A 19-Nucleotide Sequence Upstream of the 5′ Major Splice Donor Is Part of the Dimerization Domain of Human Immunodeficiency Virus 1 Genomic RNA[†,‡]

Michael Laughrea* and Louis Jetté

McGill AIDS Centre, Lady Davis Institute for Medical Research, Sir Mortimer B. Davis−Jewish General Hospital, and Department of Medicine, McGill University, Montreal, Quebec, Canada H3T 1E2

Received June 22, 1994; Revised Manuscript Received August 24, 1994[⊗]

ABSTRACT: The genome of all retroviruses, including human immunodeficiency virus type 1 (HIV-1), consists of two identical RNAs noncovalently linked near their 5′ end. Dimerization of genomic RNA is thought to modulate several steps in the retroviral life cycle, such as recombination, translation, and encapsidation. We report the results of experiments designed to identify the 5′ and 3′ boundaries of the dimerization domain of the HIV-1 genome: (1) An HIV-1 RNA starting at nucleotide 252 or at other downstream positions (four tested) does not dimerize despite the inclusion of the whole of a previously proposed dimerization domain (nucleotides 295−401); (2) an RNA starting between nucleotides 242 and 249 (five positions tested) dimerizes to a variable extent depending on the starting position; (3) an RNA starting at nucleotide 233 or at other upstream positions (five tested) is fully or >80% dimeric; (4) an RNA starting at nucleotide 1 but lacking the 233−251 or the 242−251 region is, respectively, fully monomeric or about 50% monomeric; (5) the 343−401 region contains two strings of G's (GGGGG367 and GGG384) that had been postulated to promote genome dimerization through the formation of guanine quartets. We have deleted the 379−401, 358−401, and 343−401 regions from otherwise dimeric RNAs without changing their ability to dimerize. We reach three conclusions: (1) a dimerization signal exists upstream of the major 5′ splice donor (nucleotide 290); (2) the previously proposed downstream dimerization domain is insufficient to promote dimerization and has a 3′ half that is not necessary to obtain fully dimeric RNAs; (3) the 5′ boundary of the HIV-1 dimerization domain is located somewhere between nucleotides 233 and 242, and the 3′ boundary is located no farther than at nucleotide 342, making it possible that the 5′ and 3′ boundaries of the HIV-1 dimerization domain are both located within the leader sequence. We speculate that the 248−270 or 233−285 region forms a hairpin that is the core dimerization domain of HIV-1 RNA.

Retroviruses are a family of RNA viruses that replicate through a DNA intermediate, and human immunodeficiency virus type 1 (HIV-1[1]) is a retrovirus acknowledged as a cause of acquired immunodeficiency syndrome (AIDS). The genome of all retroviruses, including HIV-1, consists of two identical single-stranded RNA molecules that are noncovalently linked near their 5′ ends [Marquet et al. (1994) and references therein]. Dimerization of this diploid genome might play an important role in several steps of the retroviral life cycle, such as (i) translation of the genome into viral proteins (Bieth et al., 1990; Baudin et al., 1993), (ii) preferential packaging (or encapsidation) of two genomic RNAs within the budding viral shell (or capsid) against a vast excess of cellular RNAs (Darlix et al., 1990), (iii) recombination between the two homologous genomic RNAs (Temin, 1993; Jones et al., 1994; Hu et al., 1993) in order to increase the rate of viral evolution and to bypass occasional breaks in one of the genomic RNAs (Clavel et al., 1989), and (iv) stabilization of the genome against degradation. Dimerization of genomic RNA could constitute a suitable anti-AIDS therapeutic target.

Evidence for the existence of a dimerization signal downstream of the 5′ major splice donor site (SD) has been described in at least six different animal retroviruses (Bieth et al., 1990; Prats et al., 1990; Darlix et al., 1992; Katoh et al., 1991, 1993; Jones et al., 1993; Torrent et al., 1994) and in human retroviruses such as HIV-1 MAL (Darlix et al., 1990; Marquet et al., 1991) and HIV-1 NL43 (Sundquist & Heaphy, 1993). Some of these downstream sequences (Marquet et al., 1991; Sundquist & Heaphy, 1993; Awang & Sen, 1993; Prats et al., 1990; Darlix et al., 1992; Torrent et al., 1994) have been shown to dimerize to a small or moderate extent in the absence of added proteins. The mechanism of retroviral genome dimerization is unclear, as Marquet et al. (1991) suggested a role for PuGGAPuA motifs that are, however, absent from dimerized HIV-2 RNA fragments (Berkhout et al, 1993) and Awang and Sen (1993) implicated, in an HIV-1 MAL RNA transcript, a highly conserved GGG stretch that has been shown by Sundquist and Heaphy (1993) to be entirely dispensable in a similarly sized RNA transcript from HIV-1 NL43.

The putative downstream dimerization signal of HIV-1 has occasionally been called dimer linkage structure (Darlix et al., 1990), even though linkage of any of the putative

[†] This work was funded by a grant from the Medical Research Council of Canada to M.L.

[‡] This paper is dedicated to the memory of Howard M. Temin (1934−1994).

* Author to whom correspondence should be addressed at the Lady Davis Institute for Medical Research, 3755 Cote-Ste-Catherine Rd., Montreal, Quebec, Canada H3T 1E2. Telephone: (514) 340-8260. Fax: (514) 340-7502.

[⊗] Abstract published in *Advance ACS Abstracts,* October 1, 1994.

[1] Abbreviations: HIV-1, human immunodeficiency virus type 1; AIDS, acquired immunodeficienty syndrome; SD, 5′ major splice donor site (directs splicing between G289 and sequences located farther downstream); DLS, name of a sequence starting very near the SD (typically at nucleotide 295) and ending at nucleotide 401 (no other connotation is to be associated with this name in the context of the present paper); PBS, primer binding site.

0006-2960/94/0433-13464$04.50/0

sequences or of only such sequences has not been demonstrated by electron microscopy or other visualization techniques. To be more consonant with our degree of ignorance, we variously designate the sequence of Darlix et al. (1990) by the more agnostic names of DLS or downstream dimerization signal (downstream relative to the SD). (In this paper, DLS is the name of a sequence, not an acronym.) This makes it clear that none of the individual nucleotides involved in the physical linkage of the two genomic RNAs have yet been identified and prepares the grounds for what follows.

Sequences upstream of the SD have been shown to play a role in the dimerization of Rous sarcoma virus (RSV) and HIV-2 and HIV-1 MAL RNA transcripts, with the caveats that the first 10−20 nucleotides of the RSV and HIV-2 RNA constructs were nongenomic (Bieth et al., 1990; Berkhout et al., 1993) and that the 311 nucleotide long HIV-1 MAL RNA transcript included six nucleotides downstream of the SD (Marquet et al., 1994). In HIV-1, this putative upstream dimerization signal has a strong impact on the structure and dimerization properties of the DLS (Marquet et al., 1994).

The primer binding site (PBS) and SD of HIV-1 RNA are separated by 90 nucleotides extending from U200 to G289 in strain LAI (previously named LAVbru). These residues represent two-thirds of the leader sequence and their function is ill-characterized. What is most striking phylogenetically is a very highly conserved sequence (U228−U253) located halfway between the PBS and the SD. We introduced a series of deletions in the leader sequence of HIV-1 RNA to identify the 5′ boundary of the dimerization domain of HIV-1 RNA. We found, both by nested deletion analysis and by internal deletion analysis, that a 19-nucleotide sequence located within this highly conserved stretch is essential for the *in vitro* dimerization of HIV-1 RNA transcripts ending well within the polyprotein Pr55gag coding sequence. This suggests that a short sequence entirely upstream of the SD is needed to dimerize HIV-1 RNA. We also showed that the 3′ boundary of an HIV-1 RNA transcript able to dimerize easily is located within the leader sequence, or at most at the sixth nucleotide of the Pr55gag coding sequence.

MATERIALS AND METHODS

1. Buffers. Buffer L (Marquet et al., 1991): 50 mM sodium cacodylate (pH 7.5), 40 mM KCl, and 0.1 mM $MgCl_2$. Buffer H (Marquet et al., 1991): 50 mM sodium cacodylate (pH 7.5), 300 mM KCl, and 5 mM $MgCl_2$. TBM (Laughrea & Moore, 1977; Marquet et al., 1994): 89 mM Tris, 89 mM borate, and 0.1 mM $MgCl_2$. TBK (Sundquist & Heaphy, 1993): 89 mM Tris, 89 mM borate, and 25 mM potassium acetate. TBN: 89 mM Tris, 89 mM borate, and 5 mM ammonium acetate.

2. Materials. Plasmid pSP72 is from Promega (Madison, WI). Plasmid pBENN7 is a clone of HIV-1 LAI obtained from the AIDS Research and Reference Reagent program of the NIH (Bethesda, MD). Plasmid pEA is a gift from E. J. Arts, McGill AIDS Centre. It is a pSP72 plasmid containing a *Bgl*II−*Pst*I HIV-1 DNA fragment from pBENN7 inserted in the multiple cloning site of pSP72, immediately downstream from a T7 RNA polymerase promoter. This insert codes for nucleotides 20−965 of HIV-1 RNA LAI or, more precisely, HIV-1 RNA LAIpB2 (see Discussion). Restriction enzymes were from Pharmacia (Montreal, PQ), except for *Tru*9I (Boehringer Mannheim, Montreal, PQ) and *Hga*I (New England Biolabs, Mississauga, ON). Water treated with 0.1% diethyl pyrocarbonate was used for all solutions likely to come in contact with RNA. Phagemid pBluescript KS$^-$ (Stratagene, La Jolla, CA) was a gift from Dr. Gabriel Gingras. Agarose, electrophoresis grade, was from ICN. The 0.24−9.5 kb RNA ladder (0.24, 1.4, 2.4, 4.4, 7.5, and 9.5 kb) was from Gibco BRL (Burlington, ON).

