SUPPLEMENTARY MATERIALS

Structural Basis for Transcriptional Start Site Control of HIV-1 RNA Fate

Joshua D. Brown[1], Siarhei Kharytonchyk[2], Issac Chaudry[1], Aishwarya S. Iyer[1], Hannah Carter[1], Ghazal Becker[1], Yash Desai[1], Lindsay Glang[1], Seung H. Choi[1], Karndeep Singh[1], Michael W. Lopresti[1], Matthew Orellana[1], Tatiana Rodriguez[1], Ubiomo Oboh[1], Jana Hijji[1], Frances Grace Ghinger[1], Kailan Stewart[1], Dillion Francis[1], Bryce Edwards[1], Patrick Chen[1], David A. Case[3], Alice Telesnitsky[2,*], and Michael F. Summers[1,*]

[1]Howard Hughes Medical Institute and Department of Chemistry and Biochemistry, University of Maryland Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250.

[2]Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor MI 48109-5620

[3]Department of Chemistry & Chemical Biology and BioMaPS Institute, Rutgers University, 610 Taylor Road, Piscataway, NJ 08854-8087

*Correspondence to: summers@hhmi.umbc.edu, ateles@umich.edu.

Plasmids for virology

Construction of the HIV-1 NL4-3 based retroviral vector GPP-pA, which encodes all HIV-1 proteins except Env and contains a SV40 polyadenylation site instead of the 3′-LTR, has been described previously (*12*). MAL-GPP-pA is similar to GPP-pA but contains HIV-1 MAL sequences from 1 to 368bp in place of the corresponding region of the NL4-3 leader. This was achieved using a series of overlap extension PCR reactions to generate a DNA fragment containing the HIV-1 MAL leader (residues 1-368bp) flanked with HIV-1 NL4-3 U3 and *gag* sequences. This PCR fragment was cloned into NL4-3 GPP-pA using PmlI and SpeI.

Viruses and cells

293T cells were grown at 37 °C with 5% CO_2 in DMEM media, containing 10% fetal bovine serum and 50 µg/ml gentamicin. For virus production, 293T cells at 70% confluence in 10 cm plates were transiently transfected with 6 µg of HIV-1 MAL-GPP-pA plasmid DNA using polyethylenimine as described [https://doi.org/10.1073/pnas.92.16.7297]. Virus media and cells were harvested 48 h after transfection.

RNA extraction

Virus containing media was filtered through 0.22 µm filters and then ultra-centrifuged (25,000 rpm) through a 20% sucrose cushion. The virus pellet was lysed with TRIzol Reagent (Ambion). To obtain cellular RNA, cells were washed once with PBS and lysed with 2.5 ml of TRIzol Reagent (Ambion) per 10 cm plate, and RNA was extracted according to the manufacturer's protocol. After ethanol precipitation, RNA was dissolved in 100 µl of TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0) and treated 15 min with 1 unit of RQ1 DNase (Promega). RNA then was re-extracted with

phenol-chloroform-isoamyl alcohol (PCI), ethanol precipitated, dissolved in TENS buffer (10 mM Tris, 1 mM EDTA, 150 mM NaCl, 0.5% SDS), and stored at -80 °C.

RNase protection assay. A riboprobe complementary to the 5′ end of the HIV-1_{MAL} RNA was transcribed using SP6 polymerase and a short PCR fragment as a template. RNase A was used for digestion of non-hybridized RNA. The products were resolved on a denaturing 15% PAGE (29:1 acrylamide/bis-acrylamide) at 1750V for ~ 5 h. Products were quantified by phosphorimaging. Data from four experimental repetitions revealed relative abundances of ^{Cap}1G and ^{Cap}3G RNAs of 34.6±1.5% and 65.4 ±1.5% in cells and 96±1.5% and 4±1.5% in virions, respectively.

DNA Template preparation and RNA *in vitro* Transcription

The pUC19 vectors containing the different RNA clones were PCR amplified (EconoTaq PLUS 2X Master Mix, Lucigen) using a forward primer 20-30 nucleotides upstream to the T7 promoter site and a reverse primer with the first two nucleotides containing 2′ O-methyl modifications. RNAs were synthesized by *in vitro* transcription using 7.5-15 mL reactions, each containing 80 mM Tris (pH 9.0), 2 mM DTT, 20% (vol/vol) DMSO, 2 mM spermidine, 20 mM $MgCl_2$, 3–6 mM NTPs (4x conc. of GMP was added to the reaction to prepare 5′MP RNA), ~0.5 mg of PCR-amplified DNA template, and 0.15 mg T7 RNA polymerase. The reaction was quenched after a 6 h incubation at 37 °C with 0.5 mM EDTA. RNA samples were heated for 5 minutes (100 °C) and snap cooled (0 °C, 5 minutes) prior to addition of 50% glycerol-water (volume/volume) and purification by electrophoresis on urea-containing polyacrylamide denaturing gels (SequaGel, National Diagnostics; 20 W for 12-18 h). The gel bands were visualized by UV-shadowing, excised, and eluted using the Elutrap electroelution system (Whatman) at 150 V overnight (16-20

h). The eluted RNAs were concentrated and washed with 2 M NaCl and then desalted using a 3 to 30-kDa MWCO Amicon Ultra-4 Centrifugal Filter Device (Millipore). The concentration of each sample was determined by measuring the optical absorbance at 260 nm.

Expression and Purification of the Vaccinia Virus Capping Enzyme

Plasmid containing the His-tagged vaccinia virus capping enzyme was a kind gift from the Stephen Cusack's lab at the European Molecular Biology Laboratory (EMBL) (*34*). The plasmid was transformed into BL21(DE3)pLysS cells (Life Technologies), grown in Terrific broth, induced with 0.5 mM IPTG at an OD of ~0.88 and grown at 20 °C overnight at 150 rpm. Cells were centrifuged at 8,000 rpm at 4 °C for 10 min and lysed in VVCE Lysis Buffer (40 mM Tris, 200 mM NaCl, 10 mM Imidazole, 5 mM TCEP, pH 8.0). The lysate was centrifuged at 18,000 rpm at 4 °C for 25 min. The supernatant was applied to a cobalt resin column and rocked for 1 h at 4 °C. The column was washed with VVCE Lysis Buffer and eluted with VVCE Elution Buffer (40 mM Tris, 200 mM NaCl, 250 mM imidazole, 5mM TCEP, pH 8.0). The elutions were dialyzed overnight at 4 °C in VVCE Dialysis Buffer (20 mM Tris, 100 mM NaCl, 0.100 mM EDTA, 10% glycerol, 1 mM DTT, pH 8.0).

Preparation and Purification of Capped RNAs

Capping reactions were performed in buffer containing 50 mM Tris, pH 8, 5 mM KCl, 1 mM $MgCl_2$, and 1 mM DTT with 20 μM RNA. 0.5 mM GTP, 0.1 mM SAM, and varying amounts of capping enzyme according to the previously determined substrate to enzyme ratio were added. The reactions were incubated at 37 °C for 1 h. Capping reactions were stopped with 0.5 mM EDTA followed by boiling for three minutes and snap cooling on ice water for three minutes. Capping of

small RNAs were determined by gel shift on a 20% denaturing acrylamide gel (SequaGel, National Diagnostics) run at 220 V for 2-4 h. Capped RNA was purified by gel electrophoresis as described above.

Native Agarose Gel Electrophoresis

The RNA (10 µM) was incubated at 37 °C (16-20 h) in the designated buffers; PI (physiological ions; 10 mM KH_2PO_4, pH 7.4, 1 mM $MgCl_2$, 122 mM KCl) or low ionic strength conditions (10 mM NaCl, 10 mM Tris, pH 7.4). Samples were removed from the incubator and immediately placed on ice. Native agarose gel loading solution containing 0.17% Bromophenol Blue and 40% (vol/vol) sucrose was added to each sample and mixed. 250-1000 ng RNA of each sample was loaded onto 1% (mass/vol) agarose tris borate without $MgCl_2$ (TB) or tris borate containing 0.2 mM $MgCl_2$ in the gel and running buffer (TBM). Ethidium bromide was added to the gel to a final concentration of 0.5 µg/mL. Gels were resolved at room temperature at 115 V for 75-100 min and visualized by UV illumination.

