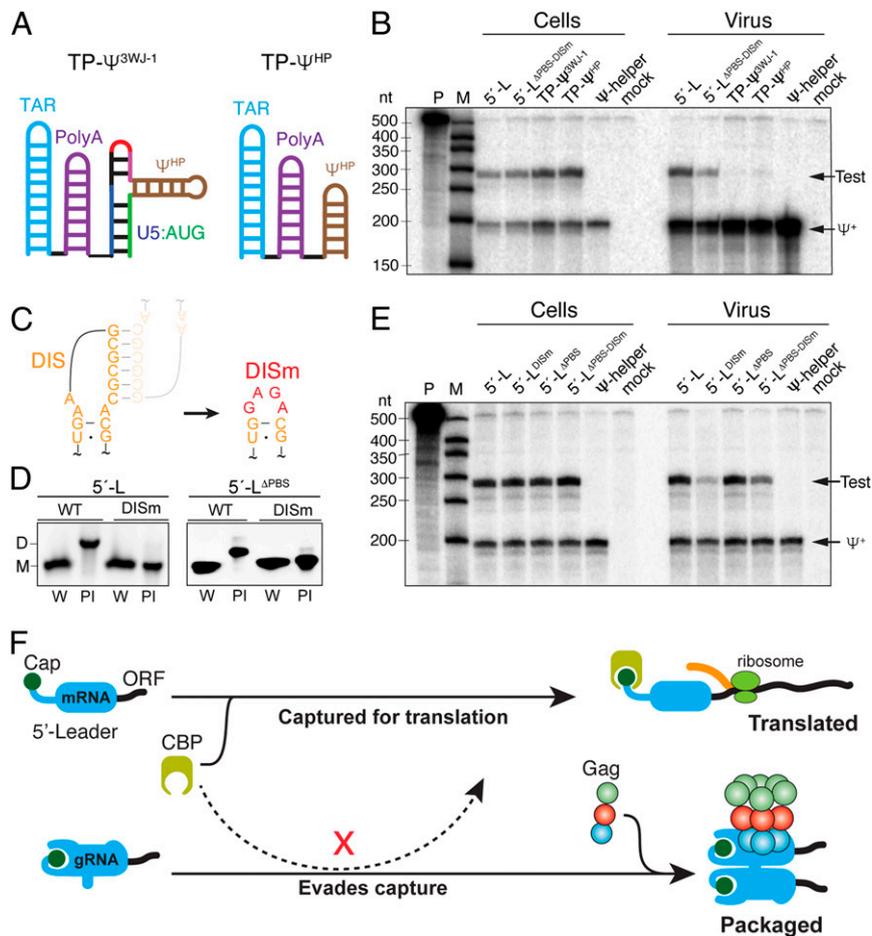
packaging, and that gRNAs avoid this capture by structural sequestration of their 5' cap (Fig. 5F).

The negative influence of cap exposure on packaging may also help explain why certain host noncoding RNAs synthesized by RNA polymerase III, including 7SL, U6 small nuclear RNA, and transfer RNAs, are enriched in retroviruses (53–56). These RNAs lack a 7-methylguanosine cap, and their expression and processing pathways are independent of the cap-dependent RNA processing and translation machinery (53, 56) (Fig. 5F). Cap sequestration may also be important for RNA packaging by other viruses; for example, the 5' cap of rotavirus plus-strand RNA associates with eIF4E when functioning as mRNA but is sequestered by the copackaged viral polymerase VP1 when functioning as gRNA (57, 58). Suppression of plus-strand viral RNA translation appears to be coupled with genome packaging by hepatitis B virus and flaviviruses and is also mediated by the binding of the viral polymerase to sites proximal to the 5' cap (59–61).

In summary, our studies reveal that cap sequestration is an essential determinant of HIV-1 genome packaging, that the TAR and polyA hairpins indirectly promote packaging by sequestering the cap, and that all additional RNA elements required for authentic packaging fidelity and efficiency reside fully within the intact  $\Psi^{\text{CES}}$  region of the dimeric 5'-L. Attenuated packaging of cap-exposed RNAs likely results from cap-dependent sequestration by cellular RNA processing and translation machinery. Since cap exposure can be induced at relatively modest energetic cost (a few kcal/mol) (23), the TAR–polyA cassette could serve as an attractive antiviral target.