3. Plasmid Construction and Nested Deletion Mutagenesis. Unless otherwise noted, plasmids were constructed by standard techniques (Sambrook et al., 1989), and nucleotide positions are those of the HIV-1 LAI genomic RNA. Nested deletions from plasmid pEA were done using the double-stranded nested deletion kit of Pharmacia. Briefly, pEA was cut at *Bgl*II (nucleotide 19 of HIV-1 RNA LAI) and the recessed 3′ ends were filled in with thio nucleotides. The protected linear plasmid was then cut at *Afl*II (*Bfr*I) (nucleotide 63) and digested with exonuclease III at 22−22.5 °C in 77 mM Tris (pH 8.0), 5.7 mM $MgCl_2$, and 56 mM NaCl. Time points were taken every minute or so from 5 to 25 min of incubation. After S1 nuclease digestion of the single-stranded regions generated by exonuclease III, the truncated pEA plasmids were recircularized. Transformation of *Escherichia coli* JM109 by these plasmids ensued, followed by plating on agar, selection of 2−10 colonies per plate, growth of these colonies in 1.5 mL cultures, and minipreparations of plasmid DNA using the alkali method. Each plasmid preparation was cut with *Hpa*I (51 nucleotides upstream of the former *Bgl*II site) and *Acc*I (nucleotide 508 of HIV-1 RNA LAI), and each resulting fragment was sized on a 3% agarose gel. Molecular weight markers were the *Acc*I−*Pst*I (459 bp) and *Bgl*II−*Pst*I (944 bp) fragments from pEA. We wished the *Hpa*I−*Acc*I fragments to vary in size from 150 to 490 bp. Fragments of the appropriate size were identified, and the parent plasmids were sequenced on 6% denaturing acrylamide gels to precisely identify the size of the deletion. For sequencing, procedure C (standard annealing of primer to double-stranded template) of the Pharmacia T7 DNA polymerase sequencing kit was used.

4. Site-Directed Mutagenesis. Three HIV-1 DNA fragments were subcloned into a phagemid environment, generating phagemids pBSEA, pBS69-2, and pBS165-1. To construct pBSEA, pBluescript KS$^-$ was cleaved with *Sac*I in its multiple cloning site, blunt-ended with mung bean nuclease, and cleaved with *Pst*I in the multiple cloning site. Plasmid pEA was cleaved with *Bgl*II, filled in with Klenow DNA polymerase, and cleaved with *Pst*I. The resulting HIV-1 DNA fragment (20−965 of HIV-1 RNA LAI) was purified by using Geneclean II (Bio101, La Jolla, CA) and inserted into pBluescript KS$^-$ processed just as described. To construct pBS69-2 and pBS165-1, advantage was taken of the fact that plasmids pLJ230 and pLJ252 happened to be cleavable by *Bgl*II between their T7 RNA polymerase promoter and the start of the HIV-1 RNA coding sequences (see Table 1). pLJ230 and pLJ252 thus were cleaved with *Bgl*II, blunt-ended with mung bean nuclease, and cleaved with *Pst*I. The resulting HIV-1 DNA fragments (233−965[2] and 253−965, respectively) were purified by using Geneclean II and inserted into pBluescript KS$^-$ processed as described

[2] The fragment was meant to start at 232, but mung bean nuclease had overdigested the *Bgl*II 3′ overhang by 1 nucleotide.

for the preparation of pBSEA. Site-directed mutagenesis of pBSEA, pBS69-2, and pBS165-1 was done using the U-DNA mutagenesis kit of Boehringer Mannheim to yield plasmids pBL, pB233 and pB253, coding for HIV-1 RNA transcripts containing only 1−3 nucleotides upstream of the HIV-1 sequence (Table 1). The structure of all mutants was verified by sequencing the mutagenized region. For sequencing purposes, plasmid purification was done using the QIAprep8 plasmid kit of Qiagen (Chatsworth, CA). In pBL, the 8 nucleotides located between the end of the T7 RNA polymerase promotor of pBSEA and the start of its HIV-1 coding sequence have been replaced by 16 HIV-1 nucleotides, so that the resulting transcript starts at position +1 of the LAI sequence. RNA transcripts from pB233 and pB253 start with pppGGG and are immediately followed by HIV-1 sequences (Table 1). To engineer the pBSEA→BL, pBS69-2→pB233, and pBS165-1→pB253 mutations, the following mutagenic oligonucleotides respectively, were used: CTC CCA GGC TCA GAT CTG GTC TAA CCA GAG AGA CCC TAT AGT GAG TCG, CCG AGT CCT GCG TCG CCC TAT AGT GAG TCG, and GCC GTG CGC GCT TCA GCC CTA TAG TGA GTC G. To construct pBLΔ233−251 and pBLΔ242−251, mutagenic primers GCC GTG CGC GCT TCAG AGA GAG CTT CTC TGGT and CCG TGC GCG CTT CAG CCT GCG TCG AGA GAG were used to delete sequences CGACG CAGGA CUCGG CUUG251 and ACUCGGCUUG251 from pBL (Table 2).

5. RNA Synthesis and Recovery. All plasmids described in this paper have a fraction of the HIV-1 genome cloned immediately (or a few nucleotides) downstream from a strong T7 RNA polymerase promoter. Prior to transcription, plasmids were usually cleaved with *Hae*III (nucleotide 401 of HIV-1 RNA LAI) or *Acc*I (nucleotide 508) or, in some experiments, with *Sau*3AI, *Tru*9I, or *Hga*I (nucleotides 378, 359, and 342). Transcription and RNA recovery were done using the MEGAscript kit of Ambion (Austin, TX). In short, after transcription had proceeded for 6 h, DNase I was added, followed 15 min later by ammonium acetate stop solution and phenol/chloroform extraction.

6. In Vitro Dimerization of HIV-1 RNA Fragments. In a standard experiment, 1−3 μg of RNA dissolved in 8 μL of water was heated for 2 min at 90 °C and chilled for 2 min on ice. Two microliters of 5 × concentrated buffer L or H was added, and the samples were incubated for 30 min at 0 °C in buffer L or at 37 °C in buffer H. The samples were cooled on ice and loaded on 3% agarose gels (Hoch & Lewallen, 1977) after the addition of 2 μL of loading buffer containing glycerol. Denatured samples were obtained by incubation at room temperature for 10 min in the presence of 64% formamide and 0.3% formaldehyde. Gels were run in the cold room using the GNA-100 submarine gel electrophoresis unit of Pharmacia, an apparatus that favors efficient cooling of the gel so that its temperature is comparable to that of the electrode buffer. The electrophoresis buffers were TBM, TBK, or TBN as indicated. After electrophoresis, the gels were stained in 2 μg/mL ethidium bromide, and the percentage of dimerization was estimated by visual inspection.

7. Chemical Probing. Dimethyl sulfate and kethoxal probing was done in buffer L as described in Laughrea and Tam (1991), except that the DMS was reacted for 10 min at a concentration of 0.5%. Primer extension was done with GCCGAACGACTTCGCG, which hybridizes to the 245−260 region of HIV-1. HIV-1 MAL, chemically probed by Baudin et al. (1993), has, relative to HIV-1 LAI, a 20-nucleotide insert between LAI nucleotides 215 and 216.

RESULTS

The 5′ Boundary: Nested Deletion Analysis. By exonuclease III digestion of appropriately cleaved pEA plasmids, we have generated 17 shortened plasmids that can be transcribed into HIV-1 RNA fragments lacking progressively more nucleotides at the 5′ end. This is shown in Figure 1; the number in the name of a plasmid refers to the position of the first HIV-1 RNA nucleotide coded by its transcript. Our goal was to define the 5′ boundary of the dimerization domain of the HIV-1 genome. Darlix et al. (1990) had suggested that the 3′ boundary of the dimerization domain of HIV-1 MAL was upstream of nucleotide 416 (402 in HIV-1 LAI). Accordingly, plasmids pLJ216−pLJ359 were linearized with *Hae*III to yield transcripts ending at nucleotide 401. After these RNAs were incubated in either low (buffer L) or high (buffer H) concentrations of KCl and $MgCl_2$, the degree of dimerization was assessed on a 3% agarose gel containing buffer TBM, buffer TBK, or buffer TBN. H and L ionic conditions were adopted because these were used by Marquet et al. (1991) to demonstrate rapid dimerization of HIV-1 RNA transcripts in the absence of added proteins. Buffer TBM was chosen because electrophoresis in Tris−borate−EDTA in the absence of Mg^{2+} yields artifactual (multiple) interactions of *E. coli* ribosomal protein S1 with the *E. coli* 30S ribosomal subunit, something never seen in TBM (Laughrea & Moore, 1977). The resolution of the HIV-1 RNA transcripts was better in 0.1 mM than in 1 mM Mg^{2+} (data not shown). It is interesting, for comparative purposes, that electrophoresis in TBM was recently adopted by Marquet et al. (1994). Buffer TBK was used to facilitate comparisons with the data of Sundquist and Heaphy (1993) and Awang and Sen (1993). Finally, some electrophoreses were done in TBN because of a report (Weiss et al., 1993) raising the possibility that HIV-1 RNAs might perhaps dimerize much better in the presence of NH_4^+ than in the presence of K^+.