Sequence Conservation Analysis

Conservation of the polyA closing base pair (C57:G103) was studied by comparing sequences from the LANL HIV sequence database (*35*). A total of 1905 depositions with complete 5′-UTR sequences was retrieved and aligned with the Clustal algorithm (*36*) using Jalview v2.11.0 (*37*). Sequences that did not contain the full coverage of the polyA hairpin region (C57 through G103) were removed. The final alignment included 1092 sequences.

Selective C8 protonation and deuteration of ATP and GTP

The partially deuterated and perdeuterated NTP reagents used for *in vitro* transcription were obtained from Cambridge Isotope Laboratories (CIL, Andover, MA). Protonation at the C8 position of perdeuterated rGTP and rATP was achieved by incubation with triethylamine (TEA, 5 equiv) in H_2O (60 °C for 24 h and for 5 days, respectively). Deuteration of the C8 position of fully protonated GTP and ATP was achieved by analogous treatment with D_2O (99.8% deuteration; CIL). TEA was subsequently removed by lyophilization.

NMR spectroscopy

Under conditions of low ionic strength (sample buffers containing 10 mM KH_2PO_4), the ^{Cap}2G-L^{371} and ^{Cap}3G-L^{371} RNAs formed homogeneous populations of monomeric species at concentrations optimum for NMR experiments. (Note: small amounts of dimeric species were detected by gel electrophoresis and NMR at higher ionic strengths due to the high RNA concentrations required for two-dimensional NMR experiments (≥100 μM)). In order to study a homogenous population of the dimeric conformation at physiological salt conditions (10 mM KH_2PO_4, pH 7.4, 1 mM $MgCl_2$, 122 mM KCl), the RNA construct was truncated (359 nts) to prevent formation of the AUG hairpin and favor the intramolecular U5:AUG interaction, which subsequent favors DIS exposure. Samples for NMR studies of the monomeric RNA (^{Cap}2G-L^{371} and ^{Cap}3G-L^{371}) and its control oligos (550 μL of ~100 μM RNA in D_2O [99.8%; CIL] in a 5 mm Bruker NMR sample tube) were prepared in 10 mM KH_2PO_4 buffer (pD = 7.4), lyophilized and D_2O exchanged. Samples for NMR studies of the dimeric RNA (^{Cap}1G-L^{359}; 340 μL of ~100-150 μM RNA in D_2O [99.8%; CIL] in a 5 mm Bruker Shaped sample tube) and its control oligos (180 μL of ~300 μM RNA in D_2O [99.8%; CIL] in a 3 mm Bruker sample tube) were prepared in PI (physiological ions; 10 mM KH_2PO_4, pD 7.4, 1 mM $MgCl_2$, 122 mM KCl), lyophilized and D_2O

exchanged. Samples were also prepared in 90% H_2O + 10% D_2O to observe the exchangeable H8 of the cap residue. NMR data of the large RNAs (>120 nts) were collected with a Bruker AVANCE spectrometer (800 MHz, ^1H, 35 °C), Non-exchangeable ^1H assignments were obtained from 2D NOESY data (NOE mixing time = 650 ms, relaxation delay = 3.0 - 4.4 s, T = 35 °C). NMR data of the smaller control oligos (<120 nts) were collected with a Bruker AVANCE spectrometer (600 MHz, ^1H, 35 °C). Non-exchangeable ^1H assignments were obtained from 2D NOESY data (NOE mixing time = 300 ms, relaxation delay = 2.0 s, T = 35 °C). All NMR data were processed with NMRFx (*38*) and analyzed with NMRViewJ (*39*).

NMR spectral analysis

Assignment of 2D ^1H-^1H NOESY, ^1H-^1H TOCSY, and ^1H-^{13}C HMQC spectra were made for fully protonated RNA fragments using the standard NOE-based sequential assignment strategy (*20*). Assignments of these control RNAs were validated by comparisons with chemical shift values in the BioMagResBank NMR repository using NMRViewJ (*22, 40, 41*). Assignments from control RNAs were transposed from the control dataset to a 5′-L dataset and NOE patterns were verified and assigned. 2D NOESY spectra were recorded for the 5′-L RNA with a variety of labeling schemes. Prefixes denote sites of protonation, all other sites deuterated; e.g., G^H = fully protonated guanosines, A^{2r} = adenosines protonated at C2 and ribose carbons. Intraresidue NOEs were generally assigned from A^H-, C^H-, G^H-, and U^H-labeled samples. Cross-strand H_2-H_1', H_2-H_2, and sequential H_2(i)-H_1' (i+1) NOEs were assigned from A^H-, $A^{2r}G^r$-, $A^{2r}C^r$-, and $A^{2r}U^r$-labeled samples. Sequential H_8(i)-ribose(i-1), sequential H_6(i)-ribose(i-1), sequential aromatic [$H_{8/6}$(i)-$H_{8/6}$(i+1), H_2(i)-$H_{8/6}$(i+1), and H_2(i)-H_2(i+1)], sequential $H_{6/8}$(i)-H_5(i+1) and H_5(i)-H_5(i+1), and

sequential H_1'-H_1' NOEs were assigned from A^H-, C^H-, G^H-, U^H-, A^HG^H-, A^HC^H-, A^HU^H-, C^HG^H-, G^HU^H-, C^HU^H-labeled samples.

NOESY data for the ^{Cap}1G-L^{TPUA} structure (130 nucleotides) determination were obtained for RNAs with the following labeling schemes: A^H, C^H, U^H, G^H, $A^{2r}C^r$, $A^{2r}G^r$, $A^{2r}U^r$, C^{6r}, $C^{6r}U^{6r}$, U^{6r}, A^HC^H, C^HG^H, G^HU^H, A^HU^H, A^HG^H and G^HU^{6r}. NOESY data for ^{Cap}3G-TAR-F1 structural studies was derived from fully protonated RNA samples. Non-exchangeable ^{1}H assignments were obtained from 2D ^{1}H-^{1}H NOESY spectra (RNA = 150-300 µM, 10 mM KH_2PO_4, pD 7.4, 1 mM $MgCl_2$, 122 mM KCl, NOE mixing time = 650 ms, relaxation delay = 4.4 s, T = 35 °C). The ^{Cap}3G-TAR^m structure (33 nucleotides) was derived from fully protonated samples. Non-exchangeable ^{1}H assignments were obtained from 2D ^{1}H-^{1}H NOESY, ^{1}H-^{1}H TOCSY, and ^{1}H-^{13}C HMQC spectra (RNA = ~200 µM, 10 mM KH_2PO_4 buffer, pH = 7.4, NOE mixing time = 300 ms, relaxation delay = 2.0 s, T = 35 °C).

Structure calculations

Initial ensembles of structures were generated with CYANA. NOE-derived distance restraints were binned using standard categorization (upper ^{1}H-^{1}H distance limits of 2.7, 3.2, and 5.0 Å for NOE cross peaks of strong, medium, and weak intensity, respectively) except as follows, to allow greater conformational sampling: Medium-intensity intraresidue H8/6 to H2′ and H3′ NOEs = 4.0 and 3.0 Å, respectively; weak adenosine-H2 noes = 7.0 Å (for highly deuterated samples). Torsion angle restraints were employed for regions of A-helical geometry, allowing for ± 25° deviations from ideality (α=−62°, β=180°, γ=48°, δ=83°, ε=−152°, ζ=−73°). Hydrogen bonding restraints were used, and cross-helix P–P distance restraints (with 20% weighting coefficient) were employed for A-form helical segments to prevent collapse of major grooves (*42*).

The 20 CYANA-minimized structures with lowest target function were subjected to molecular dynamics simulations and energy minimization with AMBER (*43*). The upper limit NOE and H-bond distance restraints were employed, along with restraints to enforce planarity of aromatic residues and standard atomic covalent geometries and chiralities (*42, 44*). No backbone torsion angle or inter-phosphate restraints were employed during AMBER refinement. Calculations were performed using the RNA.OL3 (*45*) and generalized Born (*46*) force fields with parameters appropriate for RNA (igb=8, saltcon=0.15, rgbmax=25, cut=1000). The force field was modified to accommodate the Cap residue by using pre-defined parameters for 7-methylguanosine (from tRNA, based on (*47*)) and guanosine diphosphate (based on (*48*)). MD simulations included 2000 steps of un-restrained minimization followed by slow heating to a temperature of 300K (500 ps), MD simulation at 300K (5 ns), slow cooling to 0K (500 ps), and MD minimization at 0K (500 ps). Structures were analyzed with PyMOL (*49*).

