NMR detection of intermolecular interaction sites in the dimeric 5′-leader of the HIV-1 genome

Sarah C. Keane[a,b], Verna Van[a,b], Heather M. Frank[a,b], Carly A. Sciandra[a,b], Sayo McCowin[a,b], Justin Santos[a,b], Xiao Heng[c,1], and Michael F. Summers[a,b,1]

[a]Howard Hughes Medical Institute, University of Maryland Baltimore County, Baltimore, MD 21250; [b]Department of Chemistry and Biochemistry, University of Maryland Baltimore County, Baltimore, MD 21250; and [c]Department of Biochemistry, University of Missouri, Columbia, MO 65211

Contributed by Michael F. Summers, September 7, 2016 (sent for review August 18, 2016; reviewed by Hashim M. Al-Hashimi and Bruno Sargueil)

HIV type-1 (HIV-1) contains a pseudodiploid RNA genome that is selected for packaging and maintained in virions as a noncovalently linked dimer. Genome dimerization is mediated by conserved elements within the 5′-leader of the RNA, including a palindromic dimer initiation signal (DIS) that has been proposed to form kissing hairpin and/or extended duplex intermolecular contacts. Here, we have applied a ^{2}H-edited NMR approach to directly probe for intermolecular interactions in the full-length, dimeric HIV-1 5′-leader (688 nucleotides; 230 kDa). The interface is extensive and includes DIS:DIS base pairing in an extended duplex state as well as intermolecular pairing between elements of the upstream Unique-5′ (U5) sequence and those near the *gag* start site (AUG). Other pseudopalindromic regions of the leader, including the transcription activation (TAR), polyadenylation (PolyA), and primer binding (PBS) elements, do not participate in intermolecular base pairing. Using a ^{2}H-edited one-dimensional NMR approach, we also show that the extended interface structure forms on a time scale similar to that of overall RNA dimerization. Our studies indicate that a kissing dimer-mediated structure, if formed, exists only transiently and readily converts to the extended interface structure, even in the absence of the HIV-1 nucleocapsid protein or other RNA chaperones.

HIV-1 | 5′-leader | RNA | 5′-untranslated region | structure

Like all retroviruses, HIV type-1 (HIV-1) selectively packages two copies of its unspliced, positive-sense RNA genome (1). Both strands are used for strand transfer-mediated recombination during reverse transcription (2–4), a process that promotes antiviral resistance by enhancing genetic diversity (5) and enables transcription at RNA breaks that occur naturally or are induced by restriction nucleases (6). Genomes are selected for packaging as dimers (7, 8), and studies indicate that dimerization and packaging are mechanistically coupled (9–11). Dimerization is promoted by a conserved "dimer initiation signal" (DIS) located within in the 5′-leader of the viral RNA (7, 12–19), the most conserved region of the viral genome (20). Elements that promote transcriptional activation (TAR) and splicing (SD) are also located within the 5′-leader, and these functions similarly appear to be modulated by dimerization (21–24) [although the role of dimerization on translation has been questioned (25)].

The DIS is a pseudopalindrome that contains a central, 4–6-nucleotide GC-rich element. Structural studies with oligoribonucleotides have shown that the DIS can adopt a dimeric kissing loop structure, in which residues of the GC-rich loops participate in intermolecular base pairs and the flanking residues form intramolecular base pairs within the stem of the hairpin (26–30). The DIS is also capable of adopting an extended duplex conformer in which all base pairing is intermolecular, and structures of DIS oligonucleotides in this extended duplex state have also been determined (29, 31, 32). Notably, DIS oligonucleotides with relatively long stems that include nonpaired bulges are capable of forming kissing dimers that can be converted to thermodynamically more stable extended duplex dimers (33), a process that may be facilitated by conformational fluctuations in the stem (34, 35). In addition, genomes isolated from immature virions can be readily dissociated by mild heating (36), whereas genomes from mature virions are more heat stable, with stability that increases with the age of the virus (37). The combined data are consistent with a model in which genomes are selected for packaging as less stable kissing dimers that convert to the more stable extended dimer conformers during viral maturation (13, 38–41). Other elements within the 5′-leader, including the pseudopalindromic TAR hairpin, the SD element, and residues overlapping the Unique-5′ (U5) and *gag* start codon regions (AUG; Fig. 1), have also been implicated in dimerization (42–44).

In this study, we developed a ^{2}H-edited NMR strategy to identify the intermolecular interface of the HIV-1 dimeric 5′-leader RNA (5′-L). The approach is conceptually similar to an ^{15}N-edited NMR line-shape method used to distinguish between inter- and intramolecular base pairing in a 39-nucleotide DIS oligonucleotide (45), but provides additional ^{1}H NMR chemical shift and NOE information and affords sensitivity and resolution sufficient to allow probing of the intact, dimeric HIV-1 5′-leader (NL4-3 strain; 688-nucleotide dimer; 230 kDa). Our studies reveal that, under physiological-like conditions and in the absence of nucleocapsid or other RNA chaperones, the DIS element adopts an extended duplex structure and U5 participates in intermolecular base pairing with AUG. The time scale for formation of this extended interface is similar to that of overall RNA dimerization, suggesting that kissing structures, if they exist, occur transiently along the dimerization pathway.

Significance

A nucleotide-specific ^{2}H-edited NMR approach was used to determine the nature of the intermolecular interface in the intact, dimeric HIV type-1 (HIV-1) 5′-leader RNA (230 kD). The studies distinguish between previously proposed extended duplex and kissing hairpin models and identify additional intermolecular interaction sites. A one-dimensional ^{2}H-edited NMR method that allows temporal monitoring of intermolecular base-pair formation revealed that the observed "extended dimer interface" forms rapidly, even in the absence of RNA chaperones. In addition to addressing long-standing questions about retroviral genome dimerization, these studies illustrate the utility of ^{2}H-edited NMR for determining the structures and folding kinetics of relatively large RNAs.

Author contributions: S.C.K., X.H., and M.F.S. designed research; S.C.K., V.V., H.M.F., C.A.S., S.M., J.S., and M.F.S. performed research; S.C.K. contributed new reagents/analytic tools; S.C.K., V.V., H.M.F., C.A.S., S.M., J.S., and M.F.S. analyzed data; and S.C.K., V.V., H.M.F., C.A.S., S.M., J.S., X.H., and M.F.S. wrote the paper.

Reviewers: H.M.A.-H., Duke University Medical Center; and B.S., Laboratoire de Cristallographie et RMN Biologiques.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

[1]To whom correspondence may be addressed. Email: summers@hhmi.umbc.edu or hengx@missouri.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1614785113/-/DCSupplemental.