Figure 1 summarizes the results as seen in electrophoresis buffer TBM. (1) The R, U5, and PBS regions (nucleotides 1−199) were not needed to obtain dimeric RNAs. (2) The first 33 nucleotides downstream from PBS were dispensable. (3) The 233−251 region, or elements of it, was essential to obtain dimeric RNAs. (4) An HIV-1 RNA starting at nucleotide 252, 258, 261, 262, or 267 did not dimerize, despite starting considerably upstream from the start sites of RNA seen by others (Marquet et al., 1991; Sundquist & Heaphy, 1993; Awang & Sen, 1993) to be moderately dimeric.

Figure 2 (electrophoresis in TBM) and Figure 3 (electrophoresis in TBK) document these results in more detail. The proportion of dimer often was not strikingly different in L or H incubation conditions or after electrophoresis in TBM, TBK, or TBN (TBN data not shown). Lanes 1−6 of Figure 2 and lanes 1−8 of Figure 3 show that RNA LJ233 or RNAs starting upstream of 233 were fully or almost fully dimeric, regardless of the electrophoretic or incubation conditions. The average from all of our experiments shows that RNA LJ233, which was as dimeric as LJ216, -230, or -232 in TBM, was 10% less dimeric than them in TBK. Lanes 17

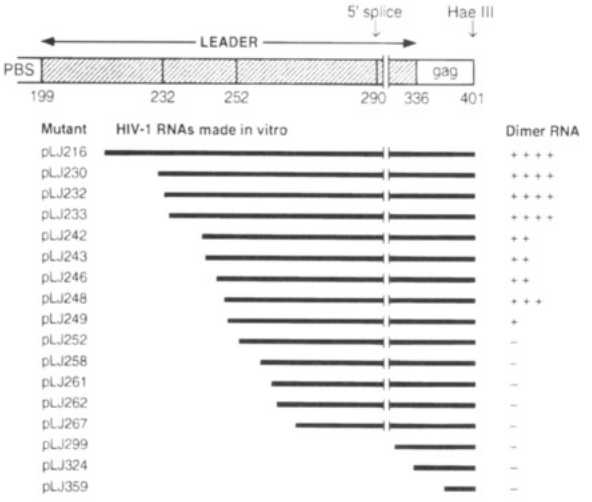

FIGURE 1: Schematic representation of most of the nest-deleted HIV-1 transcripts. The 3′ end of the $tRNA^{Lys-3}$ (primer) binding site ends at +199. The initiation codon of the Pr55gag polyprotein is at +336 (off scale), and the RNA transcripts end at position 401 (*Hae*III cut). Numbering is in reference to the genomic RNA cap site (+1). RNAs were generated *in vitro* by T7 RNA polymerase transcription of the pLJ plasmids enumerated on the left. The column of symbols headed dimer RNA indicates the average level of dimeric RNA present after *in vitro* synthesis, incubation in buffers L or H, and electrophoresis in TBM. For example, 50% dimer in buffer L and 90% dimer in buffer H would be scored as an average level of 70% dimer: −, <5% dimer; +, 10−30% dimer; ++, 40−60% dimer; +++, 60−80% dimer; ++++, ≥80% dimer.

and 18 of Figure 2 and lanes 17−22 of Figure 3 show that RNA LJ252 or RNAs starting downstream were fully monomeric. RNAs starting between 242 and 249 had more complex behavior (lanes 8−17 of Figure 2 and lanes 9−16 of Figure 3). They can be divided into two groups: (1) RNAs LJ242, LJ243, and LJ246, which were similar in property, i.e., 40−60% dimeric under all conditions, except for RNA LJ246 under H/TBM conditions which was 70% dimeric; (2) RNAs LJ248 and LJ249, which were dissimilar in property despite differing by the sole presence of the absence of C248. RNA LJ248 was predominantly dimeric and RNA LJ249 was predominantly monomeric. Specifically, RNA LJ248 incubated under H conditions was 100% dimeric in TBM and 90% dimeric in TBK or TBN. Under L conditions, it was 60−70% dimeric in TBM and 50−60% dimeric in TBK or TBN. The corresponding numbers for RNA LJ249 were 30%, 10%, 10%, and 0%. RNA LJ248 was always more dimeric than RNAs LJ242, -243, or -246. In other experiments, RNA LJ249 was incubated in buffers L and H supplemented with 5 mM ammonium acetate before electrophoresis in TBN; 2 times smaller amounts of LJ233 and 3 times larger amounts of RNAs LJ216, -242, and -243 were electrophoresed in TBM. These manipulations did not affect the dimer yields (data not shown).

While most RNA bands were sharper in TBM than in TBK or TBN, RNA LJ252 was slightly sharper in TBK and TBN than in TBM, as if it were fully monomeric in TBK and TBN but on the threshold of dimerizing in TBM. The fact that for all RNAs electrophoresis in TBK or TBN reduced or did not change the dimerization yield observed in TBM does not seem to support the idea that K^+ (Sundquist & Heaphy, 1993; Awang & Sen, 1993) or NH_4^+ (Weiss et al.,

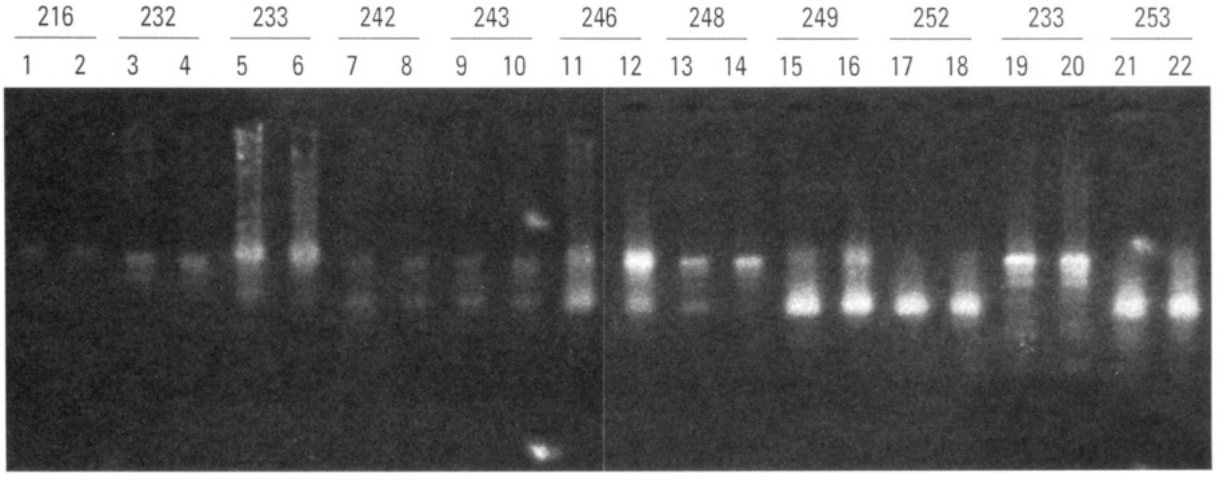

FIGURE 2: *In vitro* dimerization of RNA fragments lacking more and more nucleotides from the 5′ end of the HIV-1 genome. All RNAs end at nucleotide G401. Odd (even) numbered lanes: RNAs incubated in buffer L (H) before agarose gel electrophoresis (10 V/cm, 1.5 h) in TBM. Lanes 1−18: RNAs LJ216, LJ232, ..., LJ252. Lanes 19−22: RNAs B233 and B253.

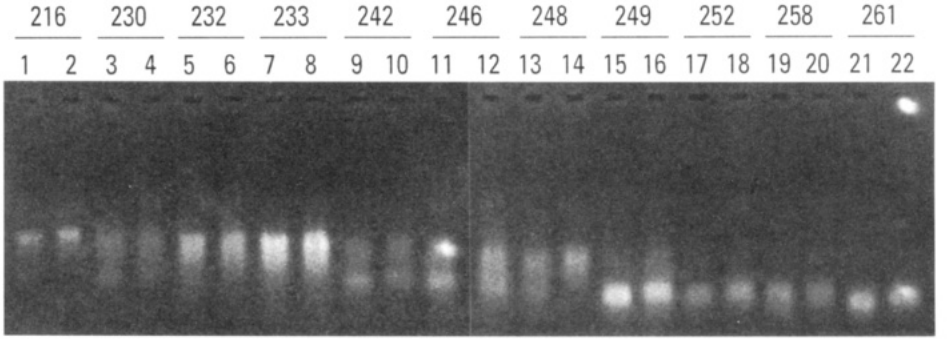

FIGURE 3: *In vitro* dimerization as in Figure 2, except the electrophoresis (7 W, i.e., 6−8 V/cm, 2 h) was done in TBK, and some new RNAs were run, such as RNAs LJ230 (lanes 3 and 4), LJ258 (lanes 19 and 20), and LJ261 (21 and 22).