Secondary Structure Predictions

Secondary structures were predicted with MFOLD (*50*) or RNAStructure (*51*).

Expression and Purification of eIF4E

A plasmid encoding the human eukaryotic translation initiation factor 4E (eIF4E) with an N-terminus His-GB1-TEV-Tag was chemically synthesized (Genewiz, NJ, USA) and expressed in BL21 *E. coli* cells in LB medium. After reaching an OD600 of 0.6 cells were induced by addition of 1 mM IPTG and grown overnight at 250 rpm at 28 °C. Cells were centrifuged at 8,000 rpm at 4 °C for 10 min and resuspended and lysed in eIF4E Lysis Buffer (50 mM Tris, 100 mM NaCl, pH 8.0). The lysate was centrifuged at 18,000 rpm at 4 °C for 25 min. 10% PEI was added dropwise

to the lysate supernatant and stirred at 4 °C for 30-60 min. The PEI supernatant was applied to a cobalt resin column and rocked overnight at 4 °C. The column was washed with eIF4E Lysis Buffer and eluted with eIF4E Elution Buffer (50 mM Tris, 100 mM NaCl, 300 mM imidazole, pH 8.0). The protein was dialyzed into TEV-cleavage buffer (25 mM Tris, 100 mM NaCl, 5mM BME, pH 8.0) and cleaved with TEV protease over two days. The untagged protein was applied to a second cobalt resin column. The flow-through was dialyzed into AIEX Buffer A (50 mM Tris, 100 mM NaCl, 2 mM EDTA, pH 8.0) and applied to a HiTrap Q-column. The column was washed with AIEX Buffer A and eluted with AIEX Buffer B (50 mM Tris, 1 M NaCl, 2 mM EDTA, pH 8.0). The protein was then concentrated to 5 mL and purified on a Superdex 200 column.

Gel Shift Assay for Detecting eIF4E-RNA Binding

All gel shift assays were performed in triplicate. RNA samples (1 μM) were incubated at at 37 °C in low ionic strength conditions (10 mM KH_2PO_4, pH 7.4, 1 mM $MgCl_2$, 122 mM KCl, Fig. 4B,C) or physiological-like conditions (PI buffer = 122 mM K^+, 1 mM Mg^{2+}, 10 mM KH_2PO_4, pH 7.4, Fig. 4D) (16-20 h) followed by a 1 h incubation with XRN-1 (New England BioLabs Inc., No. M0338L; XRN-1 concentration = 1 unit per μL where 1 unit is defined as the amount of enzyme required to fully digest 1 μg of RNA in 60 min at 37 °C). The eIF4E cap-binding protein (No. 15137 Cayman Chemicals, Fig 4B,C, or in-house expressed and purified untagged eIF4E; Fig. 4D) was incubated at 25 °C for 60 min. The samples were mixed at varying [eIF4E]:[RNA] ratios (0.5 to 4:1μM) and incubated at 37 °C for 2 h. Native agarose gel loading solution containing 0.17% Bromophenol Blue and 40% (vol/vol) sucrose was added to each sample and 300-500 ng of each

sample was loaded onto a 1% (mass/vol) TB gel containing ethidium bromide at a final concentration of 0.5 μg/mL. The gel was run at 115 V at 25 °C for 80 minutes and visualized by UV illumination.

Decapping and Exonuclease Digestion

^{Cap}3G-L^{371} or ^{Cap}1G-L^{Lock} RNAs (250-500 ng) were mixed with the catalytic domain of the human decapping complex, hDcp2 (Enzymax LLC, No. 86, Note: enzyme activity varied with lot number. The same lot number was used for all studies), XRN-1 (New England BioLabs Inc., No. M0338L; XRN-1 concentration = 1 unit per μL where 1 unit is defined as the amount of enzyme required to fully digest 1 μg of RNA in 60 min at 37 °C), and reaction buffer (1x Buffer: 10 mM Tris pH = 7.5, 100 mM NaCl, 2 mM $MgCl_2$, 1 mM DTT). The mixture was incubated at 37 °C for various time points (Fig. 4E, [hDcp2]/[RNA] = 0.43) or for 50 min (Fig. 4F). The reaction was quenched with the addition of 250 mM EDTA and 50% glycerol-water (volume/volume) and samples were loaded onto a 6% polyacrylamide denaturing gel. The gel was resolved at 220 V at 25 °C for 50 min, and stained via a carbocyanine dye (StainsAll, SigmaAldrich, No. E9379).

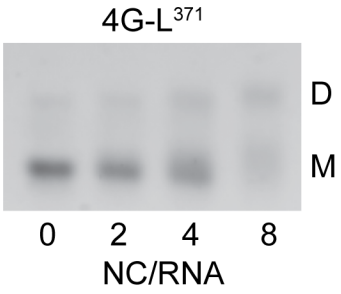

Figure S1. The monomer-dimer equilibrium of the HIV-1_{MAL} leader RNA is insensitive to the presence of the cognate NC protein. The monomer band retains intensity until the 8:1 NC:RNA molar ratio, where it begins to smear. The dimer band begins to shift by the 4:1 ratio of NC:RNA while the monomer band is the same suggesting that NC preferentially binds the dimer without shifting the monomer to the dimer.

