**The Dimer Initiation Site Hairpin Mediates Dimerization of the Human Immunodeficiency Virus, Type 2 RNA Genome**

Annette M. G. Dirac‡§, Hendrik Huthoff‡, Jørgen Kjems‡, and Ben Berkhouţ¶

From the ‡Department of Molecular and Structural Biology, Aarhus University, DK-8000 Aarhus C, Denmark and the §Department of Human Retrovirology, Academic Medical Center, University of Amsterdam, 1100 DE Amsterdam, The Netherlands

The untranslated leader of retroviral RNA genomes encodes multiple structural signals that are critical for virus replication. In the human immunodeficiency virus, type 1 (HIV-1) leader, a hairpin structure with a palindrome-containing loop is termed the dimer initiation site (DIS), because it triggers in vitro RNA dimerization through base pairing of the loop-exposed palindromes (kissing loops). Controversy remains regarding the region responsible for HIV-2 RNA dimerization. Different studies have suggested the involvement of the transactivation region, the primer binding site, and a hairpin structure that is the equivalent of the HIV-1 DIS hairpin. We have performed a detailed mutational analysis of the HIV-2 leader RNA, and we also used antisense oligonucleotides to probe the regions involved in dimerization. Our results unequivocally demonstrate that the DIS hairpin is the main determinant for HIV-2 RNA dimerization. The 6-mer palindrome sequence in the DIS loop is essential for dimer formation. Although the sequence can be replaced by other 6-mer palindromes, motifs that form more than two A/U base pairs do not dimerize efficiently. The inability to form stable kissing-loop complexes precludes formation of dimers with more extended base pairing. Structure probing of the DIS hairpin in the context of the complete HIV-2 leader RNA suggests a 5-base pair elongation of the DIS stem as it is proposed in current RNA secondary structure models. This structure is supported by phylogenetic analysis of leader RNA sequences from different viral isolates, indicating that RNA genome dimerization occurs by a similar mechanism for all members of the human and simian immunodeficiency viruses.

The genome of retroviruses is formed by two identical RNA molecules that are non-covalently linked (1, 2). For HIV-1 and other retroviruses, it has been observed by electron microscopy that the most stable contact point between the two RNAs is located within a region close to their 5’ ends, the so-called dimer linkage structure (DLS) (3, 4). The presence of RNA dimers has been suggested to affect virus replication in several ways. First, RNA dimers have been reported to be preferentially packaged, and the DLS may actually be part of the packaging signal (5, 6). Second, the dimeric state facilitates template switching during reverse transcription, thereby minimizing the effects of physical damage to the RNA genome. Third, template switching may allow recombination and thereby increase the genetic diversity and adaptability of retroviruses (reviewed in Ref. 7).

Retroviral RNA dimerizes spontaneously in vitro when incubated in cation-containing buffers (6–9). For HIV-1, it was initially suggested that guanine quartets located downstream of the splice donor site, the “downstream DLS,” are responsible for dimerization (8, 16, 19). The presence of similar guanine stretches in other retroviruses led to the idea that interstrand guanine tetrad formation may serve as a general mechanism for retroviral dimerization (16). However, in vitro studies with bovine leukemia virus and HIV-2 RNA indicated that dimerization occurs independently of the downstream DLS (9, 20), an observation that has also been supported by studies with infectious HIV-1 virus particles (14, 21).

Extensive mutational analysis of the HIV-1 untranslated leader identified the so-called dimer initiation site (DIS) as an essential motif for in vitro dimerization (15, 17, 18, 22). The HIV-1 DIS is located immediately upstream of the splice donor site and consists of a hairpin structure with an exposed palindromic sequence within its loop. According to the proposed model for dimer formation, contact between two DIS hairpins is initiated by base pairing of the palindromes, termed loop-loop kissing (15, 17, 18). Heat treatment or incubation with the viral nucleic acid-binding nucleocapsid protein, which acts as a chaperone, melts the DIS stem and triggers the formation of a dimer with extended interstrand base pairing. This transition is usually referred to as the conversion of the loose dimer (kissing-loop complex) into the tight dimer form (extended duplex) (23, 24).