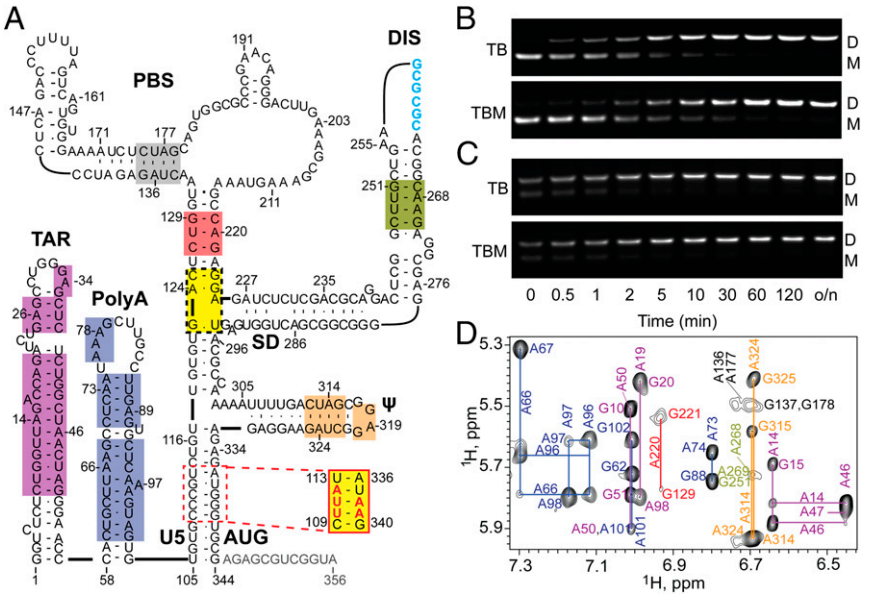

Fig. 1. (*A*) Secondary structure of the HIV-1 5′-L_{356} in its DIS-exposed, dimer-promoting state (adapted from ref. 57; residues truncated in 5′-L_{344} colored gray). The 5′-L_{344} NMR signals with resolved and assigned 2D ^{1}H–^{1}H NOESY cross-peaks are denoted by color shaded boxes; yellow denotes sites that were only assignable in mutant constructs that either lacked the upper PBS loop (5′-$L_{344\Delta PBS}$) (57) or contained an lrAID substitution in the U5:AUG stem (5′-$L_{344\text{-}UUA}$) (48). Intact (*B*) and truncated (*C*) 5′-leader constructs rapidly dimerize when incubated in PI buffer at 37 °C. Dimerization profiles on TB and TBM gels are indistinguishable, indicating that the dimers are nonlabile. o/n, overnight. (*D*) Region of the 2D ^{1}H–^{1}H NOESY spectrum of $A^{2r}G^rC^r$-labeled 5′-L_{344} RNA. Assignments (adenosine-H2 and ribose-H_1' assignments labeled vertically and horizontally, respectively) are color-coded to match the highlighted elements within the secondary structure in *A*.

Results

Sample Preparation and Characterization. The structural and dimerization properties of retroviral 5′-leader RNAs have been studied under a range of experimental conditions using RNAs prepared and purified by different methods. Previous studies in our laboratory used plasmids that had been linearized with restriction endonucleases as templates for RNA transcription, producing RNAs that contained several nonnative nucleotides at the 3′-end. We recently found that the inclusion of nonnative nucleotides at the 3′-end of 5′-leader samples alters the dimerization properties of those RNAs (46), consistent with earlier observations (42). Here, we used a method for preparation of HIV-$1_{NL4\text{-}3}$ 5′-leader RNA samples that precluded the use of nonnative 3′-residues (47). This approach also inhibited self-templated run-on during transcription, which was a readily detectable and significant problem for one of the constructs used in the current study (see *SI Materials and Methods* and Fig. S1). Samples prepared using this approach afforded ^{1}H NMR spectra with narrower signals, consistent with improved sample homogeneity. The full-length 5′-leader prepared by this method (residues 1–356, 5′-L_{356}; Fig. 1*A*) readily formed dimers when incubated under conditions of physiological-like ionic strength (PI) (PI buffer = 140 mM K^+, 10 mM Na^+, and 1 mM Mg^{+2}) and detected by native agarose gel electrophoresis (Fig. 1*B*). Equilibrium was reached ∼50-fold faster compared with constructs containing three nonnative 3′-cytidines (48).

Previous studies have shown that HIV-1_{LAI} leader constructs are capable of forming a "labile" dimer that can only be distinguished from monomers by electrophoresis when Mg^{+2} is included in the Tris–borate gel and running buffers [Tris-borate magnesium (TBM)] (49) and that the labile dimer can be converted to a nonlabile dimer (detectable in agarose gels using Tris–borate buffers that lack magnesium; TB) by incubation at 60 °C (50). The labile dimer was proposed to contain a kissing DIS interface and the nonlabile dimer an extended duplex DIS interface (50). Some fragments of the HIV-1_{MAL} leader (51), as well as intact leaders of other lentiviruses including HIV-2 and SIV (46), also form labile dimers that can only be detected by including Mg^{+2} in the electrophoresis running buffer. In contrast, we did not observe significant differences in the dimerization rate of the HIV-$1_{NL4\text{-}3}$ 5′-leader when measured using TB or TBM running buffers (Fig. 1*B*), consistent with reports that the leaders of some strains of HIV-1 do not form labile dimers (51).

The 3′-end of the HIV-1 5′-leader contains a conserved stretch of residues that are capable of either forming a local hairpin structure or base pairing with U5 (43, 48, 52). NMR studies indicate that the hairpin species is adopted in the monomeric form of the RNA and that U5:AUG pairing occurs in the dimer (48). Deletions or mutations that prevent formation of the AUG hairpin but do not affect U5:AUG pairing promote dimerization (48). The present studies were therefore performed with 3′-truncated leader constructs that include all residues required for U5:AUG base pairing but lack the 3′-residues necessary for formation of the competing AUG hairpin structure (residues 1–344, 5′-L_{344}; Fig. 1*A*) (48). The 5′-L_{344} formed nonlabile dimers, and the rate of dimerization exceeded that of the intact leader (time to reach equilibrium, ∼10 min and ∼30 min, respectively; Fig. 1 *B* and *C*).

The improved NMR spectral quality obtained with the present samples enabled us to extend assignments of adenosine-H2 NMR signals observed in ^{2}H-edited 2D NOESY NMR spectra (53). A portion of the 2D NOESY spectrum obtained for a dimeric 5′-L_{344} sample containing nonexchangeable protons on the adenosine C2 and ribose (A^{2r}), guanosine ribose (G^r), and cytosine ribose (C^r) carbons (all other nonexchangeable sites deuterated; 5′-L_{344}-$A^{2r}G^rC^r$) is shown in Fig. 1*D*, and assigned residues are labeled in Fig. 1*A*. Some regions that could not be directly examined due to severe signal overlap were observable in constructs that either lacked the upper PBS loop (5′-$L_{344\Delta PBS}$) or contained conservative base-pair substitutions in the U5:AUG region that resulted in resolved adenosine-H2 signals (long-range Adenosine Interaction Detection, lrAID) (48) (Fig. 1*A*).