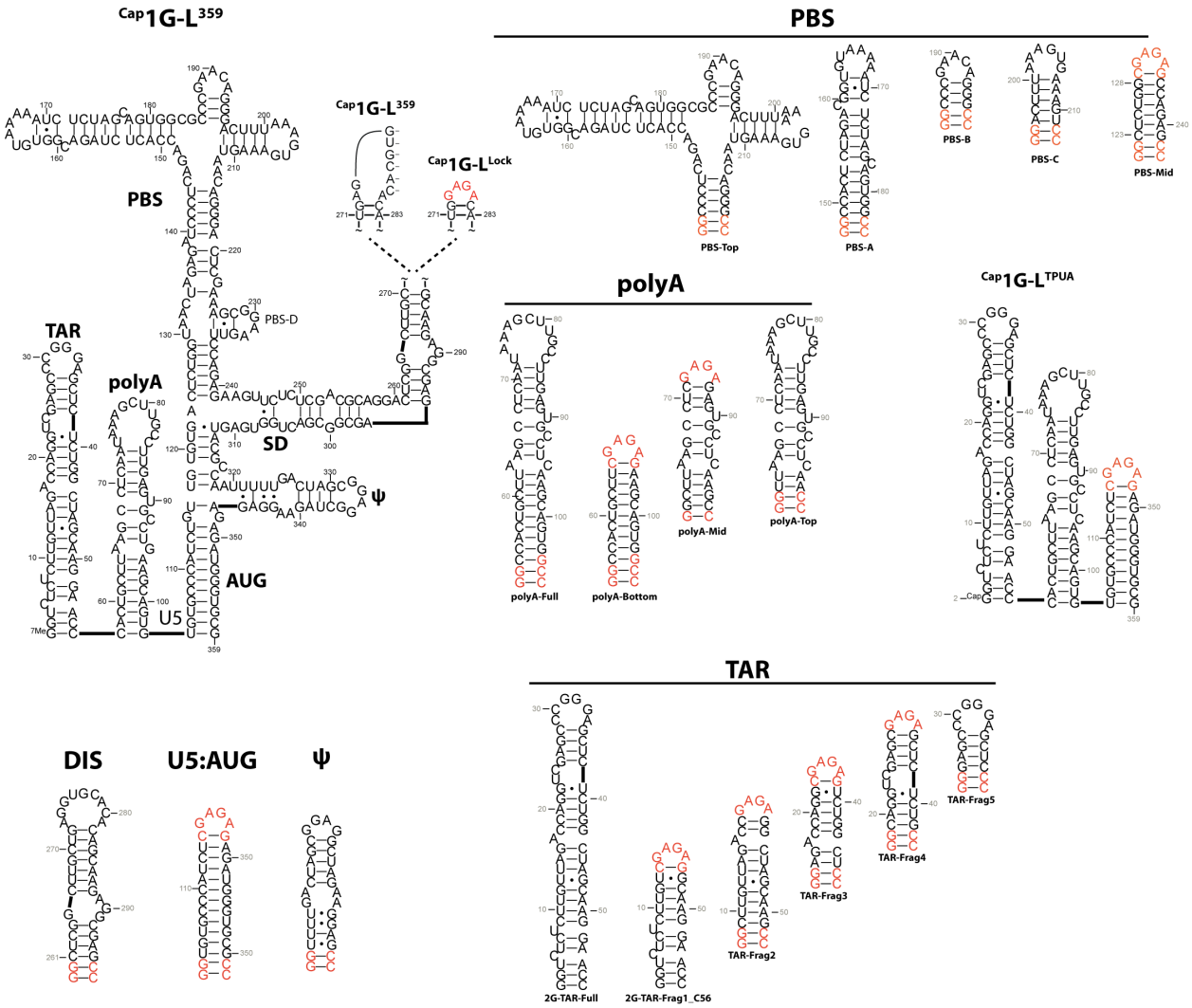

Figure S2. RNA constructs used to facilitate NMR assignment of HIV-1_{MAL} $[^{Cap}1G\text{-}L^{359}]_2$ and $^{Cap}1G\text{-}L^{TPUA}$. Non-native nucleotides used to promote *in vitro* transcription or stabilize helical structures are colored red. Fragments were designed with assistance of free energy predictions (MFOLD (*50*) and/or RNAStructure (*51*)). Improved spectral quality was obtained for a construct in which the DIS palindrome (GUGCAC) was mutated to GAGA ($^{Cap}1G\text{-}L^{Lock}$), which prevented dimerization while maintaining the secondary structure of the dimer (*27*).

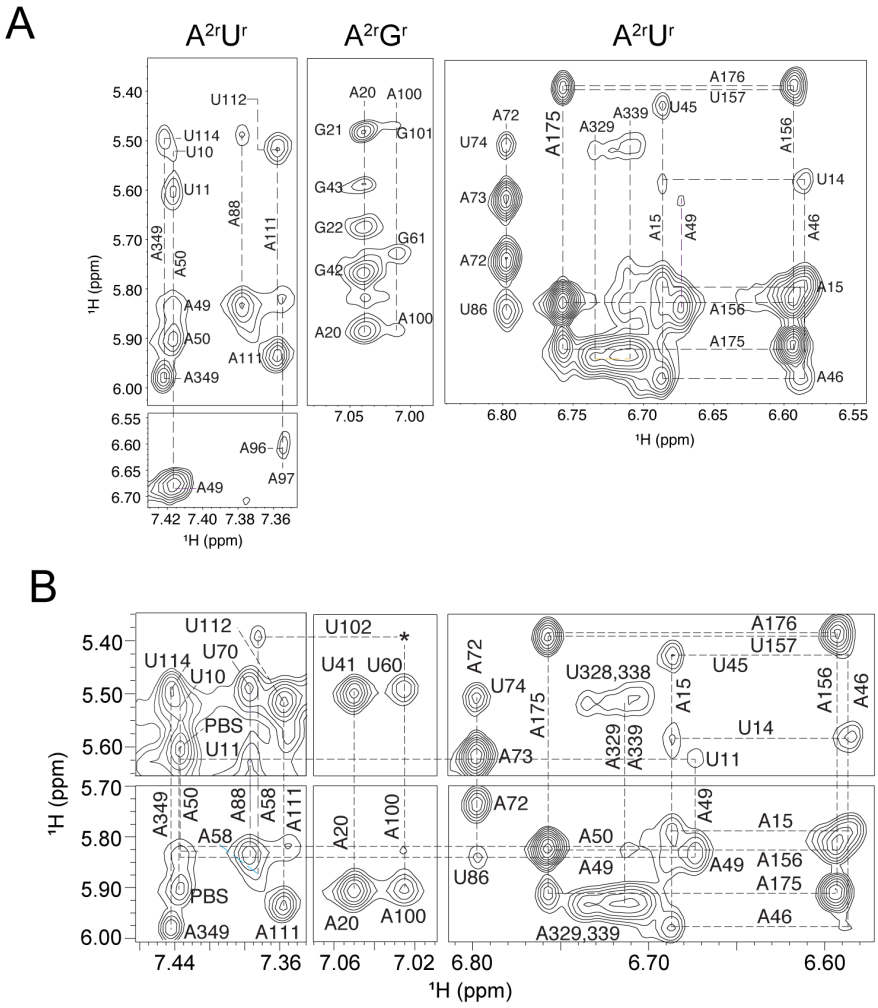

Figure S3. Representative 2D NOESY spectra obtained for HIV-1_{MAL} $[^{Cap}1G\text{-}L^{359}]_2$. **(A)** $A^{2r}U^{r}$- and $A^{2r}G^{r}$-$[^{Cap}1G\text{-}L^{359}]_2$ NOESY spectra **(B)** $A^{2r}U^{r}$-$[^{Cap}1G\text{-}L^{359}]_2$ NOESY data. These spectral regions show assignments for residues of the TAR, polyA and PBS hairpins and the U5:AUG helix.