Whereas the requirements for in vitro dimerization of HIV-1 RNA are well defined, the in vivo mechanism seems more complex. Studies of maturing HIV-1 and Moloney murine leukemia virus particles have indicated the existence of an early and a late dimeric RNA form with a different thermal stability and electrophoretic mobility (5, 25). These two RNA conformations may correspond to the kissing-loop complex and the extended dimer, the loose and tight dimer, respectively. The effect of mutations in the DIS hairpin on virus replication is not very severe, rendering viral infectivity somewhat reduced because of diminished packaging of the RNA genome. Remarkably, packaged RNA exists as dimers with a thermal stability that is similar to that of the wild type genome (21, 26, 27). Relatively minor dimerization defects have been reported in

---

*This work was funded by the European Molecular Biology Organization, the Netherlands Organization for Scientific Research (NWO-CW), the Norfa Foundation, Plasmid Fondet, Aarhus University, and SSVF (Danish National Foundation of Health). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† To whom correspondence should be addressed: Dept. of Human Retrovirology, Academic Medical Center, University of Amsterdam, P. O. Box 22700, 1100 DE Amsterdam, The Netherlands. Tel.: 31-20-6916531; E-mail: b.berkhout@amc.uva.nl.

‡ The abbreviations used are: HIV-1, human immunodeficiency virus, type 1; DLS, dimer linkage structure; DIS, dimer initiation site; PBS, primer binding site; SIV, simian immunodeficiency virus; PCR, polymerase chain reaction; TBE, Tris borate-EDTA; DMS, dimethyl sulfate; DEPC, diethyl pyrocarbonate; TAR, transactivation region.
DIS Hairpin Mediates Dimerization of the HIV-2 RNA Genome

**EXPERIMENTAL PROCEDURES**

*In Vitro Transcription*—All transcripts were generated using the Ambion Megashortscript T7 transcription kit according to the accompanying protocol. In the case of radioactively labeled transcripts, 1 μl of [α-32P]UTP (0.33 MBq/μl) was added to the transcription mixture. Transcripts were purified on 4% denaturing polyacrylamide gels and visualized by either UV shadowing or autoradiography. Upon overnight elution in TE buffer (10 mM Tris-HCl, pH 7.5, 1 mM EDTA), the RNA absorbance measurements or scintillation counting. Transcript stock solutions were renatured by heating to 85°C and slow cooling to room temperature. Aliquots of all transcripts were stored at −20°C. The templates for transcription were made by PCR amplification of the plasmid HIV-2 ROD large, which contains a T7 promoter directly upstream of the +1 position of wild type HIV-2 ROD sequence (31). All primer sequences are listed in Table I. Templates for transcripts starting at the +1 position were generated using the upstream T7 +1 primer and a downstream primer that ends at the position corresponding to the transcript name. Templates for transcription of RNA starting at +126 and +397 were generated using upstream primers starting at the indicated positions to which the T7 promoter sequence had been fused (Table I; T7 +126 and T7 +397) and downstream primers as described above. Templates for transcription of 1/44 DIS mutants were made by PCR mutagenesis, using the upstream T7 +1 primer and downstream primers in which the wild type antisense DIS palindromic sequence was altered to produce the desired mutation (Table I). PCR fragments were excised from agarose gels, purified using the QIAEX II DNA isolation system, and used in T7 transcription.

_In Vitro Dimerization Assays*—Unless otherwise indicated, −2 pmol of radiolabeled RNA was incubated in 10 μl of dimerization buffer (5 mM MgCl₂, 10 mM Tris-HCl, pH 7.5, 40 mM NaCl) for 10 min at 65°C, followed by slow cooling to room temperature. 5 μl of native gel loading buffer (30% glycerol with bromphenol blue dye) was added, and samples were analyzed on 4% native polyacrylamide gels containing 0.25% Tris borate-EDTA (TBE). Gels were run at 150 V at room temperature followed by drying and autoradiography. In the MgCl₂ titration experiments, RNA was incubated either in buffer containing 100 mM NaCl,
Fig. 1. HIV-2 leader RNA structure and T7 transcripts. Top, RNA secondary structure model of the part of the untranslated leader RNA of HIV-2 isolate ROD that is investigated in this study. The model is adapted from Berkhourt (29). The hairpin structures are named after their putative function in HIV-1 replication, and sequence elements that are essential for their function are highlighted in gray. Bottom, the leader RNA domains are drawn to scale and symbolize the different transcripts that were tested in dimerization assays. The in vitro dimerization activity of these transcripts is indicated on the right side.