DIS Adopts an Extended Duplex Structure in the Dimeric HIV-1 5′-Leader. Although the above NMR approach enabled direct detection of secondary structure elements in the dimeric leader,

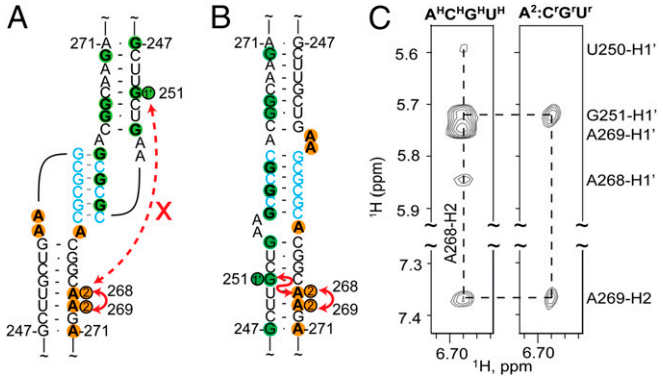

Fig. 2. DIS-containing RNAs prepared with different deuterium labeling schemes can form heterodimers with kissing (*A*) or extended duplex (*B*) interfaces. The A268-H2 to G251-H_1' separation is too great for NOE detection in the kissing conformer but short enough to be readily detected in extended duplex conformer. (*C*) Portions of the 2D ^{1}H–^{1}H NOESY spectra obtained for uniformly protonated $DIS_{237–281}$ oligo-RNA ($A^HC^HG^HU^H$) (*Left*) and an equimolar mixture of A^2 and $C^rG^rU^r$-labeled $DIS_{237–281}$ (*Right*).

it did not distinguish between inter- and intramolecular interactions. To directly probe for intermolecular base pairing, we used a modified ^{2}H-edited NMR method that involved 2D ^{1}H–^{1}H NOESY studies of differentially labeled heterodimers. With this approach, two RNA samples with identical sequences were prepared separately using different nucleotide-specific ^{2}H labeling schemes and purified by denaturing gel electrophoresis, then mixed, washed with salt, and incubated under conditions that promote refolding (PI buffer; 37 °C; 1 h). Samples prepared in this manner contain a statistical 50:50 distribution of heterodimers and homodimers, and labeling schemes are chosen such that cross-helix NOEs would only be observable for the heterodimeric interaction being probed.

As a proof of concept, this strategy was applied to a 47-nucleotide oligo-RNA corresponding to the isolated HIV-1_{NL4-3} DIS element [$DIS_{237–281}$; residues G237-C281, plus a nonnative 5′-G to improve transcription yield (54) and 3′-C to close the helix]. The 2D NOESY spectrum of the fully protonated $DIS_{237–281}$ dimer exhibited NOEs between H2 protons of adenosines on one side of the helix minor groove and H_1' protons of residues across the minor groove, which are characteristic of A-form helices and indicate that the proton pairs are separated by ∼5 Å or less. To determine if these are intramolecular NOEs indicative of a kissing dimer or intermolecular NOEs indicative of an extended duplex dimer, a mixture was prepared in which one RNA contained protons only on the adenosine–C2 carbons ($DIS_{237–281}$-A^2) and the other only on the guanosine, cytosine, and uridine ribose carbons ($DIS_{237–281}$-$G^rC^rU^r$), with all remaining nonexchangeable sites of both RNAs being deuterated (Fig. 2*A*). The A268-H2 and G251-H_1' protons are separated by more than 40 Å in the kissing dimer and would not give rise to an A268-H2 to G251-H_1' intermolecular NOE, whereas these protons would be in close proximity (<4 Å) and give rise to a significant NOE signal in the extended duplex conformer (Fig. 2*A* and *B*). The observation of a well-resolved A268-H2 to G251-H_1' intermolecular NOE in the spectrum obtained for the mixed sample provides strong direct evidence that $DIS_{237–281}$ adopts an extended duplex structure under the conditions used (Fig. 2*C*).

To investigate the nature of the base pairing within the DIS element in the HIV-1 5′-leader, two HIV-1 5′-L_{344} RNAs were prepared with different ^{2}H labeling strategies. An A^2-labeled 5′-L_{344} RNA was mixed with a G^r-labeled 5′-L_{344} RNA. The 2D ^{1}H–^{1}H NOESY spectrum collected on the mixed sample was compared with that collected on a uniformly A^2G^r-labeled sample. Consistent with our results from the isolated $DIS_{237–281}$ RNA, an NOE from A268-H2 to G251-H_1' is observed (Fig. 3*A*). The combined NOE and chemical shift information provide direct evidence that DIS adopts an extended intermolecular duplex structure in the context of the dimeric 5′-leader.

Identification of Other Intermolecular Interaction Sites. By using different labeling combinations, it was possible to probe for intermolecular base pairings for all adenosines that gave rise to resolved ^{1}H NMR signals. The TAR hairpin contains an unusual triplet of base pairs, [5′-^{12}UUA^{14}]:[5′-^{45}UAA^{47}], that gives rise to a well-resolved A46-H2 NMR signal (∼6.5 ppm) (48). In a homodimeric 5′-L_{344} sample with $A^{2r}G^rC^r$ labeling, A46-H2 exhibits a typical cross-strand NOE with the H_1' proton of A14, as well as a sequential NOE with A47-H2 (Fig. 3*B*). Spectra obtained for an A^2:A^r mixed 5′-L_{344} sample (Fig. 3*B*) exhibited the sequential A46-H2 to A47-H2 NOE but not the sequential A46-H2 to A47-H_1' NOE, as neither RNA molecule contains protons at both the aromatic C2 and ribose carbons. Importantly, the cross-strand A46-H2 to A14-H_1' signal was also absent in the mixed sample, indicating that base pairing in the helical region of TAR is intramolecular. Similar results were obtained for samples that probed the PolyA, ψ, and PBS helices, indicating that these helical substructures also contain intramolecular base pairs (Fig. 3*A* and Fig. S2).