## Materials and Methods

**Plasmid Construction and DNA Template Preparation.** The construction of plasmids for 5'-L, 5'-L<sup>ΔPB5</sup>, and  $\Psi^{\text{CES}}$  has been described previously (24, 29). Plasmids for 5'-L<sup>DISM</sup> were generated by site-directed mutagenesis on the 5'-L using forward primer 5'-CGC AGG ACT CGG CTT GCT GGA GAC GGC AAG AGG CGA GGG GCG G-3' and reverse primer 5'-CCG CCC CTC GCC TCT TGC CGT CTC CAG CAA GCC GAG TCC TGC G-3'. Plasmids for 1<sup>G5</sup>-L and 1<sup>G5</sup>-L<sup>ΔPB5</sup> were generated by site-directed mutagenesis using forward primer 5'-CTA ATA CGA CTC ACT ATA GTC TCT CTG GTT AGA CCA G-3' and reverse primer



**Fig. 5.** Dimeric  $\Psi^{CES}$  is required for authentic packaging fidelity and efficiency. (A) TP- $\Psi^{3WJ-1}$  and TP- $\Psi^{HP}$  contain the proximal three-way junction and the extended  $\Psi$ -hairpin, respectively. TP, cap-sequestering TAR–polyA cassette. (B) Test RNAs containing TP- $\Psi^{3WJ-1}$  and TP- $\Psi^{HP}$  were not detectably packaged in competition with  $\Psi^*$ . (C–E) Substitution of the native DIS loop by GAGA (C) prevents dimerization (D) and attenuates packaging of test RNAs containing 5'-L<sup>DISm</sup> and 5'-L<sup>DPBS-DISm</sup> under competitive conditions (E). (F) Working model for HIV-1 RNA packaging. Cap-dependent capture by the RNA processing or translation machinery precludes packaging of host and viral mRNAs. gRNAs are selected by a bipartite mechanism that involves structural sequestration of the 5' cap and exposure of a cluster of high-affinity Gag binding sites. CBP, cap-binding protein; ORF, open reading frame.

5'-CTG GTC TAA CCA GAG AGA CTA TAG TGA GTC GTA TTA G-3' on parent plasmids 5'-L and 5'-L<sup>APBS</sup>, respectively. The plasmid for  $^{16}\Psi^{CES}$  was prepared by site-directed mutagenesis using forward primer 5'-CTA ATA CGA CTC ACT ATA GTG CCC GTC TGT TGT GTG-3' and reverse primer 5'-CAC ACA GAC GGG CAC TAT AGT GAG TCG TAT TAG-3' on  $\Psi^{CES}$ . The plasmid for  $^{GGAAG}5'-L$  was prepared by site-directed mutagenesis using forward primer 5'-CTA ATA CGA CTC ACT ATA GGA AGG TCT CTC TGG TTA GAC C-3' and reverse primer 5'-GGT CTA ACC AGA GAG ACC TTC CTA TAG TGA GTC GTA TTA G-3' on plasmid 5'-L. The plasmid for  $^{GAAAG}5'-L$  was generated by site-directed mutagenesis on the 5'-L using forward primer 5'-CTA ATA CGA CTC ACT ATA GAA GGT CTC TCT GGT TAG ACC-3' and reverse primer 5'-GGT CTA ACC AGA GAG ACC TTC TAT AGT GAG TCG TAT TAG-3'. DNA fragments of  $\delta\text{-}\Psi^{CES}$  and  $\delta^{mut}\text{-}\Psi^{CES}$  were synthesized and inserted into pUC57 plasmids by Genewiz. The T7 promoter sequence was included at the 5' end of the RNA-encoding sequences in all the plasmids. All the site-directed mutagenesis was performed with the QuikChange Lightning Mutagenesis Kit (Agilent Technologies).