Dimerization of HIV-2 RNA in Vitro—Radioactively labeled transcripts, mimicking different regions of the untranslated leader RNA of the HIV-2 isolate ROD, were synthesized by T7 RNA polymerase (Fig. 1). We will first compare the nested set of transcripts 1/124, 1/379, and 1/444 for their dimerization ability. These mutant transcripts contain one or more of the motifs that have previously been suggested to be responsible for HIV-2 dimerization, namely the TAR hairpin (1/124), the PBS region (1/379), and the DIS hairpin (1/444). An increasing amount of these transcripts (10–1000 ng) was incubated at 65°C for 10 min and slowly cooled to room temperature in a standard dimerization buffer with 5 mM MgCl₂ to promote the formation of tight dimers. Samples were subsequently subjected to electrophoresis on a native TBE gel, followed by autoradiography (Fig. 2). The transcripts containing either the TAR hairpin (1/124) or the PBS hairpin and the DIS hairpin (1/379) dimerize very poorly, even at the highest RNA concentration. In contrast, transcript 1/444 that includes the DIS hairpin structure, is essential for in vitro dimerization of the HIV-2 RNA.

We next tested the Mg²⁺ requirement for dimerization. Three additional transcripts 1/184, 126/379, and 126/444 were included, of which the latter two lack the TAR hairpin (Fig. 1). The transcripts were incubated either in TEN buffer with 100 mM NaCl, 10 mM Tris- HCl, pH 7.5, 1 mM EDTA or in a buffer with 40 mM NaCl, 10 mM Tris- HCl, pH 7.5, and an increasing MgCl₂ concentration (0.1, 1, or 5 mM). At 0.1 mM MgCl₂, the DIS-containing RNAs 1/444 and 126/444 form dimers, albeit at low efficiency (Fig. 3, lanes 14 and 22). In the absence of Mg²⁺, using the Mfold version 3.0 algorithm (32, 33). Sequences were sent to the Macfarlane Burnet Centre Mfold server (mfold.edu.burnet.au/) and analyzed with standard settings.

RESULTS

Mfold RNA Secondary Structure Predictions—The sequences of HIV-2 and SIV isolates were downloaded from the HIV data base (hiv-web.lanl.gov/). Secondary structure predictions were performed

10 mM Tris-HCl, pH 7.5, and 1 mM EDTA or buffer containing 10 mM Tris-HCl, pH 7.5, 40 mM NaCl, with 0.1, 1, or 5 mM MgCl₂.

Antisense Oligonucleotide Annealing—Approximately 250 ng (2 pmol) of transcript 1/444 in 10 µl of dimerization buffer (5 mM MgCl₂, 10 mM Tris-HCl, pH 7.5, 40 mM NaCl) was incubated for 10 min at 65°C in the presence of 50-fold molar excess of 20-mer antisense oligonucleotides (100 pmol), the sequence of which corresponds to the numbering of the given oligonucleotide (Table I). Incubation was followed by slow cooling to room temperature. Samples were analyzed on 0.25 × TBE gels as mentioned previously.

Chemical and Enzymatic HIV-2 RNA Structure Probing—2 µg of target RNA (transcript 1/544) was incubated for 15 min in 100 µl of buffer containing 100 mM NaCl, 10 mM Tris-HCl, pH 7.5, 1 mM EDTA, and 1.5 µg of carrier antisense human 18 S ribosomal RNA transcribed from the Ambion control template. For chemical probing, 2 µl of DMS or DEPC was added, and the sample was incubated at room temperature for 5 min. The reactions were stopped by ethanol precipitation, followed by extensive washing in 70% ethanol. Samples were dissolved in 15 µl of water. For enzymatic probing, 0.02 units of RNase One (Promega) or 2 units of RNase T1 (Roche Molecular Biochemicals) were added, and the sample was incubated at room temperature for 5 min. Reactions were stopped by addition of 10 µl of 10% SDS and phenol extraction.