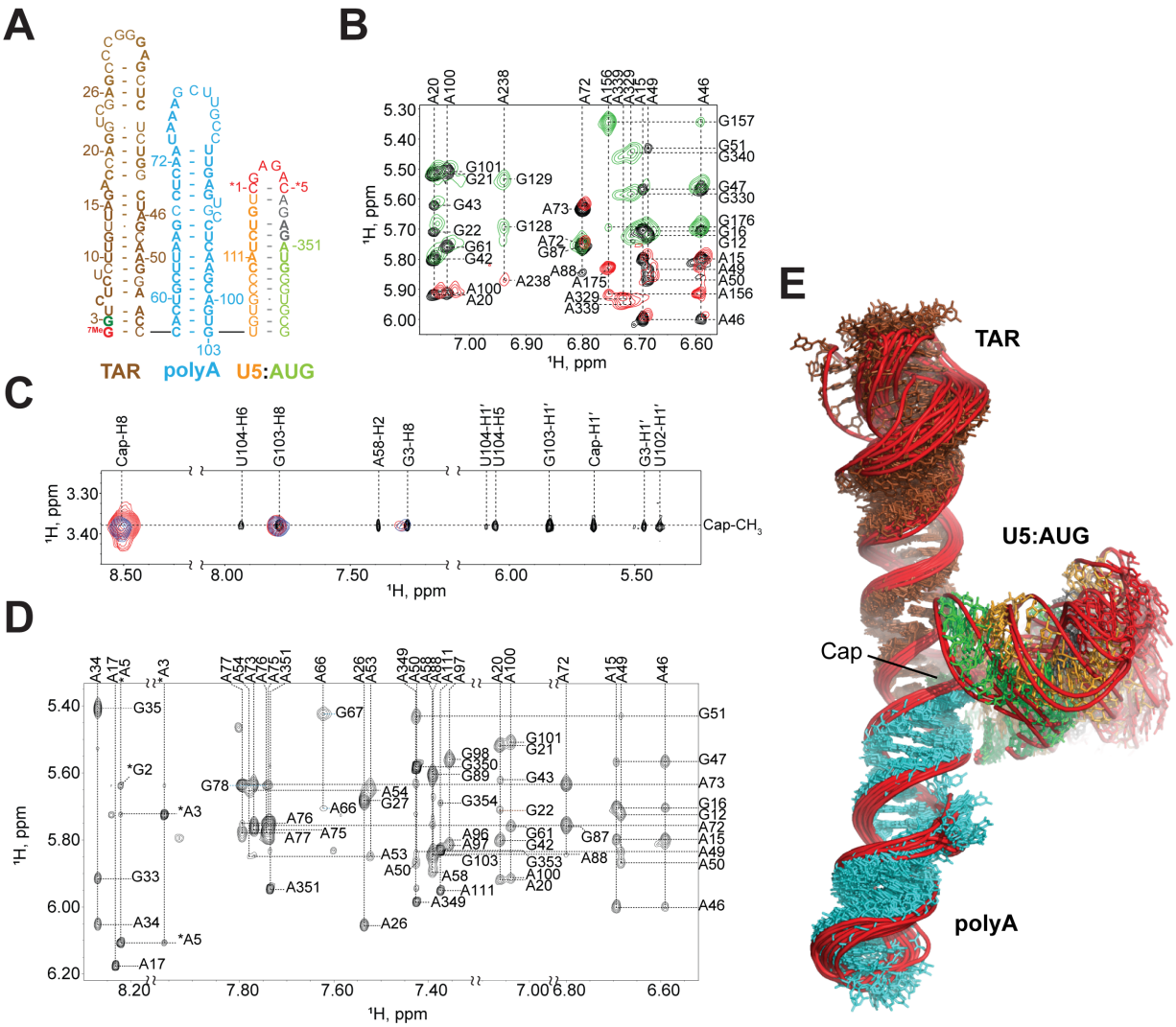

Figure S4. NMR spectra and structural findings for HIV-1_{MAL} ^{Cap}1G-L^{TPUA} **(A)** Secondary structure of the ^{Cap}1G-L^{TPUA}. **(B)** 2D ^{1}H-^{1}H NOESY spectra obtained for $A^{2r}G^{r}$-^{Cap}1G-L^{TPUA} (black) exhibits cross peaks and chemical shifts that match those of the [^{Cap}1G-L^{359}]$_2$ ($A^{2}G^{r}$; green, A^{2r}; Red). **(C)** Similarity of Cap-CH_3 NOEs in NOESY spectra obtained for fully protonated ^{Cap}1G-L^{TPUA} in D_2O (black), G^{H}-^{Cap}1G-L^{TPUA} in H_2O (red), and G^{8}-[^{Cap}1G-L^{359}]$_2$ in H_2O (blue). The NOEs (labeled) are consistent with stacking of the Cap residue between G3 of TAR and G103 of the polyA hairpin. **(D)** Portion of the 2D NOESY spectrum of $A^{2r}G^{r}$-^{Cap}1G-L^{TPUA} showing NOEs diagnostic of the TAR, polyA, and U5:AUG helices. **(E)** Three-dimensional structure of ^{Cap}1G-L^{TPUA} (see Table S1 for statistics). Residue colors match those in panel A.

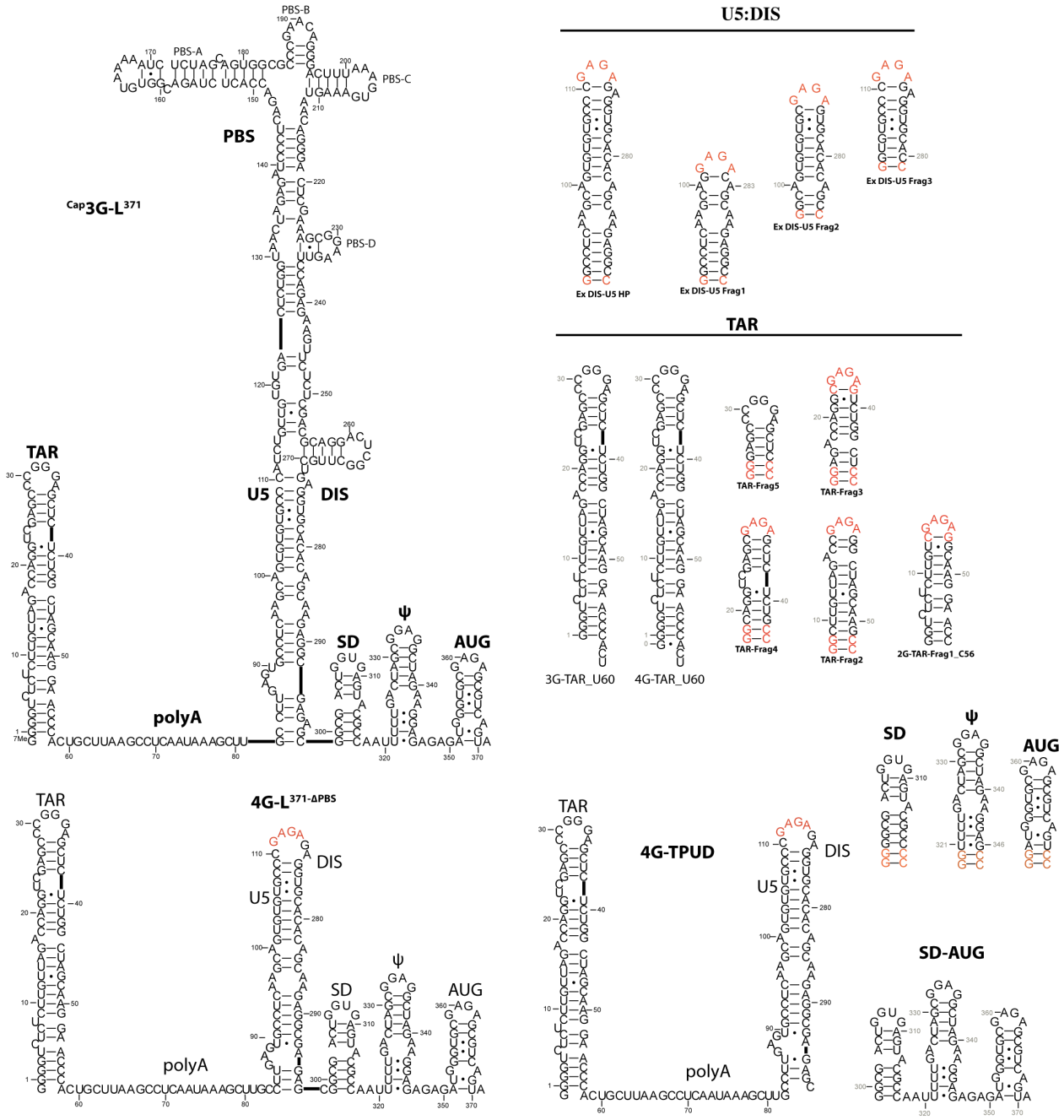

Figure S5. RNA fragments that facilitated NMR assignment of HIV-1_{MAL} ^{Cap}3G-L^{371}. Non-native nucleotides used to promote *in vitro* transcription or stabilize helical structures are colored red. Fragments were designed with assistance of free energy predictions (MFOLD (*50*) and/or RNAStructure (*51*)).