The DNA templates for RNA in vitro transcription were obtained by standard PCR amplification using the above-mentioned plasmids (EconoTaq PLUS 2X Master Mix; Lucigen). A common forward amplification primer targeting a vector region ~80 nt upstream of the T7 promoter sequence was used for all PCR reactions (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<sup>APBS</sup>, 1<sup>G</sup>5'-L, 1<sup>G</sup>5'-L<sup>APBS</sup>, GGAAG5'-L, GAAAG5'-L,  $\delta\text{-}\Psi^{CES}$ , and  $\delta^{mut}\text{-}\Psi^{CES}$ ; 5'-mGmGC ACC CAT CTC TCT CCT TCT AGC CTC C-3' for  $\Psi^{CES}$  and  $^{16}\Psi^{CES}$ ; and 5'-mCmAC TAC TTT GAG CAC TCA AGG CAA GC-3' for TAR–polyA; “m” denotes 2'-O-methyl modification to reduce nontemplated

nucleotide addition by T7 RNA polymerase (62). All the DNA oligos were synthesized by Integrated DNA Technologies.

**In Vitro Transcription.** RNAs were prepared by in vitro transcription as described previously (27). The reactions (15 to 30 mL) each contained ~0.5 mg of PCR-amplified DNA template, 20 mM MgCl<sub>2</sub>, 3 to 6 mM nucleoside triphosphates (NTPs), 2 mM spermidine, 2 mM dithiothreitol (DTT), 20% (volume [vol]/vol) dimethyl sulfoxide, 80 mM Tris-HCl (pH 9.0), and 0.3 mg T7 RNA polymerase (purified in-house). The reaction was quenched after a 5-h incubation at 37 °C by addition of an ethylenediaminetetraacetate (EDTA)/urea mixture (250 mM EDTA and 7 M urea, pH 8.0), followed by boiling for 5 min and snap cooling on ice for 5 min; 50% glycerol was added to the sample to a final concentration of 6% (vol/vol). RNAs were purified by electrophoresis on urea-containing polyacrylamide denaturing gels (SequaGel; National Diagnostics) at 20 W overnight, visualized by ultraviolet 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 washing eight times with ultrapure water using an Amicon Ultra Centrifugal Filter Device (Millipore).

**NTPs for In Vitro Transcription.** Fully protonated NTPs were purchased from Sigma. Perdeuterated NTPs were purchased from Cambridge Isotope Laboratories. A<sup>27</sup> was obtained by selective deuteration of the C8 position of fully protonated adenosine triphosphate (ATP). Briefly, 140  $\mu$ L triethylamine (TEA) was added to 0.2 g ATP dissolved in 8 mL D<sub>2</sub>O (99.8%; Cambridge Isotope Laboratories); 1-mL aliquots were incubated at 60 °C for 5 d. TEA was subsequently removed by lyophilization. G<sup>r</sup> was obtained by analogous treatments for guanosine triphosphate (GTP) with a shorter incubation time at 60 °C (24 h).

**Purification of Vaccinia Virus Capping Enzyme.** The expression vector containing the coding sequences of the His-tagged D1 and D12 subunits of vaccinia virus capping enzyme was a kind gift from the Stephen Cusack laboratory, European Molecular Biology Laboratory, Grenoble, France (63). The coexpressed D1 and D12 proteins from the plasmid were purified from BL21 (DE3) pLysS cells (Life Technologies). Briefly, the cells were cultured in Terrific broth at 37 °C with 250 rpm shaking until the OD reached ~1.0, followed by protein induction with 0.5 mM isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) at 20 °C overnight. The cells were harvested by centrifugation at 7,000 rpm at 4 °C for 20 min and resuspended in lysis buffer (20 mM Tris base, 200 mM NaCl, 20 mM imidazole, 10% glycerol, and 5 mM Tris[2-carboxyethyl]phosphine [TCEP], pH 8.0), and then lysed by freezing and thawing, followed by microfluidization. After centrifugation at 12,000 rpm at 4 °C for 30 min, the supernatant containing the His-tagged proteins was applied to an Ni-NTA affinity column (Qiagen) and eluted by elution buffer (20 mM Tris base, 200 mM NaCl, 250 mM imidazole, 10% glycerol, and 5 mM TCEP, pH 8.0). Proteins were dialyzed overnight at 4 °C in storage buffer (20 mM Tris-HCl, 100 mM NaCl, 10% glycerol, and 1 mM DTT, pH 8.0). The activity of the capping enzyme was assessed in a trial capping reaction using a 20-nt RNA. Capped RNA can be resolved from uncapped RNA on a 20% denaturing acrylamide gel (SequaGel; National Diagnostics) run at 220 V for 3 h. Only one species of capped RNA was generated by vaccinia virus capping enzyme with the cap moiety appended to the 5' end of the in vitro transcribed RNA. The appropriate amount of capping enzyme was chosen to achieve near-100% capping efficiency.