Several helical regions of the native 5′-leader could not be probed by this approach due to severe signal overlap. However, it was possible to probe for intermolecular interactions in the U5:AUG

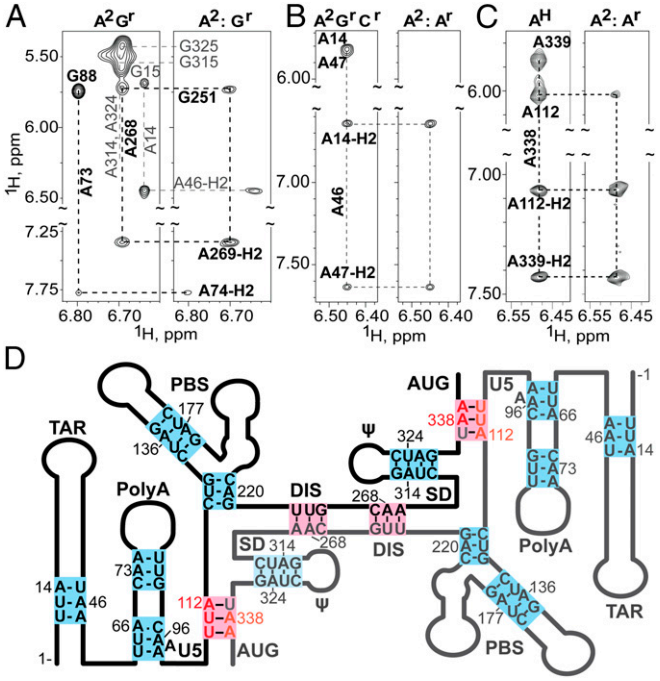

Fig. 3. (*A–C*) Portions of 2D NOESY spectra obtained for mixed, differentially ^{2}H-labeled 5′-L_{344} RNAs that probe the DIS and PolyA helices (NOEs that differentiate inter- or intramolecular base pairing are denoted with black labels and thick lines; noninformative NOEs are labeled gray). (*A*) Spectra obtained for A^2G^r (*Left*) and mixed A^2:G^r 5′-L_{344} (*Right*) samples. The A268-H2 to G251-H_1' NOE observed for the mixed sample is indicative of intermolecular base pairing at DIS (see also Fig. 2*C*). In contrast, the absence of an A73-H2 to G88-H_1' in the mixed sample indicates that base pairing in the PolyA helix is intramolecular. (*B*) Regions of spectra obtained for $A^2G^rC^r$ and A^2:A^r 5′-L_{344} samples that probe TAR. Absence of an A46-H2 to A14-H_1' cross-strand NOE for the mixed sample is indicative of intramolecular base pairing. (*C*) Spectra obtained for A^H and mixed A^2:A^r-labeled 5′-$L_{344\text{-}UUA}$ samples that probe the U5:AUG helix. The cross-strand NOE from A338-H2 to A112-H_1' observed for the mixed sample reveals that U5:AUG base pairing is intermolecular. (*D*) Secondary structure consistent with the combined NMR probing data, in which the TAR, PolyA, PBS, and ψ helices are formed by intramolecular base pairs (blue) and the U5:AUG and DIS helices comprise intermolecular base pairs (pink).

helix using an lrAID modified leader, in which three sequential base pairs ([5′-^{109}CCG^{112}]:[5′-^{337}UGG^{339}]) are conservatively substituted by a triplet of adenosine-containing base pairs ([5′-UUA]:[5′-UAA]) that gives rise to a readily detectable adenosine-H2 signal. This approach was previously used to probe for U5:AUG interactions in the context of the HIV-1 5′-L (48) and was used again to probe for analogous interactions in the dimeric HIV-2_{ROD} 5′-leader (46). The [5′-UUA]:[5′-UAA] triplet is native in the TAR hairpin; therefore, an additional mutation (A46G) was incorporated to avoid spectral overlap with the TAR signal. As a control, an lrAID-modified HIV-1 5′-L_{334} sample (5′-$L_{334\text{-}UUA}$) was prepared with only adenosines protonated (A^H). The 2D NOESY spectrum obtained for this sample exhibited sequential A338-H2 to A339-H_1' and -H2 and cross-strand A338-H2 to A112-H_1' NOEs, as expected (48). Importantly, the A338-H2 to A112-H_1' NOE was also observed in a mixed A^2:A^r 5′-$L_{344\text{-}UUA}$ sample, demonstrating that the U5:AUG helix comprises intermolecular base pairs (Fig. 3*C*). Collectively, these data provide a detailed picture of the intermolecular interface of the HIV-1 5′-L RNA (Fig. 3*D*).

Extended Dimer Interface Forms Rapidly. The above gel-based assays indicated that HIV-1 5′-L RNA dimerization occurs rapidly and reaches equilibrium at ∼30 min. However, the 2D NMR experiments that probed for intermolecular interactions required at least 72 h (but typically 144 h) of data collection, and we thus could not rule out the possibility that 5′-L might form a kissing dimer that slowly converts to the extended interface structure.

To probe the nature of the intermolecular interface at shorter time intervals, we developed a 1D ^{1}H NMR strategy that involves mixing two mutant 5′-L RNAs that differ in both sequence and ^{2}H-isotopic labeling (Fig. 4*A*). The strategy is based on the fact that the central adenosine in [UUA]:[UAA] base-pair triplets gives rise to an A-H2 signal at 6.48 ppm (Fig. 4*B*), whereas the same adenosine in the closely related [UUG]:[UAA] triplet gives rise to an A-H2 signal at 6.7 ppm (Fig. 4*C*). The U5:AUG helix was selected as the site of mutagenesis and NMR probing, as it is involved in the intermolecular interface yet is well removed from DIS, the site where dimerization is believed to nucleate. In both 5′-L_{344} RNAs, the native ^{337}UGG triplet was substituted by UAA. One RNA contained an additional ^{110}UUG substitution (5′-$L_{344\text{-}UUG}$) and in the other a ^{110}UUA (5′-$L_{344\text{-}UUA}$) substitution (Fig. 4*A*). The latter RNA corresponds precisely to the lrAID 5′-L_{344} construct used in the 2D NOESY experiments to probe for intermolecular U5:AUG base pairing. The 5′-L_{UUG} construct was prepared to be fully protonated ($A^HG^HC^HU^H$-5′-$L_{344\text{-}UUG}$), and the 5′-$L_{344\text{-}UUA}$ sample contained fully protonated G, C, and U residues and perdeuterated adenosines ($G^HC^HU^H$-5′-$L_{344\text{-}UUA}$).