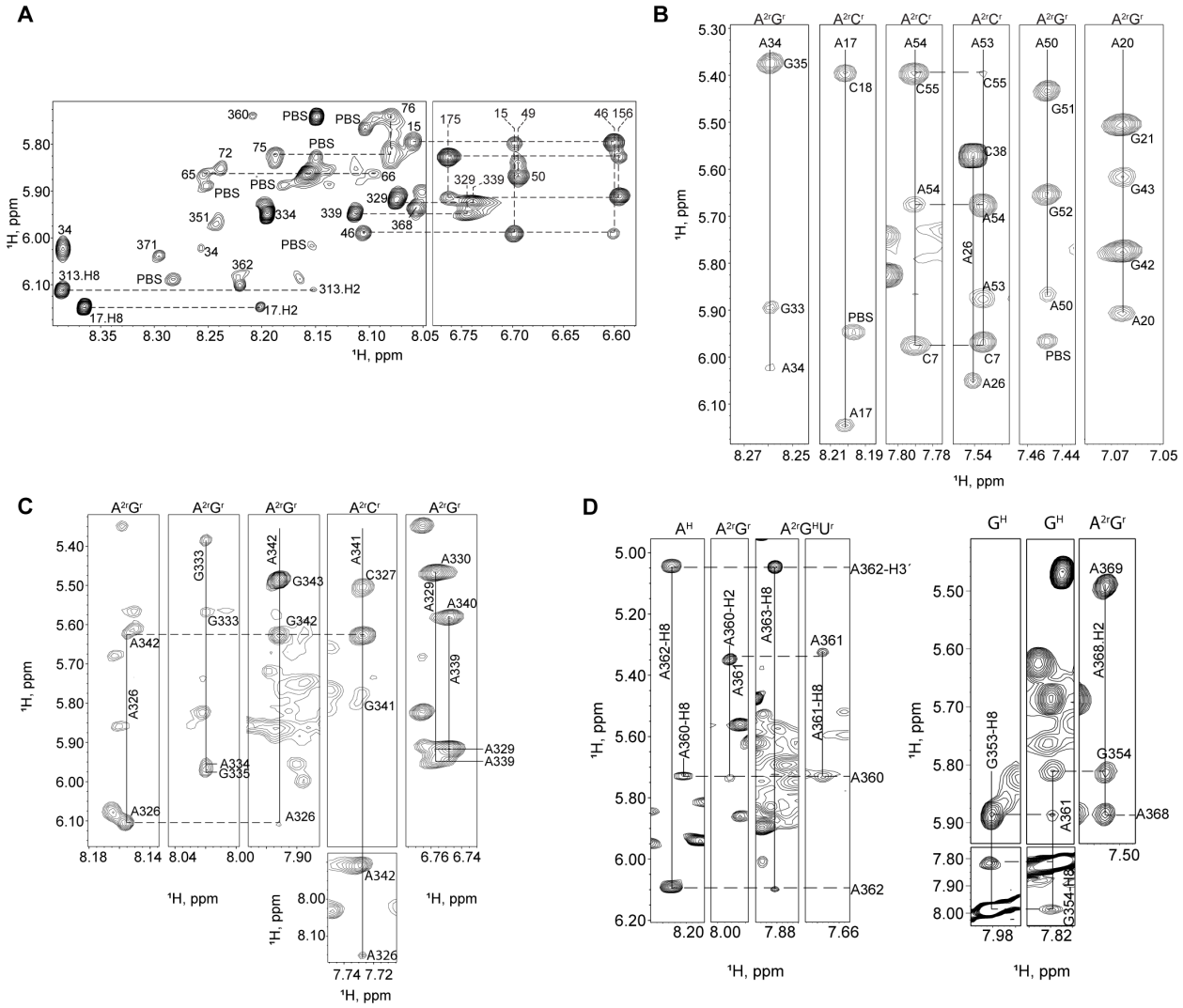

Figure S6. HIV-1_{MAL} ^{Cap}3G-L^{371} NMR spectra. **(A)** Portions of spectra for A^{H}-^{Cap}3G-L^{371} showing downfield-shifted H8 proton signals for A65, A66, A72, A73, A75, and A76 indicative of a non-base paired conformation. No cross-strand Adenosine-H2 NOEs were detected for these residues. **(B)** Portions of spectra for $A^{2r}G^{r}$- and $A^{2r}C^{r}$-^{Cap}3G-L^{371} showing assignments of the TAR hairpin. **(C)** portions of $A^{2r}G^{r}$- and $A^{2r}C^{r}$-^{Cap}3G-L^{371} spectra showing assignments of the Ψ hairpin. **(D)** Portions of spectra of A^{H}-, $A^{2r}G^{r}$-, and $A^{2r}G^{H}U^{r}$-^{Cap}3G-5′-L^{371} spectra showing assignments of the AUG hairpin and its associated GNRA.

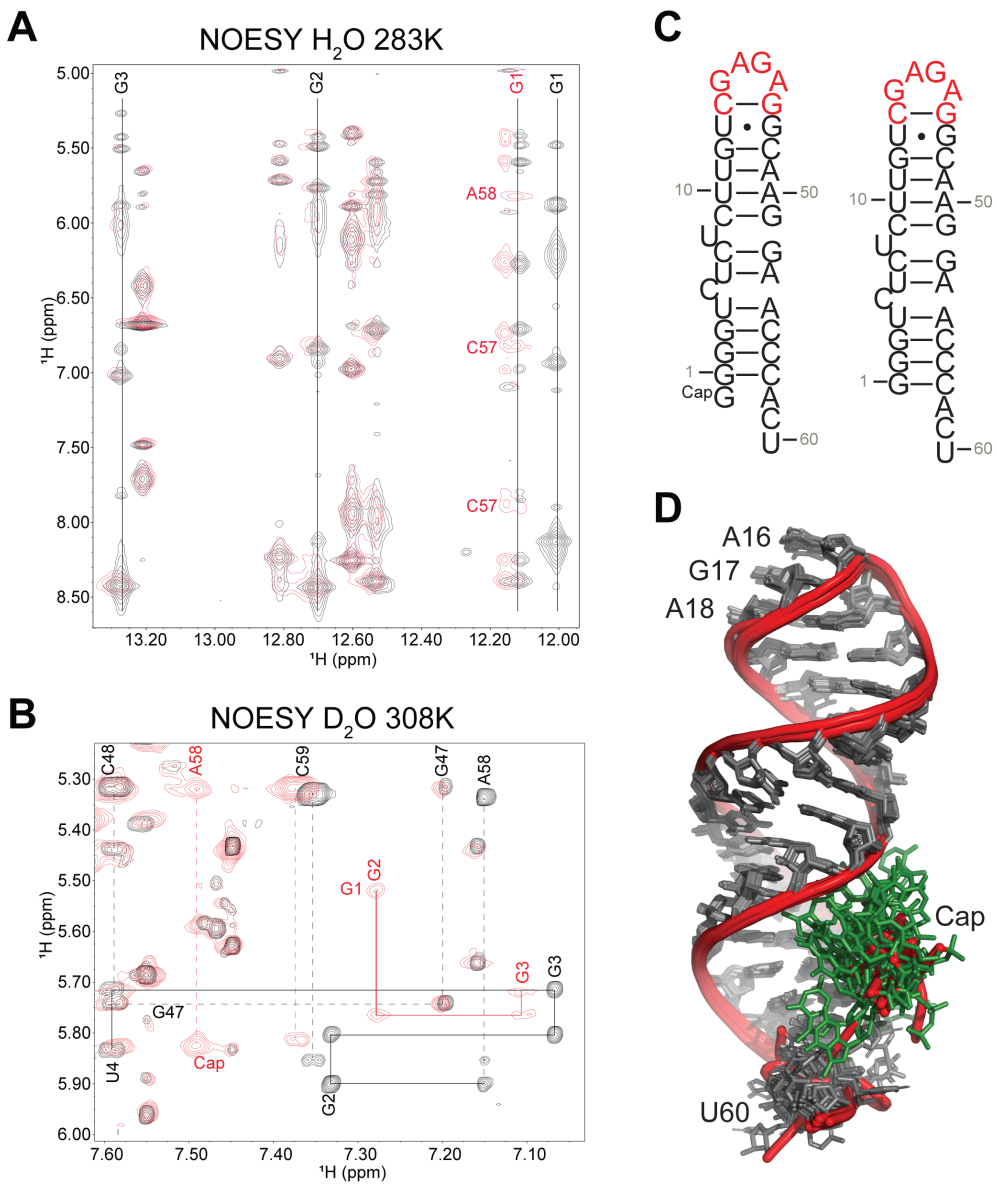

Figure S7. NMR spectra and structure of ^{Cap}3G-TAR^{m}. **(A)** Region of 2D NOESY spectra obtained for ^{Cap}3G-TAR^{m} (red) and non-capped 3G-TAR^{m} (black) showing signals associated with the imino protons (samples in 90% H_2O/10% D_2O). **(B)** NOESY spectra obtained for ^{Cap}3G-TAR^{m} (red) and non-capped 3G-TAR^{m} (black) in D_2O showing assignments for non-exchangeable protons. **(C)** NMR-derived secondary structures of ^{Cap}3G-TAR^{m} and non-capped 3G-TAR^{m}. The 5′ guanosine cap is methylated and is linked to G1 via a 5′-5′ triphosphate linkage. **(D)** Three-dimensional structure of ^{Cap}3G-TAR^{m} shown (see Table S1 for statistics).