**Preparation of Capped RNA.** The capping reaction contained 16 mL of 20  $\mu$ M in vitro T7-transcribed and purified RNA in a buffer with 50 mM Tris base, 5 mM KCl, 1 mM MgCl<sub>2</sub>, and 1 mM DTT (pH 8.0). RNA was boiled and snap-cooled in water followed by the addition of 10 $\times$  capping buffer and 0.5 mM GTP, 0.1 mM S-adenosyl methionine, and various amounts of vaccinia virus capping enzyme (determined in a 20- $\mu$ L trial reaction with a 20-nt RNA). The reaction was incubated at 37 °C for 2 h and then quenched with the addition of the EDTA/urea mixture (250 mM EDTA and 7 M urea, pH 8.0) followed by boiling for 3 min and snap cooling on ice. The capped RNA was purified the same way as described above for in vitro transcribed RNAs.

**Purification of Human eIF4E.** The pUC57 plasmid containing the codon-optimized DNA sequence for human eIF4E (UniProt ID code P06730) with an N-terminal His<sub>6</sub> and GB1 tag was synthesized by Genewiz. The protein sequence is MGHHH HHHSS GGMQY KLILN GKTLK GETTT EAVDA ATA EK VFKQY ANDNG VDGEW TYDDA TKFTF VTEIP TTEEN YFQGA MATVE PETTP TPNPP TTEEE KTESN QEVAN PEHYI KHPLQ NRWAL WFFKN DKSKT WQANL RLISK FDTVE DFVAL YNHIQ LSSNL MPGCD YSLFK DGIEP MWEDE KNRKG GRWLI TLNKG QRRSD LDRFW LETLL CLIGE SFDDY SDDVC GAVVN VRAKG DKIAI WTTEC ENREA VTHIG RVYKE RGLP PKIVI GYQSH ADTAT KSGST TKNRF VV. The coding sequence was PCR-amplified and inserted into the pLATE11 expression vector (aLICator LIC Cloning & Expression System; Thermo Fisher Scientific). BL21 (DE3) cells harboring the plasmid were cultured in Luria-Bertani broth (LB) medium at 35 °C with 250 rpm shaking until the OD reached ~0.6, and protein was overexpressed by induction with 1 mM IPTG at 28 °C overnight. The pelleted cells were lysed by freeze-thawing and microfluidization in lysis buffer containing 50 mM Tris base, 500 mM NaCl, 20 mM imidazole, and 2 mM TCEP (pH 8.0). The supernatant after centrifugation was applied to an Ni-NTA affinity column (Qiagen) and eluted by elution buffer (50 mM Tris base, 500 mM NaCl, 250 mM imidazole, and 2 mM TCEP, pH 8.0). The eluted protein was further purified by size-exclusion chromatography on a Superdex 75 column (Amersham) in 20 mM Tris base, 100 mM KCl, and 2 mM DTT (pH 7.5).

**Native Gel Electrophoresis.** RNA samples (0.4 to 2  $\mu$ M) were prepared in PI buffer (10 mM Tris-HCl, pH 7.5, 140 mM KCl, 10 mM NaCl, and 1 mM MgCl<sub>2</sub>), heat-denatured for 3 min, and slowly cooled down to room temperature. For cap-binding gel shift with eIF4E, the protein was diluted into PI buffer and incubated at 37 °C for 1 h. The complex samples were prepared by mixing appropriate amounts of protein and RNA stocks, which were further incubated at 37 °C for 1 h; 10 $\times$  native agarose gel loading solution containing 0.17% bromophenol blue and 40% (vol/vol) sucrose was added to each sample and mixed. The final samples containing ~500 ng RNA were loaded onto 1% agarose gels prestained with ethidium bromide and run at 115 V in 1 $\times$  TB buffer (44.5 mM Tris-boric acid, pH 7.5) on ice to prevent thermal denaturation during electrophoresis.