The isolated $G^HC^HU^H$-5′-$L_{344\text{-}UUA}$ construct did not give rise to ^{1}H NMR signals in the range of 6.3–6.9 ppm, as expected, as all of the adenosines were deuterated (Fig. 4*D*). The isolated $A^HG^HC^HU^H$-5′-$L_{344\text{-}UUG}$ RNA gave rise to a number of adenosine H2 NMR signals in the range of 6.7–6.9 ppm, including a signal at 6.7 ppm diagnostic of a UUG:UAA triplet (Fig. 4*E*). Note that neither isolated sample gave rise to a signal at 6.48 ppm diagnostic of the UUA:UAA triplet because (*i*) the 5′-$L_{344\text{-}UUA}$ sample lacks protonated adenosines and (*ii*) the 5′-$L_{344\text{-}UUG}$ construct can only form [UUG]:[UAA] base-pair triplets (Fig. 4*A*). However, after mixing the two samples and acquiring a 1D ^{1}H NMR spectrum (20 min total time), a well-resolved signal at 6.48 ppm was observed indicating that (*i*) heterodimerization has occurred and (*ii*) the U5:AUG helix is comprised of strands from both RNAs (Fig. 4 *A* and *F*). The intensity of this signal did not increase significantly in subsequently obtained NMR spectra (Fig. 4*G*). Thus, the intermolecular U5:AUG interface forms on the same time scale as that of overall RNA dimerization.

Discussion

The importance of the DIS element in promoting HIV genome dimerization has been well documented (12–19, 55), and there is

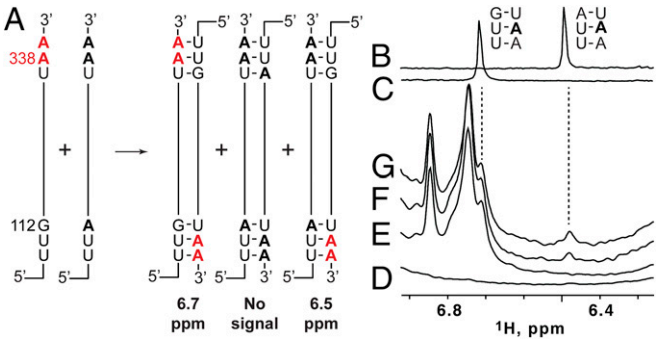

Fig. 4. Extended dimer forms rapidly, as measured using a ^{2}H-edited 1D NMR approach. (*A*) Two RNAs with different sequences and ^{2}H-labeling schemes (5′-$A^HG^HC^HU^H$-$L_{344\text{-}UUG}$ and $G^HC^HU^H$-5′-$L_{344\text{-}UUA}$) are mixed under low-salt conditions that favor the monomer. Dimerization, induced by addition of PI buffer, results in statistical mixture of [$A^HG^HC^HU^H$-$L_{344\text{-}UUG}$]$_2$ and [$G^HC^HU^H$-$L_{344\text{-}UUA}$]$_2$ homodimers and a [$G^HC^HU^H$-$L_{344\text{-}UUA}$]:[$A^HG^HC^HU^H$-$L_{344\text{-}UUG}$] heterodimer. The [$G^HC^HU^H$-$L_{344\text{-}UUA}$]:[$A^HG^HC^HU^H$-$L_{344\text{-}UUG}$] heterodimer is the only species with an adenosine-protonated [UUA]:[UAA] base-pair triplet that could give rise to a detectable adenosine-H2 NMR signal at 6.48 ppm. (*B* and *C*) Regions of ^{1}H NMR spectra obtained for control oligo-RNAs showing that the H2 signal of the central adenosine in [UUG]:[UAA] and [UUA]:[UAA] triplets occurs at significantly different frequencies (6.72 and 6.48 ppm, respectively). (*D* and *E*) Regions of 1D ^{1}H NMR spectra obtained for dimeric forms of the isolated $G^HC^HU^H$ 5′-$L_{344\text{-}UUA}$ (*D*) and $A^HG^HC^HU^H$-$L_{344\text{-}UUG}$ (*E*) RNAs (samples incubated in PI buffer for 1 h at 37 °C). (*F* and *G*) Portions of ^{1}H NMR spectra obtained for a mixed $G^HC^HU^H$ 5′-$L_{344\text{-}UUA}$ + $A^HG^HC^HU^H$-$L_{344\text{-}UUG}$ sample upon addition of dimer-promoting PI buffer. The 6.48 ppm signal appears within 20 min (*F*), with intensity similar to that observed after 12 h of incubation (*G*).

now strong genetic evidence that the central six-nucleotide palindrome is the primary determinant of RNA partner selection (7, 56). Studies with short oligoribonucleotides indicate that DIS is capable of adopting both kissing (26–30) and extended duplex (29, 31, 32) conformers, and both types of interfaces have been proposed to play roles in replication. Although the secondary structures of recombinant HIV-1 5′-leader RNAs have been extensively studied, the nature of the dimer interface has not been probed. The present studies reveal that the HIV-$1_{NL4\text{-}3}$ 5′-leader adopts an extended duplex DIS conformation and that formation of the extended dimer interface occurs on a similar time scale as that of overall RNA dimerization. Furthermore, formation of the extended duplex DIS structure in the context of the intact 5′-leader does not require the presence of nucleocapsid or other RNA chaperones and occurs without heating above physiological temperatures (37 °C).

Previous studies with truncated HIV-$1_{NL4\text{-}3}$ 5′-leader RNAs and antisense oligonucleotides indicate that SD also plays a role in dimerization (42), and this finding is consistent with the intermolecular base pairing of SD residues shown in Fig. 1*A* (57). Unfortunately, we were unable to directly detect NMR signals for these or other residues of the central helix in the intact dimeric leader due to overlap with numerous and sometimes broad adenosine-H2 signals associated with the upper PBS loop. It was, however, possible to directly detect cross-helix NOEs involving A124 in a mutant 5′-L_{344} construct that lacked the upper PBS loop (57) as well as in a 155-nucleotide "core encapsidation signal" RNA (57). Thus, the only segments of the native dimeric leader that we were unable to probe by this NMR approach, due to issues of signal overlap, were the central helix of the tandem three-way junction ([U118-G123]:[U294-A300]) and the lower region of the extended DIS stem ([A227-A242]:[G278-A292]).

Other regions of the HIV-1 5′-leader have also been proposed to play roles in dimerization. For example, deletion of the TAR hairpin has been shown to inhibit dimerization of the HIV-1_{HXB2} 5′-leader (43). Consistent with this hypothesis, antisense oligonucleotides that target the apical loop of TAR have been shown to promote dimerization (58). However, mutagenesis studies indicated

that it is the upper Tat-binding bulge and not the apical loop residues that promote dimerization (59). Other studies indicated that TAR is not essential for genome dimerization and packaging and that mutations that disrupt base pairing in the stem of TAR can lead to aberrant dimerization and packaging (which was attributed to misfolding) (60). On the other hand, chemical probing studies suggested that the base of the TAR hairpin, rather than the Tat-binding bulge or apical loop, may be involved in regulating a monomer–dimer structural switch (61). The present studies provide clear evidence that residues of TAR adopt an intramolecular hairpin structure in the context of the intact 5′-leader. However, we cannot rule out the possibility that residues in the loops or lower base of TAR, which we did not observe in our NMR experiments, might participate in dimer-stabilizing interactions with undefined downstream elements (59).