Table 1: RNA Transcripts Made from Plasmid pBL and Some Nest-Deleted Plasmids Such as pLJ230, pB233, pLJ252, and pB253[a]

plasmid	polylinker-coded sequence (non-HIV-1)	HIV-1 sequence
pEA	pppGGGAGACCGGCA	GAUCUGAGCCU(...) +20
pLJ230	pppGGGAGACCGGCAGA	UCUCGACGCAGA(...) +230
pBS69-2	pppGGGCGAAUUGG	CGACGCAGA(...) +233
pB233	pppGGG	CGACGCAGA(...) +233
pLJ252	pppGGGAGACCGGCAGAU	CUGAAGC(...) +252
pBS165-1	pppGGGCGAAUUGG	UGAAGCGCGC(...) +253
pB253	pppGGG	UGAAGCGCGC(...) +253
pBSEA	pppGGGCGAAUUGG	GAUCUGAGCCU(...) +20
pBL	pppGGGUCUCUCUGGUUAGACCAGAUCU(...) +1 +20	

[a] (...) designates the presence of a normal HIV-1 RNA sequence following the last nucleotide explicitly typed in this table. For ease of display, the various HIV-1 sequences have been arbitrarily separated from their respective upstream polylinker-coded sequences. To indicate that the HIV-1 sequence is, in reality, immediately adjacent to the polylinker-coded sequence, they are joined by a line.

1993) has a privileged role in the dimerization of HIV-1 RNA.

The 233−251 region can thus be divided into three subregions: 233−241, 242−247, and 248−251. The first played a moderate role in the dimerization process: its absence reduced the proportion of dimers by 35−50% (compare RNAs LJ233−401 and LJ242−401). The second played either no role (seen in buffer L) or a negative role (seen in buffer H) in the dimerization process. The third played a predominant role: its absence reduced the dimer yield 10-fold or more (compare RNAs LJ242−401 or -248−401 to LJ252−401). It will be shown that, in the absence of the 242−251 region, the 233−241 region can play a crucial role in the dimerization process.

For technical reasons related to exonuclease III digestion and the location of the T7 promoter of pEA, pLJ transcripts had 14−16 additional, non-HIV-coded nucleotides at their 5′ end (Table 1). More precisely, they had the first two or three nucleotides of the HIV-1 genome at their 5′ ends followed by 11−14 non-HIV-coded nucleotides and a fully HIV-1 coded sequence from there on to their 3′ ends. To verify whether these supernumerary nucleotides influenced the dimerization process, plasmids pB233 and pB253 were constructed. RNA B233 is RNA LJ233 without the extraneous AGACCGGCAGA located between its 5′ end pppGGG and the HIV-1 sequence. RNA B253 has one HIV-1 nucleotide less than RNA LJ252 and lacks its extraneous AGACCGGCAGAU (Table 1). Figure 2 (lanes 19−22) shows that RNA B233 was at least as dimeric as RNA LJ233 and that RNA B253 was as monomeric as RNA LJ252 under both L and H incubation conditions. Thus, the general trend manifested in Figure 1 is unlikely to be due to some non-HIV-1 sequence near the 5′ end of the pLJ transcripts, unless the remaining pppGGG (present to maximize transcription yields) plays a freak role, promoting dimerization in LJ216, -230, -232, and -233 but mysteriously not in LJ249, -252, -258, -261, and -267, RNAs that include much more than the DLS. RNAs LJ299, -324, and -359 support this point as long as the 5′ half of the DLS plays no role in dimer formation, a concept for which we have evidence (M. Laughrea and L. Jetté, manuscriptin preparation).

The 5′ Boundary: Oligonucleotide-Directed Mutagenesis. Some secondary structure models and chemical probing studies can be taken to suggest that the 60−280 region of HIV-1 RNA folds on itself (Baudin et al., 1993). Our nested deletions thus might have created boundary conditions leading to artifactual dimerization properties: the absence of R, U5, PBS, and part of the leader sequence might alter the conformation of downstream sequences so as to allosterically incapacitate the physiological dimerization site, or the altered conformation might itself be a dimerization site. According to this scenario, RNAs LJ216, -230, and -233, being prevented from hydrogen bonding with the 60−215 region, adopt an altered conformation that by chance favors dimerization, and RNAs LJ252, -258, -261, -262, and -267, for identical reasons but with opposite effects, fold into an abnormal structure that alters the native dimerization site. Accordingly, phagemid pBL was constructed and suitably

Table 2: RNA Transcripts Made from Derivatives of Plasmid pBL[a]

plasmid	
pBL	pppGGG(...)CUCUCUCGACGCAGGACUCGGCUUGCUGA(...) +1 227 252
pBLΔ233-251	pppGGG(...)CUCUCU(----------)CUGA(...) +1 227 252
pBLΔ242-251	pppGGG(...)CUCUCUCGACGCAGG(--------)CUGA(...) +1 227 241 252

[a] (...) designates the presence of a normal HIV-1 RNA sequence immediately following the nucleotide on the left of (...) up to the nucleotide immediately on the right of (...). (- - - -) designates an RNA deletion, residues on the left and right of (- - - -) now being adjacent in the mutant HIV-1 RNA.

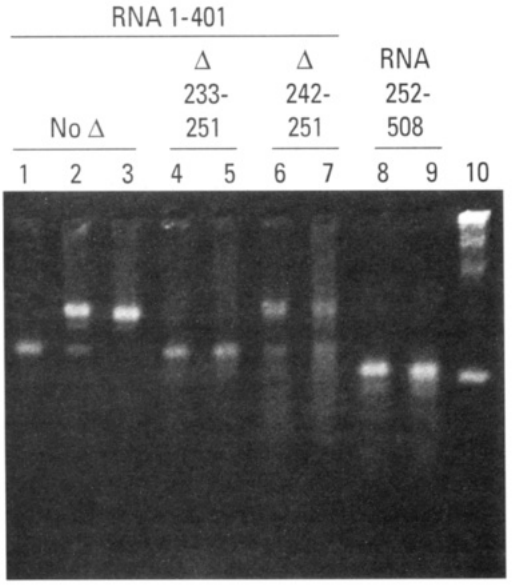

FIGURE 4: *In vitro* dimerization of internally deleted and 3′ extended RNAs. Gel conditions as in Figure 2: lanes 1−3, RNA BL1−401; lanes 4 and 5, RNA BL1−401Δ233−251; lanes 6 and 7, RNA BL1−401Δ242−251; lanes 8 and 9, RNA LJ252−508; lane 10, 0.24−9.5 kb RNA ladder. All RNAs end at G401, except in lanes 8 and 9. Samples were denatured in lane 1, incubated in buffer L in even-numbered lanes, and incubated in buffer H in lanes 3, 5, 7, and 9.

mutated (Table 2): upon transcription, pBL yields an RNA (RNA BL1) whose first 965 nucleotides are those of the average precapped HIV-1 genome, except that it might have one extra G at the 5′ end.[3] RNAs BL1−401Δ233−251 and BL1−401Δ242−251 are identical to RNA BL1, except that they terminate at nucleotide 401 (*Hae*III cut) and they lack the 233−251 and 242−251 regions, respectively. Contrary to RNA BL1−401, which was 60−75% dimeric in buffer L and ≥95% dimeric in buffer H, RNA BL1−401Δ233−251 was monomeric in both L and H incubation conditions (Figure 4, lanes 1−5). An identical pattern was seen in TBN, except that RNA BL1−401 was 50−60% dimeric under L conditions and 90% dimeric under H conditions (data not shown). This confirms the data with RNAs LJ233, B233, LJ252, and B253. Under both L and H incubation conditions, RNA BL1−401Δ242−251 was 60% dimeric in TBM (Figure 4, lanes 6−7) and 40% dimeric in TBN (data not shown), results qualitatively but not quantitively similar to the LJ RNA data. Specifically, the LJ data showed that removal of the 242−251 region reduces dimerization by at least 10-fold, instead of the 1.5−2-fold displayed by the BL data. It is inferred that the 233−241 region of BL1−401Δ242−251 compensates for the lack of a 242−251 sequence, a compensation not available to RNA LJ252. A similar degree of dimerization was thus obtained as long as the 252−401 region was immediately downstream of either the 233−241 or the 242−251 region. The effect of these regions was less than additive: in the presence of the 242−251 region, the 233−241 region played a moderate role in the dimerization process, but a crucial role in its absence.

It is noteworthy (Figure 4, together with unshown 3 h electrophoreses of RNAs BL1−401 and BL1−508 to make the data more quantifiable) that, at least with RNAs BL1−401 and BL1−508, two dimer conformations can be distinguished: a slow dimer obtained under L incubation conditions and a normal dimer under H incubation conditions. The slow dimer moves like an RNA whose volume/charge ratio is 5−12% larger than that of the normal dimer, assuming no change in shape. RNA BL1−508 comes from an *Acc*I cut of pBL prior to transcription. We also note that RNA BL1−401 was less dimeric in L conditions than RNAs B233−401, LJ216−401, LJ230−401, LJ232−401, and LJ233−401 (Figures 2−5). There seems to be some generality to this result as RNAs BL1−508, EA20−401, and EA20−508 were 60−75% dimeric under L/TBM conditions, 40−60% dimeric under L/TBK conditions, and completely dimeric in H conditions (data not shown). RNAs EA20−401 and EA20−508 came from an *Hae*III and an *Acc*I cut of pEA prior to transcription, respectively.