**Purification of HIV-1<sub>NL4-3</sub> NC.** The NC protein was overexpressed and purified from BL21 (DE3) pLysE cells with a pET-3a plasmid containing the coding sequence of the HIV-1<sub>NL4-3</sub> NC (amino acid sequence: MQKGN FRNQR KTVKC FNCgK EGHIA KNCRA PRKKG CWKCG KEGHQ MKDCT ERQAN). The detailed purification steps have been described previously (27).

Ding et al.

5'-Cap sequestration is an essential determinant of HIV-1 genome packaging

**Isothermal Titration Calorimetry.** Isothermal titration calorimetry (ITC) experiments were carried out using an ITC200 (MicroCal). Both the protein and RNA samples were dialyzed into ITC buffer (10 mM Tris-HCl, pH 7.0, 140 mM KCl, 10 mM NaCl, 5 mM MgCl<sub>2</sub>, and 5 mM TCEP) overnight before use. The injection syringe contained 40  $\mu$ L ~240  $\mu$ M NC, and the calorimetry cell was loaded with 200  $\mu$ L ~1.0  $\mu$ M RNA. After thermal equilibration at 30 °C and the initial 60-s delay, a single injection of 0.2  $\mu$ L followed by 19 serial injections of 2  $\mu$ L was made into the calorimetry cell.

**NMR Spectroscopy.** NMR data were collected with Bruker AVANCE 600- or 800-MHz spectrometers equipped with cryoprobes. Samples for NMR studies of the capped RNAs were prepared in 90% H<sub>2</sub>O/10% D<sub>2</sub>O with 10 mM Tris-d<sub>11</sub> (pH 7.5), 50 mM KCl, and 1 mM MgCl<sub>2</sub> in standard 5-mm NMR tubes (500  $\mu$ L ~100  $\mu$ M). The NMR sample for 7mG (Sigma-Aldrich) was 5 mM. Two-dimensional NOESY spectra were collected at 35 °C with a 300-ms NOE mixing time and 3.2-s relaxation delay using a jump-return pulse sequence for water suppression. NMR data were processed with NMRfX (64) and analyzed with NMRViewJ (65).

**$\Psi^+$ -Helper and Test Vector Plasmid Construction.** The  $\Psi^+$ -helper NL4-3 GPP was described previously and consists of an HIV-1<sub>NL4-3</sub> genome with *vpu*, *vpr*, and partial *env* deletions and with *nef* replaced by a puromycin-resistance expression cassette (20). The native test vector plasmid (aka L688-RRE-Puro-LTR) contains 688 base pairs of the parental HIV-1<sub>NL4-3</sub> leader, the RRE, and a puromycin-resistance cassette as described previously (24). All changes in test vector 5'-L sequences were introduced by replacement of the corresponding region in the native test vector with a fragment generated using overlap extension PCR.

**Virus Production.** Human 293T cells were grown at 37 °C under 5% CO<sub>2</sub> in Dulbecco's modified Eagle's medium containing 10% fetal bovine serum and 50  $\mu$ g/mL gentamicin. HIV-1 particles were produced by 293T cells after cotransfection with plasmid DNA mixtures that included helper (NL4-3 GPP) and one test vector plasmid at an ~1:1 molar ratio using polyethylenimine, as described previously (27). Forty-eight hours later, transfection media were collected and filtered through 0.22- $\mu$ m filters. Cells were directly lysed with TRIzol reagent (Ambion) and viral particles were first concentrated by ultracentrifugation (25,000 rpm) through a 20% sucrose cushion and then lysed with TRIzol. RNA was extracted from TRIzol lysates according to the manufacturer's protocol, treated with DNase, and stored at -80 °C.