Our data are incompatible with an early proposal that residues of PolyA form an intermolecular duplex (62). The binding of $tRNA^{Lys3}$ (and an 18-nucleotide DNA analog) to the PBS was also reported to promote dimerization, suggesting that the PBS could play a direct or indirect role in dimerization (63). The present studies show that the lower stem of PBS comprises intramolecular base pairs in the dimeric 5′-leader. Although we cannot rule out the possibility that other regions of the PBS make intermolecular contacts in the dimer, the fact that 5′-leader–containing vector RNAs that lack the tRNA-binding site within PBS are packaged with avidity similar to that of the intact 5′-leader (53, 57) is consistent with studies indicating that the PBS does not contribute to dimerization (53).

Studies also indicate that U5:AUG interactions promote genome dimerization. Helper RNAs are unable to rescue packaging of RNAs with AUG mutations (44), and this is likely due to a defect in U5:AUG-dependent exposure of the DIS (48). Mutations in U5 and/or AUG that would disrupt U5:AUG pairing inhibited dimerization (43, 64), and it was suggested that this is most likely due to an indirect effect and not intermolecular base pairing (43). Consistent with this hypothesis, addition of short oligonucleotides complementary to the U5 region of the HIV-1_{NL4-3} 5′-leader promoted 5′-leader dimerization, and this was attributed to displacement of monomer-stabilizing intramolecular U5:DIS base pairing (48). However, the present study shows that, at equilibrium, U5:AUG base pairing is intermolecular in the dimeric 5′-leader.

The collective data are consistent with a dimerization model in which the DIS is originally sequestered by base pairing with U5 in the monomeric form of the 5′-leader (Fig. 5*A*). Subsequent intramolecular base pairing of the downstream AUG element with U5 would displace and expose the DIS, thus leading to a monomeric RNA that is poised to form DIS-mediated kissing interactions with another RNA molecule (Fig. 5*B*). The next species formed along this dimerization trajectory would be the labile kissing dimer species, in which the U5:AUG helix comprises intramolecular base pairs and only the DIS:DIS interface contains intermolecular base pairs (Fig. 5*B*). Finally, the labile kissing dimer undergoes rearrangement to the thermodynamically nonlabile extended dimer that includes intermolecular DIS and U5:AUG helices (Fig. 5*C*). In the case of the HIV-1_{NL4-3} leader, the DIS-exposed monomer and kissing dimer species are undetected by the methods used here and therefore, if they exist, must only be transiently formed along the dimerization pathway. We now plan to apply this approach to the 5′-leader RNAs of retroviruses that form labile dimers to determine if they adopt kissing or "labile-extended" species.

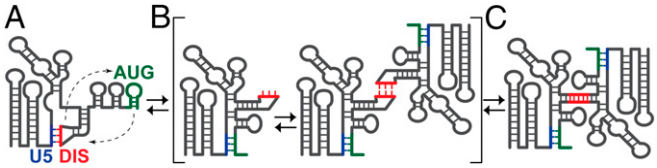

Fig. 5. Mechanism of HIV-1 5′-leader dimerization. (*A*) In the monomeric conformer, the DIS hairpin is sequestered by intramolecular base pairing with U5 (48). (*B*) AUG base pairs with U5, displacing the DIS and enabling formation of a labile kissing intermolecular interface, which rapidly transitions to the extended dimer conformer (*C*). NMR signatures for the kissing species were not observed in the present studies, indicating that it represents a transient species along the dimerization pathway.

The present studies illustrate both the potential and weaknesses of the ^{2}H-edited adenosine-detected ^{1}H NMR method for differentiating inter- and intramolecular interactions in relatively large RNAs. The sensitivity and chemical shift dispersion was sufficient for detection of 28 out of 85 total adenosine residues, which enabled probing of all of the known helices of the native, dimeric 5′-leader except the lower helix of the extended DIS stem, the central helix of the tandem three-way junction, and the native U5:AUG helix. The majority of unassigned adenosines (50%) reside within the upper PBS loop, a region for which we do not have appropriate control oligo-RNAs to aid in assignment and which may be conformationally heterogeneous in the absence of the primer tRNA binding partner. The lrAID approach enabled probing of U5:AUG; however, we were unable to use this method to probe the short, central helix of the tandem three-way junction due to apparent RNA misfolding caused by the lrAID mutations. It thus appears that the lrAID approach may only be readily applicable if the helical elements being probed are sufficiently stable to tolerate incorporation of a short stretch of sequential adenosines. Finally, the combination of mixed ^{2}H-labeling with mixed lrAID mutations provides a powerful approach for temporal monitoring of intermolecular helix formation in dimeric RNAs as large as 230 kDa.

Materials and Methods

RNAs were synthesized by in vitro transcription using T7 RNA polymerase with slight modifications to published procedures (57) (plasmids and primers are summarized in Tables S1 and S2, respectively). NMR data were collected with a Bruker AVANCE spectrometer (800 MHz, ^{1}H, 37 °C). For studies involving mixed RNAs with different labeling schemes, RNAs were transcribed and purified independently. After elution from the denaturing gel, equal molar amounts were pooled, washed, and lyophilized before addition of dimer-promoting PI buffer. Samples for time-dependent NMR studies were prepared as follows: Equal quantities (~17 nmol) of $G^{H}C^{H}U^{H}$-5′-L_{UUA} and $A^{H}G^{H}C^{H}U^{H}$-5′-L_{UUG} RNA were lyophilized from water and independently reconstituted in 170 μL D_2O. The RNAs were pooled together, added to a tube containing lyophilized NMR buffer, and immediately placed into the spectrometer. A series of identical 1D proton experiments were queued to run sequentially. For details, see *SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank the Howard Hughes Medical Institute (HHMI) staff at University of Maryland Baltimore County for technical assistance. This research was supported by National Institute of General Medical Sciences (NIGMS) Grants R01 GM42561 and P50 GM 103297 (to M.F.S.). S.M. and J.S. were supported by NIGMS Grant MARC U*STAR 2T34 GM008663 for enhancing minority access to research careers and by an HHMI undergraduate education grant.

1. Coffin JM, Hughes SH, Varmus HE (1997) *Retroviruses* (Cold Spring Harbor Lab Press, Plainview, NY).
2. Coffin JM (1979) Structure, replication, and recombination of retrovirus genomes: Some unifying hypotheses. *J Gen Virol* 42(1):1–26.
3. Hu W-S, Temin HM (1990) Genetic consequences of packaging two RNA genomes in one retroviral particle: Pseudodiploidy and high rate of genetic recombination. *Proc Natl Acad Sci USA* 87(4):1556–1560.
4. Hu WS, Temin HM (1990) Retroviral recombination and reverse transcription. *Science* 250(4985):1227–1233.
5. Nora T, et al. (2007) Contribution of recombination to the evolution of human immunodeficiency viruses expressing resistance to antiretroviral treatment. *J Virol* 81(14):7620–7628.
6. Onafuwa-Nuga A, Telesnitsky A (2009) The remarkable frequency of human immunodeficiency virus type 1 genetic recombination. *Microbiol Mol Biol Rev* 73(3): 451–480.