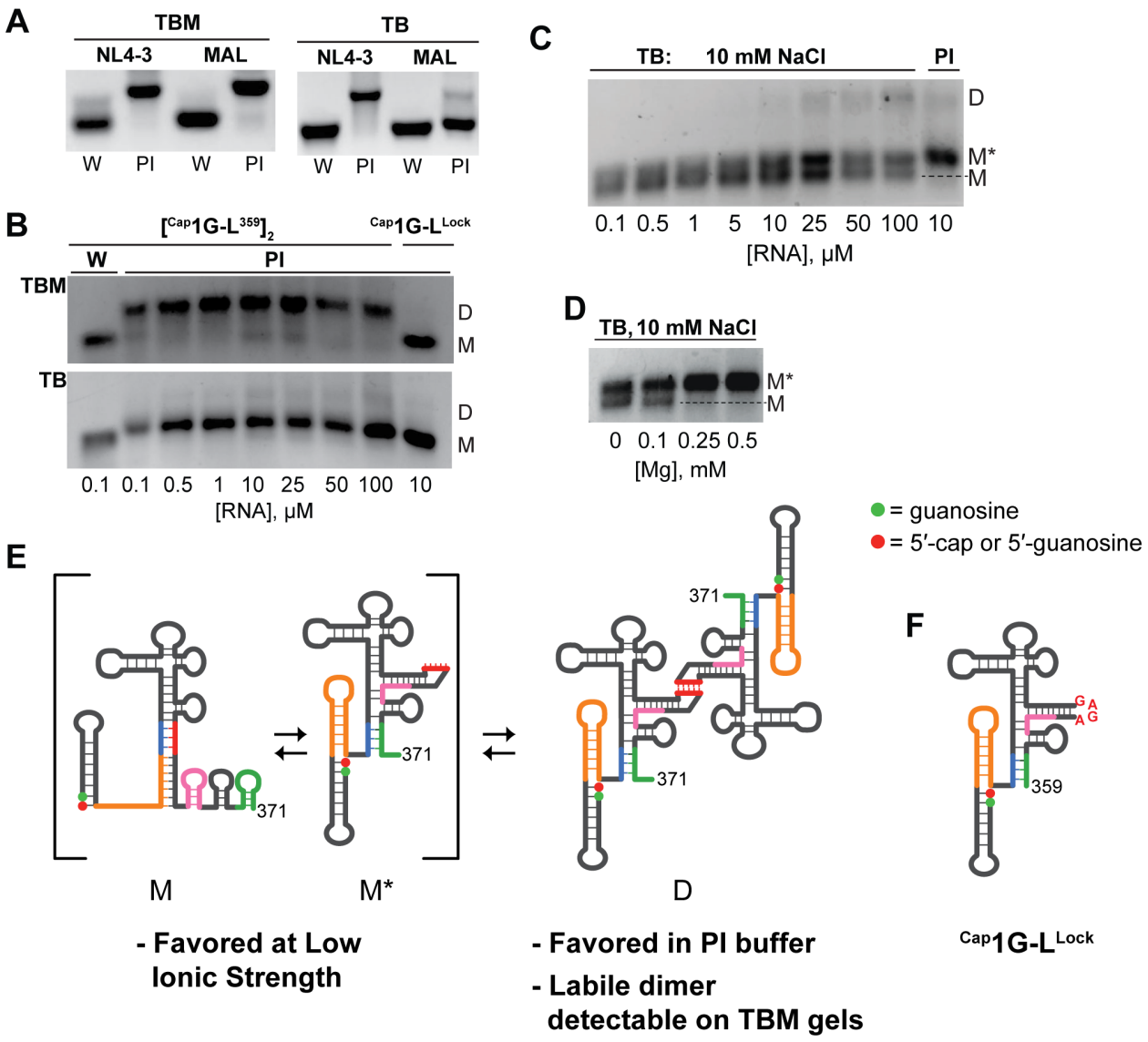

Figure S8.

Dimerization properties of ^{Cap}1G-L RNA (**A**) Non-capped HIV-1$_{MAL}$ 2G-L^{359} and the corresponding HIV-1$_{NL4-3}$ 2G-L^{344} leader RNAs migrate as monomers on both TBM (0.2 mM $MgCl_2$) and TB (no Mg^{2+}) gels when incubated in water. However, upon incubation in PI buffer, HIV-1$_{NL4-3}$ 2G-L^{344} migrates as a dimer on both gels whereas HIV-1$_{MAL}$ 2G-L^{359} migrates as a dimer on TBM gels but as a monomer on TB gels. These data indicate that, as observed for other HIV and SIV leader RNAs that contain A and U residues in the DIS palindrome, the HIV-1$_{MAL}$ 2G leader forms a labile dimer (*19, 52*). **(B)** As observed for the non-capped 2G analog, HIV-1$_{MAL}$

^{Cap}1G-L^{359} in PI buffer migrates as a dimer over a range of RNA concentrations (0.1-100 μM) but as a monomer on TB gels. The ^{Cap}1G-L^{Lock} construct, engineered to remain monomeric while retaining conformation of the dimer, does not form dimers on either TB or TBM gels. **(C, D)** Under non-physiological low salt conditions (10 mM NaCl, no Mg^{+2}), ^{Cap}1G-L^{359} adopts two monomeric conformers, M and M*, that are resolvable on TB gels with extended run time (~2 h, 115 V). The M* conformer is stabilized by addition of Mg^{+2} and appears to adopt the secondary structure the dimer (i.e., a pre-kissing dimer conformer and sequestered cap residue). The M conformer is unstable in the presence of Mg^{+2} and appears to adopt a structure similar to that of the ^{Cap}2G/^{Cap}3G leader (i.e., U5:DIS helix and exposed cap residue). **(E, F)** Cartoon representation of the M, M*, D conformers and ^{Cap}1G-L^{Lock} mutant.

Table S1. NMR Restraints and Structure Statistics

Cyana[1]	^{Cap}1G-L^{TPUA}	^{Cap}3G-TAR^{m}
NOE-derived restraints		
Intraresidue NOEs	252	62
Sequential NOEs	650	159
Medium range (|i – j < 6) NOEs	2	0
Long range (|i – j| > 6) NOEs	30	9
H-bond restraints[2]	534	124
NOE restraints/residue	12.4	12.4
H-bond restraints/residue	8.2	7.8
Backbone torsion angle restraints/residue	6.0	6.0
Total restraints/restrained residue	26.6	26.2
Target function (Å^2)	0.63 ± 0.01	0.25 ± 0.00
Upper distance viol. (Å^2)	0.01 ± 0.00	0.01 ± 0.00
Lower distance viol. (Å^2)	0.00 ± 0.00	0.00 ± 0.00
Sum VDW viol. (Å^2)	2.0 ± 0.0	0.9 ± 0.0
Amber[3]		
Amber energy (± stdev)	-32,702 ± 8	-9,112 ± 10
Distance	62.6 ± 1.4	57.1 ± 3.6
Torsion	9.4 ± 0.4	2.9 ± 0.2
Convergence (rmsd ± stdev, Å)[4] (superposed residues in parentheses)		
q	1.11 ± 0.35 Tar (2-56)	0.26 ± 0.09 Lower Tar (1-4,54-57)
	1.73 ± 0.47 polyA (57-103)	0.37 ± 0.16 Tar (1-57)
	1.56 ± 0.43 PolyA stem (57-73,85-103)	1.28 ± 0.42 All (0-60)
	0.24 ± 0.06 U5:AUG (104-359)	
	2.04 ± 0.71 Tar-PolyA stem (2-73,85-103)	
	5.53 ± 1.70 All (2-359)	

MolProbity analysis[5]		
Clashscore	0.30 ± 0.15, 99th %ile	0.74 ± 0.38, 99th %ile
Probably wrong sugar pucker (%)	0.45 ± 0.60, 0.35%	0.10 ± 0.31, 0.29%
Bad backbone conformation (%)	7.10 ± 1.77, 5.50%	0.55 ± 0.60, 1.67%
Bad bonds (%)	(1.00 ± 0.00)/3067, 0.03%	(1.00 ± 0.00)/787, 0.13%
Bad angles (%)	(1.15 ± 0.75)/4776, 0.02%	(0.05 ± 0.22)/1223, 0.00%

[1] Statistics for the 20 structures with lowest target function.

[2] Four restraints (two upper limit and two lower limit) per hydrogen bond employed during Cyana calculations; only the two upper H-bond distance limit restraints were employed during Amber refinement.

[3] Statistics for the 20 lowest energy structures.

[4] RMSDs versus mean coordinate positions (calculated with PyMol using an in-house script) for all C, N, O, and P atoms of residues shown in parentheses.