**RNA Content Analysis.** The RNA content of transfected cells and viral particles was analyzed by RPA as previously described (37). RNA samples were hybridized with riboprobe in 1 $\times$  hybridization buffer (80% formamide, 40 mM piperazine-N,N'-bis[2-ethanesulfonic acid], pH 6.4, 400 mM NaCl, and 1 mM EDTA) by denaturing at 95 °C for 5 min followed by a 16-h incubation at 65 °C. Samples were then digested in a mixture containing 300 mM NaCl, 10 mM Tris (pH 7.5), 10 mM EDTA, 10  $\mu$ g/ $\mu$ L salmon sperm DNA, 40  $\mu$ g/mL RNase A (Roche), and 10 U RNase T1 (Ambion) at 37 °C for 30 min. RNase-digested samples were further processed by the addition of sodium dodecyl sulfate (SDS) to a final concentration of 1.3% and 0.04 mg/ $\mu$ L proteinase K (Invitrogen) before incubation at 37 °C for 15 min, followed by phenol/chloroform extraction and ethanol precipitation. Pellets were resuspended in 2 $\times$  loading dye (80% formamide, 10 mM EDTA, 0.025% bromophenol blue, 0.025% xylene cyanol, and 0.1% SDS) and denatured at 95 °C for 5 min before loading onto a prewarmed 8% polyacrylamide/8 M urea gel and running at 500 V for 1.5 h. The gel was dried and exposed to a phosphorimager screen overnight. Imaging was performed using a Typhoon FLA 9500 (GE Healthcare) instrument, and radiolabeled bands were quantified using ImageQuant software (GE Healthcare). Packaging efficiency was quantified by determining the molar ratio of test vector to  $\Psi^+$ -helper RNAs present in viral particles as compared with their ratio in cells, normalized to packaging efficiency for a test vector with a native NL4-3 strain 5'-L. Packaging efficiencies are reported as the mean  $\pm$  SD from three individual packaging experiments. The chimeric riboprobe used in this study, HIVgag/CMV (cytomegalovirus), was described previously and protects a 201-nt fragment unique to the NL4-3 GPP helper and 289 nt of sequence unique to test vectors (27). The riboprobe was transcribed from linearized plasmid templates using T7 RNA polymerase (Promega) and [ $\alpha$ -<sup>32</sup>P]rCTP (PerkinElmer).

**Data Availability.** All data generated and analyzed during this study are included in this published article and its *SI Appendix*.

**ACKNOWLEDGMENTS.** We thank the Howard Hughes Medical Institute (HHMI) staff at the University of Maryland, Baltimore County (UMBC) and

C. Burnett at the University of Michigan Medical School for technical assistance, and J. Marchant at UMBC for helpful suggestions. This research was supported by research grants from the NIH (National Institute of Allergy and Infectious Diseases [NIAID] 8R01 AI50498 to M.F.S. and A.T. and NIAID U54

AI150470 to A.T.). E.C. and H.F. were supported by a National Institute of General Medical Sciences grant for enhancing minority access to research careers (MARC U\*STAR 2T34 GM008663); R.C. was supported by an HHMI undergraduate education grant and the Meyerhoff Scholars Program at UMBC.