7. Moore MD, et al. (2007) Dimer initiation signal of human immunodeficiency virus type 1: Its role in partner selection during RNA copackaging and its effects on recombination. *J Virol* 81(8):4002–4011.
8. Nikolaitchik OA, et al. (2013) Dimeric RNA recognition regulates HIV-1 genome packaging. *PLoS Pathog* 9(3):e1003249.
9. D'Souza V, Summers MF (2005) How retroviruses select their genomes. *Nat Rev Microbiol* 3(8):643–655.
10. Lu K, Heng X, Summers MF (2011) Structural determinants and mechanism of HIV-1 genome packaging. *J Mol Biol* 410(4):609–633.
11. Kuzembayeva M, Dilley K, Sardo L, Hu W-S (2014) Life of psi: How full-length HIV-1 RNAs become packaged genomes in the viral particles. *Virology* 454-455:362–370.
12. Kim H-J, Lee K, O'Rear JJ (1994) A short sequence upstream of the 5′ major splice site is important for encapsidation of HIV-1 genomic RNA. *Virology* 198(1):336–340.
13. Skripkin E, Paillart JC, Marquet R, Ehresmann B, Ehresmann C (1994) Identification of the primary site of the human immunodeficiency virus type 1 RNA dimerization in vitro. *Proc Natl Acad Sci USA* 91(11):4945–4949.
14. Berkhout B, Vastenhouw NL, Klasens BI, Huthoff H (2001) Structural features in the HIV-1 repeat region facilitate strand transfer during reverse transcription. *RNA* 7(8):1097–1114.
15. Clever JL, Wong ML, Parslow TG (1996) Requirements for kissing-loop-mediated dimerization of human immunodeficiency virus RNA. *J Virol* 70(9):5902–5908.
16. Paillart J-C, Skripkin E, Ehresmann B, Ehresmann C, Marquet R (1996) A loop-loop "kissing" complex is the essential part of the dimer linkage of genomic HIV-1 RNA. *Proc Natl Acad Sci USA* 93(11):5572–5577.
17. Paillart JC, et al. (1996) A dual role of the putative RNA dimerization initiation site of human immunodeficiency virus type 1 in genomic RNA packaging and proviral DNA synthesis. *J Virol* 70(12):8348–8354.
18. McBride MS, Panganiban AT (1996) The human immunodeficiency virus type 1 encapsidation site is a multipartite RNA element composed of functional hairpin structures. *J Virol* 70(5):2963–2973.
19. Berkhout B, van Wamel JL (1996) Role of the DIS hairpin in replication of human immunodeficiency virus type 1. *J Virol* 70(10):6723–6732.
20. Foley B, et al. (2013) *HIV Sequence Compendium* (Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM).
21. Miele G, Mouland A, Harrison GP, Cohen E, Lever AM (1996) 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(2):944–951.
22. Greatorex J (2004) The retroviral RNA dimer linkage: Different structures may reflect different roles. *Retrovirology* 1:22.
23. Abbink TEM, Berkhout B (2008) RNA structure modulates splicing efficiency at the human immunodeficiency virus type 1 major splice donor. *J Virol* 82(6):3090–3098.
24. Jablonski JA, Buratti E, Stuani C, Caputi M (2008) The secondary structure of the human immunodeficiency virus type 1 transcript modulates viral splicing and infectivity. *J Virol* 82(16):8038–8050.
25. Abbink TE, Ooms M, Haasnoot PC, Berkhout B (2005) The HIV-1 leader RNA conformational switch regulates RNA dimerization but does not regulate mRNA translation. *Biochemistry* 44(25):9058–9066.
26. Mujeeb A, Clever JL, Billeci TM, James TL, Parslow TG (1998) Structure of the dimer initiation complex of HIV-1 genomic RNA. *Nat Struct Biol* 5(6):432–436.
27. Dardel F, Marquet R, Ehresmann C, Ehresmann B, Blanquet S (1998) Solution studies of the dimerization initiation site of HIV-1 genomic RNA. *Nucleic Acids Res* 26(15):3567–3571.
28. Ennifar E, Walter P, Ehresmann B, Ehresmann C, Dumas P (2001) Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site. *Nat Struct Biol* 8(12):1064–1068.
29. Baba S, et al. (2005) Solution RNA structures of the HIV-1 dimerization initiation site in the kissing-loop and extended-duplex dimers. *J Biochem* 138(5):583–592.
30. Kieken F, Paquet F, Brulé F, Paoletti J, Lancelot G (2006) A new NMR solution structure of the SL1 HIV-1Lai loop-loop dimer. *Nucleic Acids Res* 34(1):343–352.
31. Mujeeb A, Parslow TG, Zarrinpar A, Das C, James TL (1999) NMR structure of the mature dimer initiation complex of HIV-1 genomic RNA. *FEBS Lett* 458(3):387–392.
32. Ulyanov NB, et al. (2006) NMR structure of the full-length linear dimer of stem-loop-1 RNA in the HIV-1 dimer initiation site. *J Biol Chem* 281(23):16168–16177.
33. Takahashi K-I, et al. (2000) Structural requirement for the two-step dimerization of human immunodeficiency virus type 1 genome. *RNA* 6(1):96–102.
34. Mihailescu M-R, Marino JP (2004) A proton-coupled dynamic conformational switch in the HIV-1 dimerization initiation site kissing complex. *Proc Natl Acad Sci USA* 101(5):1189–1194.
35. Sun X, Zhang Q, Al-Hashimi HM (2007) Resolving fast and slow motions in the internal loop containing stem-loop 1 of HIV-1 that are modulated by Mg2+ binding: Role in the kissing-duplex structural transition. *Nucleic Acids Res* 35(5):1698–1713.
36. Fu W, Gorelick RJ, Rein A (1994) Characterization of human immunodeficiency virus type 1 dimeric RNA from wild-type and protease-defective virions. *J Virol* 68(8):5013–5018.
37. Song R, Kafaie J, Yang L, Laughrea M (2007) HIV-1 viral RNA is selected in the form of monomers that dimerize in a three-step protease-dependent process; the DIS of stem-loop 1 initiates viral RNA dimerization. *J Mol Biol* 371(4):1084–1098.
38. Paillart JC, Marquet R, Skripkin E, Ehresmann B, Ehresmann C (1994) Mutational analysis of the bipartite dimer linkage structure of human immunodeficiency virus type 1 genomic RNA. *J Biol Chem* 269(44):27486–27493.
39. Muriaux D, Girard P-M, Bonnet-Mathonière B, Paoletti J (1995) Dimerization of HIV-1Lai RNA at low ionic strength. An autocomplementary sequence in the 5′ leader region is evidenced by an antisense oligonucleotide. *J Biol Chem* 270(14):8209–8216.