The RNA was ethanol-purified and redissolved in 15 µl of water. DNA oligonucleotides for primer extension analysis were labeled with γ-ATP as follows: 100 ng of DNA primer 484/509 was incubated at 37°C for 30 min in 10 µl of T4 polynucleotide kinase buffer (50 mM Tris-HCl, 10 mM MgCl₂, 0.1 mM EDTA, 5 mM dithiothreitol, 1 mM spermidine, pH 8.2) with 10 units of T4 polynucleotide kinase enzyme (Roche Molecular Biochemicals) and 2 µl of γ-ATP (0.33 MBq/µl). Approximately 2 ng of labeled primer was added to 0.25 µg of target RNA in hybridization buffer (50 mM Tris-HCl, pH 7.5, 125 mM KCl). Annealing was performed by heating to 85°C for 2 min followed by slow cooling to room temperature. 5 µl of 3 × concentrated RT buffer (9 mM MgCl₂, 30 mM dithiothreitol, 30 mM of each dNTP, and 150 µg/ml actinomycin D) was added, together with 10 units of avian myeloblastosis virus reverse transcriptase (Stratagene). Reverse transcription was allowed to proceed for 15 min at 37°C and stopped by addition of 15 µl of sequencing stop buffer. Sequence analysis of the template HIV-2-Y large, corresponding to the target RNA, was performed with radiolabeled DNA primer 484/509, using the United States Biochemicals T7 sequenase kit according to the manufacturer’s instructions.

Mfold RNA Secondary Structure Predictions—The sequences of HIV-2 and SIV isolates were downloaded from the HIV data base (hiv-web.lanl.gov/). Secondary structure predictions were performed
DIS Hairpin Mediates Dimerization of the HIV-2 RNA Genome

**Fig. 2.** Concentration-dependent dimerization of HIV-2 transcripts. We compared transcripts 1/124, 1/379, and 1/444, of which 10 (lanes 1, 6, and 11), 50 (lanes 2, 7, and 12), 200 (lanes 3, 8, and 13), 500 (lanes 4, 9, and 14), and 1000 ng (lanes 5, 10, and 15) were incubated at 65 °C for 10 min in 5 mM MgCl₂, 10 mM Tris-HCl, pH 7.5, and 40 mM NaCl and slowly cooled to room temperature. The samples were analyzed on a 4% native TBE gel followed by autoradiography. The monomer (M) and dimer (D) bands are indicated. The annealing positions of the oligonucleotides on the domains of the HIV-2 transcript are indicated in Fig. 2. Although it is possible that some oligonucleotides may not remain annealed to the transcript during gel electrophoresis, we observed that several oligonucleotides affected the gel migration of the monomeric transcript. Such effects were more prominent in gels that were run for an extended period (results not shown), but some of the more profound effects are visible in Fig. 4 (e.g. with antisense oligonucleotides 80/61 and 320/303–444/425, lane 5 and lanes 15–20, respectively). This change in migration is caused either directly by oligonucleotide binding or indirectly by the induction of a conformational change of the RNA. None of the antisense oligonucleotides that target the TAR hairpin, the poly(A) hairpin, or the PBS affect the level of dimerization. The three oligonucleotides, 418/399, 426/407, and 444/425, that target the TAR hairpin are marked 1, 2, 3, 4, 5, and 6, respectively. The weaker bands that comigrate with transcripts 1/184 (lanes 5–8), 1/379 (lanes 9–12), and 126/379 (lanes 17–20) are alternatively folded conformations of the respective RNAs.