1. J. M. Coffin, S. H. Hughes, H. E. Varmus, *Retroviruses* (Cold Spring Harbor Laboratory Press, Plainview, NY, 1997).
2. R. Berkowitz, J. Fisher, S. P. Goff, RNA packaging. *Curr. Top. Microbiol. Immunol.* **214**, 177–218 (1996).
3. 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).
4. R. S. Russell, C. Liang, M. A. Wainberg, Is HIV-1 RNA dimerization a prerequisite for packaging? Yes, no, probably? *Retrovirology* **1**, 23 (2004).
5. 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).
6. M. Kuzembayeva, K. Dille, L. Sardo, W.-S. Hu, Life of psi: How full-length HIV-1 RNAs become packaged genomes in the viral particles. *Virology* **454–455**, 362–370 (2014).
7. V. D'Souza, M. F. Summers, How retroviruses select their genomes. *Nat. Rev. Microbiol.* **3**, 643–655 (2005).
8. K. Lu, X. Heng, M. F. Summers, Structural determinants and mechanism of HIV-1 genome packaging. *J. Mol. Biol.* **410**, 609–633 (2011).
9. 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).
10. R. J. Kaddis Maldonado, L. J. Parent, Orchestrating the selection and packaging of genomic RNA by retroviruses: An ensemble of viral and host factors. *Viruses* **8**, 257 (2016).
11. M. Comas-Garcia *et al.*, Dissection of specific binding of HIV-1 Gag to the 'packaging signal' in viral RNA. *eLife* **6**, e27055 (2017).
12. A. Rein, RNA packaging in HIV. *Trends Microbiol.* **27**, 715–723 (2019).
13. 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).
14. N. Jouvenet, S. M. Simon, P. D. Bieniasz, Visualizing HIV-1 assembly. *J. Mol. Biol.* **410**, 501–511 (2011).
15. B. K. Ganser-Pornillos, M. Yeager, W. I. Sundquist, The structural biology of HIV assembly. *Curr. Opin. Struct. Biol.* **18**, 203–217 (2008).
16. G. Miele, A. Moulard, G. P. Harrison, E. Cohen, A. M. Lever, The human immunodeficiency virus type 1 5' packaging signal structure affects translation but does not function as an internal ribosome entry site structure. *J. Virol.* **70**, 944–951 (1996).
17. T. E. M. Abbink, B. Berkhout, A novel long distance base-pairing interaction in human immunodeficiency virus type 1 RNA occludes the Gag start codon. *J. Biol. Chem.* **278**, 11601–11611 (2003).
18. J. Greathouse, The retroviral RNA dimer linkage: Different structures may reflect different roles. *Retrovirology* **1**, 22 (2004).
19. 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).
20. K. Lu *et al.*, NMR detection of structures in the HIV-1 5'-leader RNA that regulate genome packaging. *Science* **334**, 242–245 (2011).
21. 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).
22. 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).
23. J. D. Brown *et al.*, Structural basis for transcriptional start site control of HIV-1 RNA fate. *Science* **368**, 413–417 (2020).
24. 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).
25. 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).
26. 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).
27. P. Ding *et al.*, Identification of the initial nucleocapsid recognition element in the HIV-1 RNA packaging signal. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 17737–17746 (2020).
28. C. K. Damgaard, E. S. Andersen, B. Knudsen, J. Gorodkin, J. Kjems, RNA interactions in the 5' region of the HIV-1 genome. *J. Mol. Biol.* **336**, 369–379 (2004).
29. S. C. Keane *et al.*, RNA structure. Structure of the HIV-1 RNA packaging signal. *Science* **348**, 917–921 (2015).
30. L. Didierlaurent *et al.*, Role of HIV-1 RNA and protein determinants for the selective packaging of spliced and unspliced viral RNA and host U6 and 7SL RNA in virus particles. *Nucleic Acids Res.* **39**, 8915–8927 (2011).
31. J. Luban, S. P. Goff, Mutational analysis of cis-acting packaging signals in human immunodeficiency virus type 1 RNA. *J. Virol.* **68**, 3784–3793 (1994).
32. 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).
33. S.-W. Wang, A. Aldovini, RNA incorporation is critical for retroviral particle integrity after cell membrane assembly of Gag complexes. *J. Virol.* **76**, 11853–11865 (2002).
34. D. Muriaux, J. Mirro, D. Harvin, A. Rein, RNA is a structural element in retrovirus particles. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 5246–5251 (2001).
35. D. Muriaux, J. Mirro, K. Nagashima, D. Harvin, A. Rein, Murine leukemia virus nucleocapsid mutant particles lacking viral RNA encapsidate ribosomes. *J. Virol.* **76**, 11405–11413 (2002).
36. S. J. Rulli Jr *et al.*, Selective and nonselective packaging of cellular RNAs in retrovirus particles. *J. Virol.* **81**, 6623–6631 (2007).