The 3′ Boundary: Some Nested Deletion Experiments. Our results indicate that the 5′ boundary of the dimerization domain of HIV-1 should be moved 62 nucleotides upstream from the 5′ boundary of the DLS. Should the 3′ boundary, set by Darlix et al. (1990) and Marquet et al. (1991) at G401 or upstream, also be extended? RNA LJ252−401 was always monomeric, but RNA LJ249−401 was monomeric in L/TBK conditions and partly dimeric in TBM and H/TBK conditions, indicating that RNA LJ252−401 is a monomer at, or almost at, the threshold of dimerization. RNA LJ252−508 was prepared. We reasoned that if the 402−508 region played any direct role in the dimerization process, RNA LJ252−508 might become partly dimeric in buffer H. It remained completely monomeric (Figure 4, lanes 8 and 9). RNA LJ249−508 was also prepared. (It cannot be doubted that RNA LJ249−401 is at or slightly above the threshold of dimerization and that the slightest extra help should boost the dimer yield.) In both L/TBM and H/TBM conditions, RNA LJ249−508 was, if anything, slightly less dimeric than RNA LJ249−401 (data not shown). These results suggest that the 3′ boundary of the dimerization domain does not extend downstream of G401, unless the 402−508 region needs to interact with sequences outside the 249−508 region to have any effect. Can the 3′ boundary be located somewhat upstream of G401?

RNAs B233−359, B233−378, and LJ242−342 were produced. RNA B233−378 has lost a string of three G's implicated by Awang and Sen (1993) in the dimerization of

[3] The 5′ end of the HIV genome may vary by ±1 nucleotide (Muesing et al., 1985). Some find that HIV RNA starts on average at what would be position 0 of Table 1 (Starcich et al., 1985). Others find that it starts on average at position +1 of Table 1 (Sanchez-Pescador et al., 1985; Wain-Hobson et al., 1985; Muesing et al., 1985; Guyader et al., 1987).

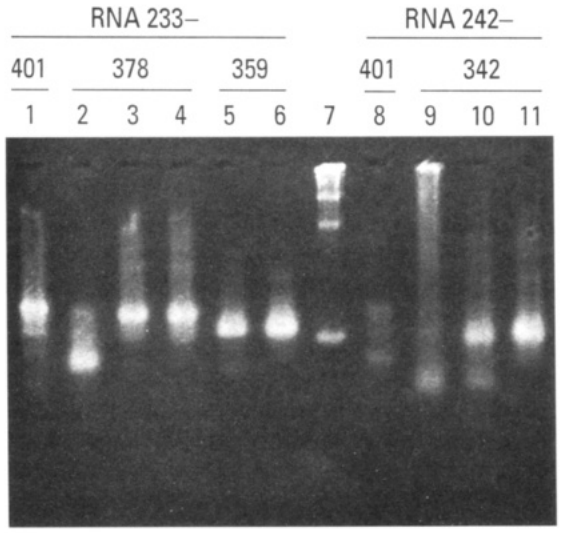

FIGURE 5: *In vitro* dimerization of RNA transcripts lacking more and more nucleotides from the 3′ end of the HIV-1 genome. Gel conditions as in Figure 2: lane 1, RNA B233−401 in buffer L; lanes 2−4, RNA B233−378 denatured, in buffer L, and in buffer H; lanes 5 and 6, RNA B233−359 in buffer L or H; lane 7, 0.24−9.5 kb RNA ladder; lane 8, RNA LJ242−401 in buffer L; lanes 9−11, RNA LJ242−342 denatured, in buffer L, and in buffer H.

the 305−415 region of HIV-1 MAL RNA (289−401 in HIV-1 LAI RNA), and RNA B233−359 has lost a second string of G's (GGGGG367) implicated by Awang and Sen (1993) in the dimerization of the 289−401 region and by Sundquist and Heaphy (1993) in the dimerization of the 278−371 region of HIV-1 NL43 RNA (same positions in HIV-1 LAI RNA). RNA LJ242−342 has lost the sequence 343−354, which Hayashi et al. (1993) base-paired with nucleotides 231−246 of the upstream dimerization signal. Figure 5 shows that the 379−401 region, the 360−378 region, or the 343−359 region had no clear-cut positive effect on the ability to dimerize RNAs starting at nucleotide 233 or 242.

Specifically, RNAs B233−401 and B233−378 were 100% dimeric in both L and H incubation conditions, while B233−359 was 100% dimeric in buffer H and 85−95% dimeric in buffer L (lanes 1−6 of Figure 5), i.e., it always displayed a faint to barely visible monomer band in L conditions. This would suggest that the 360−378 region has a weak positive effect on the ability to dimerize in buffer L, were it not for lanes 8−11 of Figure 5, which lead to a contrary conclusion. RNA LJ242−342 was 65−70% dimeric under L conditions and 100% dimeric under H conditions, exactly as RNA LJ242−359 (data not shown), while RNA LJ242−401 has always been 40−55% dimeric under both L and H incubation conditions, as if the 360−401 region was detrimental to the dimerization process. We conclude that, on the whole, the effect of the 343−401 or 360−401 region on dimer formation is neutral if not negative. It is noteworthy that a contrary conclusion was reached from the study of RNA transcripts 278−401 (Sundquist & Heaphy, 1993) and 289−401 (Awang & Sen, 1993). Experiments are in progress to verify whether the 343−359, 360−378, or 379−401 regions might confer greater thermal stability to formed dimers.

DISCUSSION

Relevance of in Vitro Studies. There are now good grounds to believe that the full-length HIV-1 genome can dimerize without nucleocapsid proteins (Fu et al., 1994; Fu & Rein, 1993; Stewart et al.,1990) and that its thermal stability is comparable to that of dimeric 5′ end transcripts such as HIV-1 RNAs 1−615 and 1−707 (Fu et al., 1994; Marquet et al., 1994; Roy et al., 1990). Since no compelling evidence exists in favor of a role for the Pr55gag polyprotein in the dimerization of HIV-1 RNA transcripts (Berkowitz et al., 1993), it is plausible that the HIV-1 genome can dimerize *in vivo* without the help of proteins. This would explain why, despite containing an average of 0.075 retroviral RNA per virion, some packaging defective Moloney murine leukemia viruses harbor retroviral RNAs that are almost exclusively dimeric (Meric & Goff, 1989). It is possible that nucleocapsid proteins contribute to a late step of the dimerization process: the transition from an immature to a mature dimer characterized by an electrophoretically faster structure and a 5 °C higher melting temperature (Fu et al., 1994; Fu & Rein, 1993). These observations strongly support the concept that results obtained *in vitro* on HIV-1 RNA transcripts in the absence of added proteins reflect some, if not much, of the biological situation, unless the transcripts adopt an idiosyncratic internal structure, i.e., one that varies with the length of the HIV-1 flanking sequences (see following). On the other hand, despite the negative *in vitro* results of Berkowitz et al. (1993), we remain aware that an *in vivo* role for the Pr55gag polyproteins (or even the nucleocapsid proteins) in the kinetics of formation of dimeric HIV-1 RNAs has not yet been completely ruled out.

Self-Consistency of Data Obtained with 30 Different HIV-1 RNAs. Our results lead us to three conclusions: (1) a dimerization signal exists in the 233−251 region, i.e., upstream of the major 5′ splice donor; (2) the previously proposed downstream dimerization domain (Darlix et al., 1990; Marquet et al., 1991; Sundquist & Heaphy, 1993; Awang & Sen, 1993) is insufficient to promote dimerization and has a 3′ half that is not necessary to obtain fully dimeric RNAs; (3) the 5′ boundary of the HIV-1 dimerization domain is located somewhere between nucleotides 233 and 242, and the 3′ boundary is located no farther than at nucleotide 342.

The evidence for conclusion 1 is unambiguous (see Results section). To explain why the 233−241 and 242−251 regions seem to have less than additive effects, we imagine that less than the total dimerization domain is needed to obtain almost complete dimerization under our ionic conditions. Under more stringent conditions, the lack of one or the other region might yield very few dimers compared to the two together.

The evidence for conclusion 2 is numerous, but the implications are perplexing. It has been speculated that the DLS, a sequence located between the SD (289−290) and nucleotide 401 of HIV-1 RNA, is the dimerization domain of the HIV-1 genome. Our experiments demonstrate (i) that RNA sequences including all of the DLS but few flanking sequences (e.g., RNAs LJ252−401, BL253−401, LJ258−401, LJ261−401, LJ262−401, and LJ267−401) do not dimerize and (ii) that RNA sequences upstream of the SD are needed for the dimerization of HIV-1 RNA transcripts, in apparent contradiction of the results of Marquet et al. (1991), Sundquist and Heaphy (1993), and Awang and Sen (1993). The simplest way to harmonize our results and those of these laboratories is to propose that our HIV-1 RNA transcripts, which together start at nucleotides 1, 20, 216, 230, 232, 233, 242, 243, 246, 248, 249, 252, 258, 261, 262, 267, 299, 324, and 359, adopt mutually consistent conforma-

tions, no matter whether they end at nucleotide 342, 359, 378, 401, or 508, and that the RNAs studied by Sundquist and Heaphy (1993), Awang and Sen (1993), and Marquet et al. (1991), which start at nucleotides 278, 289, and 295, respectively, and end near nucleotide 401, adopt an idiosyncratic internal structure. The tertiary structure of the polypurine-rich 360−401 region thus seems highly dependent on the length of its 5′ flanking sequence: a 65−80 nucleotide long flanking sequence would give this region remarkable physical properties that are lost when it is part of a more complete HIV-1 RNA sequence [see also Marquet et al. (1994)], or even shorter ones (see our RNAs LJ299, LJ324, and LJ359).