**Fig. 3.** Magnesium-induced dimerization of HIV-2 transcripts. We compared the transcripts 1/124, 1/184, 1/379, 1/444, 126/379, and 126/444. Approximately 2 pmol of radiolabeled transcript in a 10-µl volume (200 nM) was incubated at 65 °C for 10 min and slowly cooled to room temperature. Transcripts were incubated in 10 mM Tris-HCl, pH 7.5, 1 mM EDTA, and 100 mM NaCl (lanes 1, 5, 9, 13, 17, and 21), in 10 mM Tris-HCl, pH 7.5, and 40 mM NaCl, with 0.1 mM MgCl₂ (lanes 2, 6, 10, 14, 18, and 22), 1 mM MgCl₂ (lanes 3, 7, 11, 15, 19, and 23), or 5 mM MgCl₂ (lanes 4, 8, 12, 16, 20, and 24). Gel electrophoresis was performed as indicated in the legend to Fig. 2. The monomer (M) and dimer (D) bands of transcripts 1/124, 1/184, 1/379, 1/444, 126/379, and 126/444 are marked 1, 2, 3, 4, 5, and 6, respectively. The weaker bands that comigrate with transcripts 1/184 (lanes 5–8), 1/379 (lanes 9–12), and 126/379 (lanes 17–20) are alternatively folded conformations of the respective RNAs. A certain level of dimerization could also be reached by inclusion of 100 mM Na⁺ (lanes 13 and 21). Dimerization is barely detectable for any of the other transcripts, even at the highest Mg²⁺ and Na⁺ concentrations tested, whereas dimer formation of the DIS containing transcripts 1/444 and 126/444 is strongly enhanced at 5 mM MgCl₂. The absence of TAR in transcript 126/444 only decreased the dimerization efficiency marginally, confirming that the TAR hairpin is not essential for DIS-mediated RNA dimerization. These combined results indicate that sequences in the 370/444 region, which overlaps the DIS hairpin, dimerize at levels that are similar to that of the wild type transcript (lanes 4, 5, and 8). This result indicates that the DIS hairpin must contain a 6-mer palindrome in the loop to facilitate dimerization. Furthermore, the formation and/or stability of the dimers is influenced by the composition of the palindromes. Mutant transcripts with the alternative palindromes GCGGCC and GUG-CAC that are present in HIV-SIV viral isolates, or the unnatural palindrome GUCGAC, each with two or less A/U nucleotides in the 6-mer sequence, dimerize at levels that are similar to that of the wild type transcript (lanes 4, 5, and 8). In contrast, mutants GGUACC and GGUACC with four A/U nucleotides in their palindromes form dimers at an extremely low efficiency (lanes 3 and 6). Computer analysis with the Mfold

**Fig. 4.** Antisense oligonucleotide-mediated mapping of the HIV-2 leader RNA regions involved in dimerization. Approximately 2 pmol of radiolabeled transcript 1/444 was incubated with a 50-fold molar excess of antisense RNA oligonucleotide at optimal dimerization conditions as indicated in the legend to Fig. 2 and subjected to native gel electrophoresis. The monomer (M) and dimer (D) bands are indicated. The annealing position of the antisense oligonucleotides and the domains of the HIV-2 transcript are indicated on top of the autoradiogram.

Antisense Oligonucleotide-mediated Inhibition of Dimerization—Scanning of transcripts with antisense oligonucleotides provides an alternative way of mapping the sequences involved in dimerization. We incubated transcript 1/444, in dimerization buffer with 5 mM MgCl₂, with a series of antisense oligonucleotides that cover the entire transcript length. The transcript and a 50-fold molar excess of DNA oligonucleotide were incubated at 65 °C and DNA/RNA hybrids were allowed to form during slow cooling. The annealing positions of the oligonucleotides on the HIV-2 transcript are indicated in Fig. 4. Although it is possible that some oligonucleotides may not remain annealed to the transcript during gel electrophoresis, we observed that several oligonucleotides affected the gel migration of the monomeric transcript. Such effects were more prominent in gels that were run for an extended period (results not shown), but some of the more profound effects are visible in Fig. 4 (e.g. with antisense oligonucleotides 80/61 and 320/303–444/425, lane 5 and lanes 15–20, respectively). This change in migration is caused either directly by oligonucleotide binding or indirectly by the induction of a conformational change of the RNA. None of the antisense oligonucleotides that target the TAR hairpin, the poly(A) hairpin, or the PBS affect the level of dimerization. The three oligonucleotides, 418/399, 426/407, and 444/425, that target different parts of the DIS domain inhibit dimerization almost completely. Oligonucleotide 426/407, which shields the entire DIS palindrome sequence, has the most profound effect.

The DIS Palindrome Sequence Influences Dimerization—To analyze the sequence requirements for efficient dimerization, we synthesized a series of 1/444 transcripts with a mutated DIS palindromic (Fig. 5A). Compared with the wild type transcripts containing the GGUACC palindrome (Fig. 5B, lane 1), we measured a complete inability to form dimers for the substitution mutant GGUAGG and the deletion mutant GAC that have disrupted palindromes (Fig. 5B, lanes 2 and 7, respectively). This result indicates that the DIS hairpin must contain a 6-mer palindrome in the loop to facilitate dimerization.
algorithm indicated that all dimerization-defective mutants fold the DIS hairpin as shown in Fig. 5.