[5] The 20 amber-refined structures were evaluated using the MolProbity webserver (*53, 54*). The Cap residue was not included in MolProbity evaluations.

References and Notes

1. J. M. Coffin, S. H. Hughes, H. E. Varmus, *Retroviruses*. (Cold Spring Harbor Laboratory Press, Plainview, N.Y., 1997).
2. V. D'Souza, M. F. Summers, How retroviruses select their genomes. *Nature Reviews Microbiology* **3**, 643-655 (2005).
3. K. Lu, X. Heng, M. F. Summers, Structural determinants and mechanism of HIV-1 genome packaging. *J. Mol. Biol.* **410**, 609-633 (2011).
4. M. Kuzembayeva, K. Dilley, 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).
5. A. Onafuwa-Nuga, A. Telesnitsky, The remarkable frequency of human immunodeficiency virus type 1 genetic recombination. *Microbiol Mol Biol Rev* **73**, 451-480 (2009).
6. A. M. Lever, HIV-1 RNA packaging. *Adv Pharmacol* **55**, 1-32 (2007).
7. T. E. M. Abbink, M. Ooms, P. C. J. Haasnoot, B. Berkhout, The HIV-1 Leader RNA Conformational Switch Regulates RNA Dimerization but Does Not Regulate mRNA Translation†. *Biochemistry* **44**, 9058-9066 (2005).
8. K. Lu *et al.*, NMR detection of structures in the HIV-1 5′-leader RNA that regulate genome packaging. *Science* **344**, 242-245 (2011).
9. J.-L. Darlix, C. Gabus, M.-T. Nugeyre, F. Clavel, F. Barre-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).
10. 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).
11. 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).
12. 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).
13. Y. L. Chiu, E. Coronel, C. K. Ho, S. Shuman, T. M. Rana, HIV-1 Tat protein interacts with mammalian capping enzyme and stimulates capping of TAR RNA. *J Biol Chem* **276**, 12959-12966 (2001).
14. M. Zhou *et al.*, The Tat/TAR-dependent phosphorylation of RNA polymerase II C-terminal domain stimulates cotranscriptional capping of HIV-1 mRNA. *Proc Natl Acad Sci U S A* **100**, 12666-12671 (2003).
15. T. M. Menees, B. Muller, H. G. Krausslich, The major 5' end of HIV type 1 RNA corresponds to G456. *AIDS Res Hum Retroviruses* **23**, 1042-1048 (2007).
16. A. Sharma, A. Yilmaz, K. Marsh, A. Cochrane, K. Boris-Lawrie, Thriving under stress: selective translation of HIV-1 structural protein mRNA during Vpr-mediated impairment of eIF4E translation activity. *PLoS Pathog* **8**, e1002612 (2012).
17. A. Ramanathan, G. B. Robb, S. H. Chan, mRNA capping: biological functions and applications. *Nucleic Acids Res* **44**, 7511-7526 (2016).
18. M. Alizon, S. Wain-Hobson, J.-C. Gluckman, P. Sonigo, Genetic variability of the AIDS virus: nucleotide sequence analysis of two isolates from African patients. *Cell*, 63-74 (1986).

19. T. Tran *et al.*, Conserved determinants of lentiviral genome dimerization. *Retrovirology* **12**, 83 (2015).
20. K. Wüthrich, *NMR of Proteins and Nucleic Acids*. (John Wiley & Sons, New York, 1986).
21. 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).
22. J. D. Brown, M. F. Summers, B. A. Johnson, Prediction of hydrogen and carbon chemical shifts from RNA using database mining and support vector regression. *Journal of biomolecular NMR* **63**, 39-52 (2015).
23. 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).
24. T. Xia *et al.*, Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry* **37**, 14719-14735 (1998).
25. A. C. Gingras, B. Raught, N. Sonenberg, eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation. *Annu Rev Biochem* **68**, 913-963 (1999).
26. Z. Wang, X. Jiao, A. Carr-Schmid, M. Kiledjian, The hDcp2 protein is a mammalian mRNA decapping enzyme. *Proc Natl Acad Sci U S A* **99**, 12663-12668 (2002).
27. S. C. Keane *et al.*, Structure of the HIV-1 RNA packaging signal. *Science* **348**, 917-921 (2015).
28. A. Serganov, E. Nudler, A decade of riboswitches. *Cell* **152**, 17-24 (2013).
29. J. D. Lewis, E. Izaurralde, The role of the cap structure in RNA processing and nuclear export. *Eur J Biochem* **247**, 461-469 (1997).
30. B. M. Akiyama *et al.*, Zika virus produces noncoding RNAs using a multi-pseudoknot structure that confounds a cellular exonuclease. *Science* **354**, 1148-1152 (2016).
31. 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).
32. P. Carninci *et al.*, Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet* **38**, 626-635 (2006).
33. M. C. Frith *et al.*, A code for transcription initiation in mammalian genomes. *Genome Res* **18**, 1-12 (2008).
34. M. De la Pena, 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).
35. https://www.hiv.lanl.gov/.
36. J. D. Thompson, D. G. Higgins, T. J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673-4680 (1994).
37. A. M. Waterhouse, J. B. Procter, D. M. Martin, M. Clamp, G. J. Barton, Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-1191 (2009).
38. M. Norris, B. Fetler, J. Marchant, B. A. Johnson, NMRFx Processor: a cross-platform NMR data processing program. *Journal of biomolecular NMR*, 1-12 (2016).

39. B. A. Johnson, R. A. Blevins, NMRview: a Computer Program for the Visualization and Analysis of NMR Data. *J. Biomol. NMR* **4**, 603-614 (1994).
40. S. Barton, X. Heng, B. A. Johnson, M. F. Summers, Database proton NMR chemical shifts for RNA signal assignment and validation. *Journal of biomolecular NMR* **55**, 33-46 (2013).
41. J. Marchant, M. F. Summers, B. A. Johnson, Assigning NMR spectra of RNA, peptides and small organic molecules using molecular network visualization software. *Journal of biomolecular NMR* **73**, 525-529 (2019).
42. B. S. Tolbert *et al.*, Major groove width variations in RNA structures determined by NMR and impact of 13C residual chemical shift anisotropy and 1H-13C residual dipolar coupling on refinement. *J. Biomol. NMR* **47**, 205-219 (2010).
43. D. A. Case *et al.*, The Amber biomolecular simulation programs. *J. Computat. Chem.* **26**, 1668-1688 (2005).
44. I. Yildirim, H. A. Stern, J. D. Tubbs, S. D. Kennedy, D. H. Turner, Benchmarking AMBER force fields for RNA: comparisons to NMR spectra for single-stranded r(GACC) are improved by revised chi torsions. *J Phys Chem B* **115**, 9261-9270 (2011).
45. M. Zgarbova *et al.*, Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J Chem Theory Comput* **7**, 2886-2902 (2011).
46. J. Mongan, C. Simmerling, J. A. McCammon, D. A. Case, A. Onufriev, Generalized born model with a simple, robust molecular volume correction. *J. Chem. Theory Comput.* **3**, 156-169 (2007).
47. R. Aduri *et al.*, AMBER Force Field Parameters for the Naturally Occurring Modified Nucleosides in RNA. *J Chem Theory Comput* **3**, 1464-1475 (2007).
48. K. L. Meagher, L. T. Redman, H. A. Carlson, Development of polyphosphate parameters for use with the AMBER force field. *J Comput Chem* **24**, 1016-1025 (2003).
49. W. L. DeLano, "The PyMOL molecular graphics system," (DeLano Scientific, San Carlos, CA, 2002).
50. M. Zuker, Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**, 3406-3415 (2003).
51. J. S. Reuter, D. H. Mathews, RNAstructure: software for RNA secondary structure prediction and analysis. *BMC bioinformatics* **11**, 129 (2010).
52. R. Marquet, J.-C. Paillart, E. Skripkin, C. Ehresmann, B. Ehresmann, Dimerization of human immunodeficiency virus type 1 RNA involves sequences located upstream of the splice donor site. *Nucl. Acids Res.* **22**, 145-151 (1994).
53. I. W. Davis *et al.*, MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35**, W375-383 (2007).
54. V. B. Chen *et al.*, MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21 (2010).