We conclude that when the 360−401 or 360−378 region has a positive effect on dimer formation, the effect is not biologically meaningful but is peculiar to a precise set of *in vitro* conditions and a particular size of RNA transcript. Marquet et al. (1994) and Fu et al. (1994) have provided convincing data directly supporting this interpretation. Their data suggest that the dimers formed by HIV-1 MAL RNA 311−415 (295−401 in LAI) have melting temperatures of >80 vs ~47 °C for RNAs 1−615 and 1−707 and ~60 °C for full-length genomic RNA. We also draw attention to the fact that HIV-1 RNA transcripts containing only the DLS dimerize modestly at best, the dimerization level ranging from less than 5% (our results) to 25−33% (Marquet et al., 1991, 1994; Awang & Sen, 1993) or 50% (Sundquist & Heaphy, 1993).

Why shouldn't our RNA transcripts also fold in an artifactual way? We have studied 30 HIV-1 RNA transcripts containing different sequences from the 1−508 region of the HIV-1 genome, and consistent results were obtained with these RNAs; this, we argue, gives a strong measure of confidence that the common folding of these RNAs is not an artifact and that it bears on the *in vivo* conformation of the dimerization domain of the HIV-1 genome. The *in vivo* relevance of our results is presently being directly addressed by site-directed mutagenesis (M. Laughrea and L. Jetté, work in progress).

The evidence for conclusion 3 also requires discussion. The boundaries indicated are those of an HIV-1 RNA transcript able to dimerize, not necessarily one that dimerizes with the same kinetics as full-length HIV-1 genomic RNA and whose dimeric structure has thermodynamic properties equal to those of full-length genomic RNA. Hence, we do not know whether the indicated boundaries enclose a core dimerization domain or a dimerization domain, an issue presently being addressed (M. Laughrea and L. Jetté, work in progress). A dimerization domain is the shortest available dimeric sequence that mimicks the dimerization properties of the protein-free HIV-1 genome. A core dimerization domain is a subset of the dimerization domain, i.e., a smaller dimer whose relevant physicochemical properties (e.g., kinetics of dimerization, stability of the dimer) fall short of matching those of longer sequences or those of the HIV-1 genome. The kissing-loop model to be presented in the following section will suggest that the 248−270 region might be a core dimerization domain and that the 233−285 region might enclose the complete dimerization domain of HIV-1 RNA. We propose that the 233−285 region is at least a core dimerization domain of HIV-1 RNA.

Why do we arrive at a fuzzy 5′ boundary? RNA LJ242−342 was completely dimeric in H conditions and 65−70% dimeric in L/TBM conditions, much like RNAs BL1−401, BL1−508, EA20−401, and EA20−508 (see the Results section). This argues in favor of a 5′ boundary located at nucleotide 242. Surprisingly, RNAs B233−401 and B233−378 were completely dimeric in both L and H conditions. It is the lack of nucleotides 233−241 that seems to make RNA LJ242−342 less dimeric than them, because RNA LJ242−401 was even less dimeric (50% dimeric in L and H conditions) than LJ242−342. This argues in favor of a 5′ boundary located at nucleotide 233. The 5′ boundary thus is imprecisely known because the long RNAs (e.g., RNAs BL1−401, BL1−508, EA20−401 and EA20−508) were 20−40% less dimeric in L conditions than some of their shorter counterparts, such as RNAs LJ216−401, LJ230−401, LJ232−401, LJ233−401, B233−401, and B233−378. The reason for this slight discrepancy is unclear. Perhaps one should not overinterpret the dimerization yields and consider as equivalent RNAs that dimerize to a 50−100% level and distinguish them only from RNAs that dimerize to a 25−50% level, a 12−25% level, etc. Another interpretation is that some sequences located between nucleotides 20 and 216 might reduce the dimer yield. Finally, RNA LJ248−401 was as dimeric as RNA BL1−401 or RNA LJ242−342. The next section will show why RNA LJ248−401 cannot be used to argue for a 5′ boundary located at C248.

Kissing-Loop Model of HIV-1 RNA Dimerization. Monomer and dimer samples of RNA 1−707 from HIV-1 MAL have been probed chemically (Baudin et al., 1993). Major reactivity changes were observed at UU245 (CU228 in LAI), in the 269−297 region (252−280 in LAI), and in the 312−370 region (296−356 in LAI). The first two areas enclose the upstream dimerization signal identified by us (the 233−251 region) and the third is located within the DLS. The absence of a reactivity change within the 357−401 region supports our finding that the 360−401 region can be deleted without significantly changing the ability to dimerize. However, no major reactivity change was seen within the 233−251 region, as if it was base-paired to about the same sequence in monomer and dimer samples. The simplest way to harmonize our findings and the chemical probing data is to suggest that the 233−251 region base-pairs to the same sequence in monomer and dimer samples, except that in the dimer the chosen sequence belongs to the *other* genome. This putative switch from intra- to interstrand bonding is not easily imagined from the secondary structure model of Baudin et al. (1993), but is made plausible by that of Harrison and Lever (1992) derived from a mixture of computer modeling and chemical/enzymatic probing of RNA 87−635 from HIV-1 LAI/HXB2.

In the Harrison and Lever model, the 232−286 region forms an imperfect hairpin, the stem being the 232−254 region bulged out at AGGA242 and base-paired to the 264−286 region (Figure 6A). Kethoxal and dimethyl sulfate probing experiments done by John Tam in our laboratory on RNA EA20−508 of HIV-1 LAI (data not shown) are more consistent with the model of Harrison and Lever in the 233−245 region, the only part of the region of interest examined by Tam: G240, G241, and A242 were readily accessible as in the Harrison and Lever model but in variance with the Baudin model, which has the two G's located in a stem. [It should be pointed out that the experimental data of Baudin et al. (1993) seem to us about as consistent with

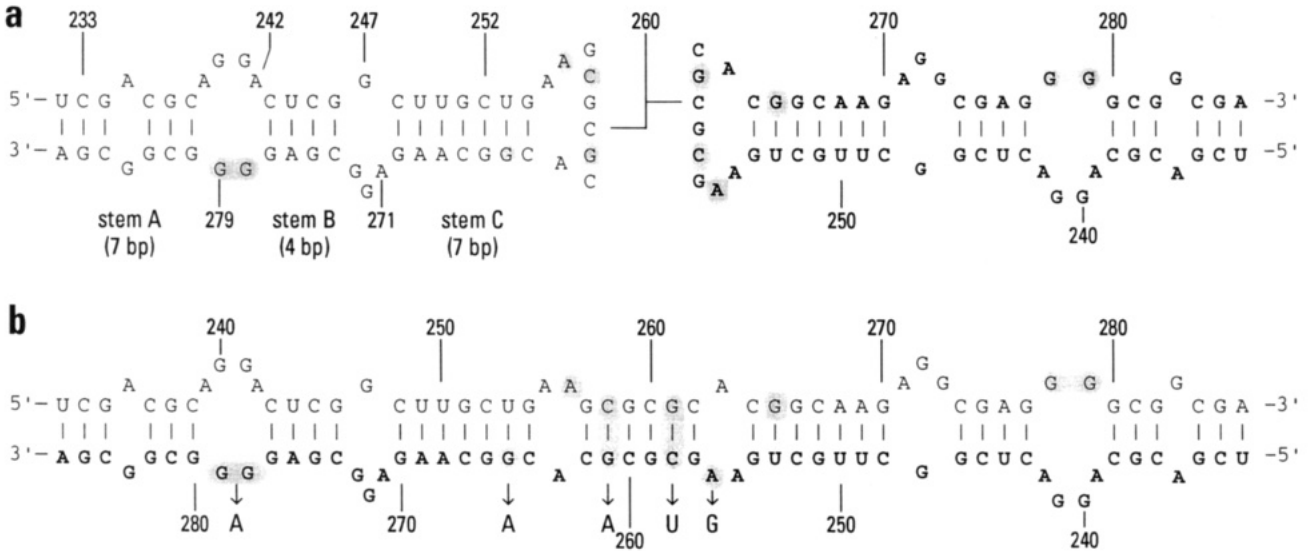

FIGURE 6: Secondary structure model for nucleotides 232−286 of HIV-1 in monomeric and dimeric forms. Nucleotides that differ in HIV-1 MAL and HIV-1 LAIpB2 are shadowed, and the corresponding MAL nucleotides are indicated by arrows in the bottom part of b. (a) Postulated stem−loop structure of two monomers with their respective GCGCGC262's almost ready to hydrogen-bond to each other. (b) Proposed structure of the dimer with one strand in bold characters and the other in normal characters.