40. Muriaux D, Fossé P, Paoletti J (1996) A kissing complex together with a stable dimer is involved in the HIV-1Lai RNA dimerization process in vitro. *Biochemistry* 35(15):5075–5082.
41. Muriaux D, De Rocquigny H, Roques BP, Paoletti J (1996) NCp7 activates HIV-1Lai RNA dimerization by converting a transient loop-loop complex into a stable dimer. *J Biol Chem* 271(52):33686–33692.
42. Deforges J, Chamond N, Sargueil B (2012) Structural investigation of HIV-1 genomic RNA dimerization process reveals a role for the Major Splice-site Donor stem loop. *Biochimie* 94(7):1481–1489.
43. Song R, Kafaie J, Laughrea M (2008) Role of the 5′ TAR stem-loop and the U5-AUG duplex in dimerization of HIV-1 genomic RNA. *Biochemistry* 47(10):3283–3293.
44. Nikolaitchik O, Rhodes TD, Ott D, Hu W-S (2006) Effects of mutations in the human immunodeficiency virus type 1 Gag gene on RNA packaging and recombination. *J Virol* 80(10):4691–4697.
45. Takahashi K, et al. (2000) NMR analysis of intra- and inter-molecular stems in the dimerization initiation site of the HIV-1 genome. *J Biochem* 127(4):681–686.
46. Tran T, et al. (2015) Conserved determinants of lentiviral genome dimerization. *Retrovirology* 12:83.
47. Helmling C, et al. (2015) Rapid NMR screening of RNA secondary structure and binding. *J Biomol NMR* 63(1):67–76.
48. Lu K, et al. (2011) NMR detection of structures in the HIV-1 5′-leader RNA that regulate genome packaging. *Science* 334(6053):242–245.
49. Laughrea M, Jetté L (1997) HIV-1 genome dimerization: Kissing-loop hairpin dictates whether nucleotides downstream of the 5′ splice junction contribute to loose and tight dimerization of human immunodeficiency virus RNA. *Biochemistry* 36(31):9501–9508.
50. Laughrea M, Jetté L (1996) Kissing-loop model of HIV-1 genome dimerization: HIV-1 RNAs can assume alternative dimeric forms, and all sequences upstream or downstream of hairpin 248-271 are dispensable for dimer formation. *Biochemistry* 35(5):1589–1598.
51. Marquet R, Paillart J-C, Skripkin E, Ehresmann C, Ehresmann B (1994) Dimerization of human immunodeficiency virus type 1 RNA involves sequences located upstream of the splice donor site. *Nucleic Acids Res* 22(2):145–151.
52. Abbink TEM, Berkhout B (2003) A novel long distance base-pairing interaction in human immunodeficiency virus type 1 RNA occludes the Gag start codon. *J Biol Chem* 278(13):11601–11611.
53. Heng X, et al. (2012) Identification of a minimal region of the HIV-1 5′-leader required for RNA dimerization, NC binding, and packaging. *J Mol Biol* 417(3):224–239.
54. Milligan JF, Uhlenbeck OC (1989) Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol* 180:51–62.
55. Laughrea M, Shen N, Jetté L, Wainberg MA (1999) Variant effects of non-native kissing-loop hairpin palindromes on HIV replication and HIV RNA dimerization: Role of stem-loop B in HIV replication and HIV RNA dimerization. *Biochemistry* 38(1):226–234.
56. Moore MD, et al. (2009) Probing the HIV-1 genomic RNA trafficking pathway and dimerization by genetic recombination and single virion analyses. *PLoS Pathog* 5(10):e1000627.
57. Keane SC, et al. (2015) RNA structure. Structure of the HIV-1 RNA packaging signal. *Science* 348(6237):917–921.
58. Reyes-Darias JA, Sánchez-Luque FJ, Berzal-Herranz A (2012) HIV RNA dimerisation interference by antisense oligonucleotides targeted to the 5′ UTR structural elements. *Virus Res* 169(1):63–71.
59. Jalalirad M, Saadatmand J, Laughrea M (2012) Dominant role of the 5′ TAR bulge in dimerization of HIV-1 genomic RNA, but no evidence of TAR-TAR kissing during in vivo virus assembly. *Biochemistry* 51(18):3744–3758.
60. Das AT, Vrolijk MM, Harwig A, Berkhout B (2012) Opening of the TAR hairpin in the HIV-1 genome causes aberrant RNA dimerization and packaging. *Retrovirology* 9:59.
61. Kenyon JC, Prestwood LJ, Le Grice SF, Lever AM (2013) In-gel probing of individual RNA conformers within a mixed population reveals a dimerization structural switch in the HIV-1 leader. *Nucleic Acids Res* 41(18):e174.
62. Höglund S, Öhagen A, Goncalves J, Panganiban AT, Gabuzda D (1997) Ultrastructure of HIV-1 genomic RNA. *Virology* 233(2):271–279.
63. Seif E, Niu M, Kleiman L (2013) Annealing to sequences within the primer binding site loop promotes an HIV-1 RNA conformation favoring RNA dimerization and packaging. *RNA* 19(10):1384–1393.
64. Russell RS, Hu J, Laughrea M, Wainberg MA, Liang C (2002) Deficient dimerization of human immunodeficiency virus type 1 RNA caused by mutations of the u5 RNA sequences. *Virology* 303(1):152–163.
65. Kao C, Zheng M, Rüdisser S (1999) A simple and efficient method to reduce nontemplated nucleotide addition at the 3 terminus of RNAs transcribed by T7 RNA polymerase. *RNA* 5(9):1268–1272.
66. Huang X, Yu P, LeProust E, Gao X (1997) An efficient and economic site-specific deuteration strategy for NMR studies of homologous oligonucleotide repeat sequences. *Nucleic Acids Res* 25(23):4758–4763.
67. Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 9(7):671–675.
68. Norris M, Fetler B, Marchant J, Johnson BA (2016) NMRFx Processor: A cross-platform NMR data processing program. *J Biomol NMR* 65(3-4):205–216.
69. Johnson BA (2004) Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol Biol* 278:313–352.
70. Barton S, Heng X, Johnson BA, Summers MF (2013) Database proton NMR chemical shifts for RNA signal assignment and validation. *J Biomol NMR* 55(1):33–46.
71. Brown JD, Summers MF, Johnson BA (2015) Prediction of hydrogen and carbon chemical shifts from RNA using database mining and support vector regression. *J Biomol NMR* 63(1):39–52.