37. Y. Miyazaki *et al.*, An RNA structural switch regulates diploid genome packaging by Moloney murine leukemia virus. *J. Mol. Biol.* **396**, 141–152 (2010).
38. S. Kharytonchyk *et al.*, Influence of gag and RRE sequences on HIV-1 RNA packaging signal structure and function. *J. Mol. Biol.* **430**, 2066–2079 (2018).
39. J. L. Clever, D. Miranda Jr, T. G. Parslow, RNA structure and packaging signals in the 5' leader region of the human immunodeficiency virus type 1 genome. *J. Virol.* **76**, 12381–12387 (2002).
40. N. Fong, M. Ohman, D. L. Bentley, Fast ribozyme cleavage releases transcripts from RNA polymerase II and aborts co-transcriptional pre-mRNA processing. *Nat. Struct. Mol. Biol.* **16**, 916–922 (2009).
41. G. L. Buchschacher Jr, A. T. Panganiban, Human immunodeficiency virus vectors for inducible expression of foreign genes. *J. Virol.* **66**, 2731–2739 (1992).
42. Y. Liu *et al.*, HIV-1 sequence necessary and sufficient to package non-viral RNAs into HIV-1 particles. *J. Mol. Biol.* **429**, 2542–2555 (2017).
43. E. Mailler *et al.*, The life-cycle of the HIV-1 Gag-RNA complex. *Viruses* **8**, 248 (2016).
44. B. Klaver, B. Berkhout, Evolution of a disrupted TAR RNA hairpin structure in the HIV-1 virus. *EMBO J.* **13**, 2650–2659 (1994).
45. A. T. Das, A. Harwig, M. M. Vrolijk, B. Berkhout, The TAR hairpin of human immunodeficiency virus type 1 can be deleted when not required for Tat-mediated activation of transcription. *J. Virol.* **81**, 7742–7748 (2007).
46. A. T. Das, M. M. Vrolijk, A. Harwig, B. Berkhout, Opening of the TAR hairpin in the HIV-1 genome causes aberrant RNA dimerization and packaging. *Retrovirology* **9**, 59 (2012).
47. A. T. Das, B. Klaver, B. I. Klasens, J. L. van Wamel, B. Berkhout, A conserved hairpin motif in the R-U5 region of the human immunodeficiency virus type 1 RNA genome is essential for replication. *J. Virol.* **71**, 2346–2356 (1997).
48. J. Chen *et al.*, Visualizing the translation and packaging of HIV-1 full-length RNA. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 6145–6155 (2020).
49. J. Chen *et al.*, Impact of nuclear export pathway on cytoplasmic HIV-1 RNA transport mechanism and distribution. *mBio* **11**, e01578-20 (2020).
50. J. G. Levin, M. J. Rosenak, Synthesis of murine leukemia virus proteins associated with virions assembled in actinomycin D-treated cells: Evidence for persistence of viral messenger RNA. *Proc. Natl. Acad. Sci. U.S.A.* **73**, 1154–1158 (1976).
51. I. Boeras *et al.*, The basal translation rate of authentic HIV-1 RNA is regulated by 5' UTR nt-pairings at junction of R and U5. *Sci. Rep.* **7**, 6902 (2017).
52. J. M. Esquiaqui, S. Kharytonchyk, D. Drucker, A. Telesnitsky, HIV-1 spliced RNAs display transcription start site bias. *RNA* **26**, 708–714 (2020).
53. 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).
54. M. J. Eckwahl *et al.*, Analysis of the human immunodeficiency virus-1 RNA packaging. *RNA* **22**, 1228–1238 (2016).
55. A. Telesnitsky, S. L. Wolin, The host RNAs in retroviral particles. *Viruses* **8**, 235 (2016).
56. M. J. Eckwahl, A. Telesnitsky, S. L. Wolin, Host RNA packaging by retroviruses: A newly synthesized story. *mBio* **7**, e02025-15 (2016).
57. S. M. McDonald, J. T. Patton, Assortment and packaging of the segmented rotavirus genome. *Trends Microbiol.* **19**, 136–144 (2011).
58. X. Lu *et al.*, Mechanism for coordinated RNA packaging and genome replication by rotavirus polymerase VP1. *Structure* **16**, 1678–1688 (2008).
59. D. K. Ryu, B. Y. Ahn, W. S. Ryu, Proximity between the cap and 5' epsilon stem-loop structure is critical for the suppression of pgRNA translation by the hepatitis B viral polymerase. *Virology* **406**, 56–64 (2010).
60. A. A. Khromykh, A. N. Varnavski, P. L. Sedlak, E. G. Westaway, Coupling between replication and packaging of flavivirus RNA: Evidence derived from the use of DNA-based full-length cDNA clones of Kunjin virus. *J. Virol.* **75**, 4633–4640 (2001).
61. T. Fajardo *et al.*, The flavivirus polymerase NS5 regulates translation of viral genomic RNA. *Nucleic Acids Res.* **48**, 5081–5093 (2020).
62. C. Kao, M. Zheng, S. Rüdiger, 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).
63. M. De la Peña, O. J. Kyrieleis, S. Cusack, Structural insights into the mechanism and evolution of the vaccinia virus mRNA cap N7 methyl-transferase. *EMBO J.* **26**, 4913–4925 (2007).
64. 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).
65. B. A. Johnson, Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol. Biol.* **278**, 313–352 (2004).