the Harrison and Lever folding as with the Baudin folding.] Why might the 232−254 and 264−286 regions of one monomer split and base-pair with the 264−286 and 232−254 regions, respectively, of another monomer? Because the two 255−263 loops (AAGCGCGCA263) thus would not form an internal loop but a strong duplex through their palindromic GCGCGC262, thereby shifting the equilibrium toward the formation of dimers and reducing the activation energy of the monomer−dimer conformational switch (Figure 6). This kissing-loop model suggests that the upstream dimerization signal extends from C232 to G286 and that its heart is the 248−270 region, i.e., hairpin C (the hairpin formed by stem C and the 255−263 loop). It explains why RNA 1−311 (1−295 in LAI) of Marquet et al. (1994) dimerizes, and why our RNAs do not dimerize when they start downstream of 251 or when they have the 233−251 region deleted. It is in harmony with our other deletion and nested-deletion experiments if hairpin 232−286 is given four plausible attributes: (i) stem B plays a smaller role in the dimerization process than stem A or stem C; (ii) the absence of stem A can be partly compensated by a 1 bp longer stem C; (iii) moderate dimerization yields can be obtained despite the absence of stem A or the lower half of stem C; (iv) the effect of these two elements is less than additive in our ionic conditions. Then the kissing-loop model explains why RNAs LJ242, LJ243, and LJ246 are half-dimeric (they have lost stem A), why they dimerize equally well (they only differ by various levels of stem B amputation), why RNA LJ248 is more dimeric than RNAs LJ242, -243, and -246 (it has the potential for an 8 bp long stem C, because its C248 is adjacent to a plasmid-coded U that we imagine base-paired with otherwise bulged-out A271), and why RNA LJ249 is predominantly monomeric and strikingly different from RNA LJ248 (its stem C is 2 bp shorter than that of RNA LJ248).

In addition, our data suggest that an intact stem A plus a 4 bp hairpin C is slightly better than a normal 7 bp hairpin C without any stem A, but slightly worse than an extra long 8 bp hairpin C without any stem A: RNA BL1−401Δ242−251 is slightly more dimeric than RNAs LJ242, LJ243, and LJ246 and slightly less dimeric than RNA LJ248. More experiments will be needed to understand the nature of the putative equivalence between stem A and half of stem C and the reasons for the exceedingly light role of stem B. We speculate that the 233−285 region is at least a core dimerization domain of the HIV-1 genome and that the dimerization process is a multistep process in which sequences of the DLS might further stabilize (Marquet et al., 1994) a process initiated upstream of the splice donor, at the level of hairpin C.

Upstream and Downstream Domains Folded into One? Hayashi et al. (1993) have published a mostly computer-derived secondary structure model of the 5′ end of the HIV-1 genome, which happened to hybridize the 231−246 region of the upstream dimerization signal to the 343−354 region of the DLS. Our results indicate that dimer formation does not depend on such an interaction: RNA LJ242−342 made such an interaction impossible without reducing the dimerization yield. The simplest explanation for this result and the fact that the 233−251 region is involved in the dimerization process is that the 231−246 region never interacts with sequences downstream of 342 in HIV-1 RNA.

Relationship between Dimerization and Encapsidation? The DLS of HIV-1 is thought to encompass an important, although not sufficient (Richardson et al., 1993; Kim et al., 1994), encapsidation/RNA stabilization signal (Hayashi et al., 1992; Clavel & Orenstein, 1990; Lever et al., 1989; Aldovini & Young, 1990; Luban & Goff, 1994). This has raised the possibility that the DLS might play overlapping roles: one in encapsidation, or stabilization (Aronoff et al., 1993), of the encapsidated RNA, and one in dimerization. This paper shows that the 3′ 60% of the DLS, including the polypurine stretches, is unnecessary for obtaining dimeric HIV-1 RNA transcripts. In contrast, deletion of the first 40 nucleotides of the Pr55gag coding sequence reduces the level of RNA encapsidation/stabilization 5-fold (Luban & Goff, 1994). What about the 5′ 40% of the DLS? Lever et al. (1989), Clavel et al. (1990), Aldovini and Young (1990), and Hayashi et al. (1992) have presented evidence suggesting that sequences located between the SD and the Pr55gag coding sequence play a role in genome packaging. In

contrast again, we have evidence (M. Laughrea and L. Jetté, manuscript in preparation) that excellent dimerization yields are obtained with transcripts whose 3′ boundary lies upstream of the SD, no matter whether their 5′ boundary is nucleotide 1 or 233. The therefore uncertain or seemingly weak role of the DLS in dimer formation [see also Marquet et al. (1994)] emphasizes that an overlap between a downstream encapsidation signal and an undefined downstream dimerization signal might be tenuous. What is conceivable and has not yet been ruled out is that the 5′ 40% of the DLS, or even its 3′ 60%, might stabilize the dimers formed [see Marquet et al. (1994)] or improve the kinetics of dimer formation.

Nevertheless, the overlap hypothesis seems to hold upstream of the SD. It has recently been shown that deletion of the 241−253 region of the HIV-1 LAI genome reduces the amount of intact genomic RNA present per virion recovered from MT4 cells by 80−85% (Kim et al., 1994). The 241−253 region is essentially a subset of the dimerization signal identified in this paper. The exact nature or meaning of this upstream neighborhood is not clear. One interpretation is that encapsidation of dimeric RNA takes precedence over encapsidation of monomeric RNA. Another interpretation is that dimerization stabilizes genomic RNA, i.e., protects it against rapid degradation both outside and inside the virus. One is also allowed to speculate that the 242−247 region might play a positive role only in RNA encapsidation, since it seems to play no detectable role in RNA dimerization.

Sequence Heterogeneity. It is worth noting the differences in sequence between the dimeric HIV-1 RNAs studied thus far, i.e., MAL transcripts (Marquet et al., 1994; Awang & Sen, 1993), NL 43 transcripts (Sundquist & Heaphy, 1993), and our transcripts. The plasmids used in the present work to synthesize our HIV-1 RNA transcripts derive from the pB2 molecular clone of an LAI isolate (Benn et al., 1985; Gendelman et al., 1986; Wain-Hobson et al., 1991). In the course of the sequencing of the 20−90 and 151−519 RNA regions required for the present studies, we have found that the pB2 sequence was about 1% different from the official LAI sequence (Wain-Hobson et al., 1985; Myers et al., 1993). Specifically, it differed at positions 224 (A instead of G in LAI), 281 (C instead of A in LAI), 303 (U instead of A in LAI and a deletion in MAL), 305 (U instead of A in LAI and a deletion in MAL), and 455 (A instead of G in LAI and MAL). Four of these five changes are in nonconserved regions of the genome, and the fifth (C281) conforms pB2 to all other HIV-1 strains. In comparison, HIV-1 HXB2, seemingly another component of the LAI sequence spectrum (Chang et al., 1993; Shaw et al., 1984; Rabson & Martin, 1985; Ratner et al., 1985; Myers et al., 1993), differs from HIV-1 LAI at seven positions within the equivalent RNA region (three point mutations at positions 23, 200, and 281, deletion of nucleotides 224−226, and insertion of CC[4] between nucleotides 260 and 261). These variations reflect the fact that any sample of HIV-1 virus isolated from an infected patient contains a spectrum of genomic sequences, each sequence typically differing from the consensus by about 2% (Rabson & Martin, 1985; Saag et al., 1988; Sakai et al., 1988; Meyerhans et al., 1989; Goodenow et al., 1989; Martins et al., 1992; Wolfs et al., 1992; Hahn et al., 1986). Cloning destroys this spectrum and amplifies one sequence at the expense of all others. The official LAI is one such sequence, HXB2 is another one, and ours, which we now call LAIpB2, is a third, each presumably originating from the same patient.

In the 228−401 region, LAIpB2 transcripts differ from MAL transcripts at 13 nucleotides, 6 of them located within the kissing-loop structure of Figure 6. These have been indicated in Figure 6. The other changes are three point mutations at C378, G379, and A386, the deletion of AAA305, and a U insertion between U309 and G310. Given the sequence disparity, it is very satisfying that a common mechanism can explain the dimerization of both LAIpB2 and MAL RNA transcripts. In the 228−401 region, LAIpB2 transcripts differ from NL43 transcripts by three point mutations at A303, A305, and A353. It seems highly unlikely that these small differences can explain some of the variance between our conclusions and those of Sundquist and Heaphy (1993).

Convergence. After the experiments described here had been fully completed and several drafts of this paper had been written, we were made aware of the work of Skripkin et al. (1994), then in press. Skripkin et al. (1994) showed that deletion of the 265−287 region of HIV-1 MAL (248−270 in LAI) or mutation of nucleotides 274−276 (257−259 in LAI) abolished dimerization. Our results and those of Skripkin et al. (1994) are complementary and mutually supportive. It is, however, intriguing that with many of our RNAs, the degree of dimerization was almost equal in the presence of low or high salt, something not seen in Skripkin et al. (1994) or Marquet et al. (1994). We do not have an insightful explanation for this variance.

ACKNOWLEDGMENT

We thank Drs. Léa Brakier-Gingras and Lawrence Kleiman for critical reading of the manuscript, Rhona Rosenzweig for the preparation of the typed text, and Christine Lalonde for the illustrations.

REFERENCES

Aldovini, A., & Young, R. A. (1990) *J. Virol. 64,* 1920−1926.
Aronoff, R., Hajjar, A. M., & Linial, M. L. (1993) *J. Virol. 67,* 178−188.
Awang, G., & Sen, D. (1993) *Biochemistry 32,* 11453−11457.
Baudin, F., Marquet, R., Isel, C., Darlix, J. L., Ehresmann, B., & Ehresmann, C. (1993) *J. Mol. Biol. 229,* 382−397.
Benn, S., Rutledge, R., Folks, T., Gold, J., Baker, L., McCormick, J., Feorino, P., Piot, P., Quinn, T., & Martin, M. (1985) *Science 230,* 949−951.
Berkhout, B., Oude Essink, B. B., & Schoneveld, I. (1993) *FASEB J. 7,* 181−187.
Berkowitz, R. D., Luban, J., & Goff, S. P. (1993) *J. Virol. 67,* 7190−7200.
Bieth, E., Gabus, C., & Darlix, J. L. (1990) *Nucleic Acids Res. 18,* 119−127.
Chang, S.-Y. P., Bowman, B. H., Weiss, J. B., Garcia, R. E., & White, T. J. (1993) *Nature 363,* 466−469.
Clavel, F., & Orenstein, J. M. (1990) *J. Virol. 64,* 5230−5234.
Darlix, J.-L., Gabus, C., Nugeyre, M.-T., Clavel, F., & Barré-Sinoussi, F. (1990) *J. Mol. Biol. 216,* 689−699.

[4] HIV-1 HXB2C does not have this CC insertion, because (like LAIpB2) it is *Bss*HII cleavable at GCGCGC262 (Berkowitz et al., 1993; Luban & Goff, 1994). In terms of the kissing-loop model, LAIpB2 and HXB2C thus have identical loops.

Darlix, J.-L., Gabus, C., & Allain, B. (1992) *J. Virol. 66,* 7245–7252.
Fu, W., & Rein, A. (1993) *J. Virol. 67,* 5443–5449.
Fu, W., Gorelick, R. J., & Rein, A. (1994) *J. Virol. 68,* 5013–5018.
Gendelman, H. E., Phelps, W., Feigenbaum, L., Ostrove, J. M., Adachi, A., Howley, P. M., Khoury, G., Ginsberg, H. S., & Martin, M. A. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83,* 9759–9763.
Guyader, M., Emerman, M., Sonigo, P., Clavel, F., Montagnier, L., & Alizon, M. (1987) *Nature 326,* 662–669.
Hahn, B. H., Shaw, G. M., Taylor, M. E., Redfield, R. R., Markham, P. D., Salahuddin, S. Z., Wong-Staal, F., Gallo, R. C., Parks, E. S., & Parks, W. P. (1986) *Science 222,* 1548.
Harrison, G. P., & Lever, A. M. L. (1992) *J. Virol. 66,* 4144–4153.
Hayashi, T., Shioda, T., Iwakura, Y., & Shibuta, H. (1992) *Virology 188,* 590–599.
Hayashi, T., Ueno, Y., & Okamoto, T. (1993) *FEBS Lett. 327,* 213–218.
Hoch, H., & Lewallen, C. G. (1977) *Anal. Biochem. 70,* 312–317.
Hu, W.-S., Pathak, V. K., & Temin, H. M. (1993) in *Reverse Transcriptase* (Skalka, A. M., & Goff, S. P., Eds.) pp 251–274, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Jones, J. S., Allan, R. W., & Temin, H. M. (1993) *J. Virol. 67,* 3151–3158.
Jones, J. S., Allan, R. W., & Temin, H. M. (1994) *J. Virol. 68,* 207–216.
Katoh, I., Kyushiki, H., Sakamoto, Y., Ikawa, Y., & Yoshinaka, Y. (1991) *J. Virol.* 65, 6845–6855.
Katoh, I., Yasunaga, T., & Yoshinaka, Y. (1993) *J. Virol. 67,* 1830–1839.
Kim, H.-J., & O'Rear, J. J. (1994) *Virology 198,* 336–340.
Laughrea, M., & Moore, P. B. (1977) *J. Mol. Biol. 112,* 399–421.
Laughrea, M., & Tam, J. (1991) *Biochemistry 30,* 11412–11420.
Lever, A., Gottlinger, H., Haseltine, W., & Sodroski, J. (1989) *J. Virol. 63,* 4085–4087.
Luban, J., & Goff, S. P. (1994) *J. Virol. 68,* 3784–3793.
Marquet, R., Baudin, F., Gabus, C., Darlix, J.-L., Mougel, M., Ehresmann, C., & Ehresmann, B. (1991) *Nucleic Acids Res. 19,* 2349–2357.
Marquet, R., Paillart, J.-C., Skripkin, E., Ehresmann, C., & Ehresmann, B. (1994) *Nucleic Acids Res. 22,* 145–151.
Martins, L. P., Chenciner, N., & Wain-Hobson, S. (1992) *Virology 191,* 837–845.
Méric, C., & Goff, S. P. (1989) *J. Virol. 63,* 1558–1568.
Meyerhans, A., Cheynier, R., Albert, J., Seth, M., Kwok, S., Sninsky, J., Morfeldt-Manson, L., Asjö, B., & Wain-Hobson, S. (1989) *Cell 58,* 901–910.
Muesing, M. A., Smith, D. H., Cabradilla, C. D., Benton, C. V., Lasky, L. A., & Capon, D. J. (1985) *Nature 313,* 450–458.
Murphy, J. E., & Goff, S. P. (1989) *J. Virol. 63,* 319–327.
Myers, G., Korber, B., Wain-Hobson, S., Smith, R. F., & Pavlakis, G. N., Eds. (1993) *Human Retroviruses and AIDS: A compilation and Analysis of Nucleic Acid and Amino Acid Sequences,* Los Alamos Natl. Lab., Los Alamos, NM.
Nishizawa, M., Koyama, T., & Kawai, S. (1985) *J. Virol. 55,* 881–885.
Prats, A.-C., Roy, C., Wang, P., Erard, M., Housset, V., Gabus, C., Paoletti, C., & Darlix, J.-L. (1990) *J. Virol. 64,* 774–783.
Rabson, A. B., & Martin, M. A. (1985) *Cell 40,* 477–480.
Ratner, L., Fisher, A., Jagodzinski, L. L., Mitsuya, H., Liou, R.-S., Gallo, R. C., & Wong-Staal, F. (1987) *AIDS Res. 3,* 57–69.
Richardson, J. H., Child, L. A., & Lever, A. M. L. (1993) *J. Virol. 67,* 3997–4005.
Roy, C., Tounekti, N., Mougel, M., Darlix, J.-L., Paoletti, C., Ehresmann, C., Ehresmann, B., & Paoletti, J. (1990) *Nucleic Acids Res. 18,* 7287–7292.
Saag, M. S., Hahn, B. H., Gibbons, J., Li, Y., Parks, E. S., Parks, W. P., & Shaw, G. M. (1988) *Nature 334,* 440–444.
Sakai, K., Dewhurst, S., Ma, X., & Volsky, D. J. (1988) *J. Virol. 62,* 4078–4085.
Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular cloning, a Laboratory Manual,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Sanchez-Pescador, R., Power, M. D., Barr, P. J., Steimer, K. S., Stempien, M. M., Brown-Shimer, S. L., Gee, W. W., Renard, A., Randolph, A., Levy, J. A., Dina, D., & Luciw, P. A. (1985) *Science 227,* 484–492.
Shaw, G. M., Hahn, B. H., Arya, S. K., Groopman, J. E., Gallo, R. C., & Wong-Staal, F. (1984) *Science 226,* 1165–1171.
Skripkin, E., Paillart, J.-C., Marquet, R., Ehresmann, B., & Ehresmann, C. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91,* 4945–4949.
Starcich, B., Ratner, L., Josephs, S. F., Okamoto, T., Gallo, R. C., & Wong-Staal, F. (1985) *Science 227,* 538–540.
Stewart, L., Schatz, G., & Vogt, V. M. (1990) *J. Virol. 64,* 5076–5092.
Sundquist, W. I., & Heaphy, S. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90,* 3393–3397.
Temin, H. M. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90,* 6900–6903.
Torrent, C., Gabus, C., & Darlix, J.-L. (1994) *J. Virol. 68,* 661–667.
Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S., & Alizon, M. (1985) *Cell 40,* 9–17.
Wain-Hobson, S., Vartanian, J.-P., Henry, m., Chenciner, N., Cheynier, R., Delassus, S., Pedrozo Martins, L., Sala, M., Nugeyre, M.-T., Guétard, D., Klatzmann, D., Gluckman, J.-C., Rozenbaum, W., Barré-Sinoussi, F., & Montagnier, L. (1991) *Science 252,* 961–965.
Weiss, S. Häus, G., Famulok, M., & König, B. (1993) *Nucleic Acids Res. 21,* 4879–4885.
Wolfs, T. F. W., Zwart, G., Bakker, M., & Goudsmit, J. (1992) *Virology 189,* 103–110.