**The Minimal DIS Domain and Heterodimer Formation—**
Transcripts that contain the 379/444 region with the DIS hairpin dimerize efficiently. To test whether this domain is sufficient for dimerization, we synthesized the minimal DIS transcript 397/444. This short transcript is also able to dimerize, illustrating that the isolated DIS element mediates dimer formation (Fig. 6, lane 2). Because an unrelated transcript band is running close to the dimer signal, we included an equivalent sample that was heated in formamide before gel electrophoresis to specifically melt the dimer (lane 1).

Using a heterodimerization assay, we verified that the slow migrating band seen in our gels is in fact a dimer of two HIV-2 RNA molecules and not an alternative conformation of the RNA. Furthermore, this assay also tested whether the DIS hairpin contains all signals required for optimal dimerization. We incubated the short transcript (397/444) and long (1/444) HIV-2 transcripts, either separately or combined (Fig. 6, lanes 3, 4, and 6, respectively). Both transcripts form homodimers (D1 in lane 3 and D2 in lane 4, respectively), and the heterodimer (D12 in lane 6) is seen as an additional band of intermediate gel mobility. Heterodimer formation seems at least as efficient as formation of the D2 homodimer (Fig. 6, lane 6), indicating that the short DIS transcript dimerizes as efficiently as the long transcript. These results further support the idea that the DIS hairpin structure is sufficient for optimal in vitro dimerization. To further prove that the long and short transcripts dimerize in an identical manner, we mixed the short transcript containing the wild type hairpin with the mutant DIS transcript 397/444. This short transcript is also able to dimerize efficiently. To test whether this domain is sufficient for dimerization, we synthesized the minimal DIS hairpin structures, either separately or combined (Fig. 6, lanes 3 and 4).

**Enzymatic and Chemical Probing of the HIV-2 DIS Hairpin Structure—**
We have established that the DIS hairpin is a critical determinant for in vitro dimerization of HIV-2 RNA. The palindrome must be exposed in the context of the complete HIV-2 leader RNA to form the initial loop-loop kissing interaction between two monomers. We examined the secondary structure of the HIV-2 DIS region in transcript 1/544, which runs up to the start codon of the gag open reading frame. In vitro transcribed RNA was renatured and subjected to enzymatic and chemical modification. This was followed by primer extension with a radiolabeled DNA primer to identify modified nucleotides and separation of the cDNA products by electrophoresis on sequencing gels. A representative gel is shown in Fig. 7. We used several reagents that specifically react with single stranded nucleotides (RNase T1, RNase One, DMS, and DEPC). The majority of the nucleotides that are predicted to be part of the DIS loop are modified (Fig. 7). Furthermore, nucleotides 403–407 are highly accessible to DMS, DEPC, and RNase T1, whereas the other nucleotides that directly flank the DIS hairpin remain unmodified. However, in the existing structure model for the HIV-2 untranslated leader all 5′-flanking nucleotides in the region 379–408 are shown as single stranded (Fig. 1). We therefore reassessed the structure of the DIS hairpin and propose an elongated stem as drawn in Fig. 7. The same secondary structure of the DIS is predicted by the Mfold 3.0 algorithm (32, 33).

**Phylogenetic Analysis of HIV-2/SIV Isolates Reveals Similar DIS Hairpin Structures—**
We analyzed the DIS regions of 11 viral isolates that belong to the HIV-2 subtype A, including the prototype ROD isolate, and three isolates of subtype B. In addition, we included 11 isolates of the related SIVsmm (sooty mangabey) virus group and the single member of the SIVst (stump tailed macaque) group (Fig. 8). This phylogenetic survey shows almost identical DIS structures for HIV-2 subtypes A and B and SIVst. All these DIS elements have the same

---

**Fig. 5.** Dimerization abilities of HIV-2 RNAs with mutated DIS palindromes. A, Mfold predictions of the DIS hairpins with mutated palindromes. All transcripts are predicted to fold a structure in which the DIS palindrome is exposed in the loop of the hairpin. The palindromes are derived from HIV-2 ROD (wild type), a substitution(s) mutant, SIV sykes (SIVsyk), HIV-1 subtype B (HIV-1b), HIV-1 subtype A (HIV-1a), SIV chimpanzee (SIVcpz), SIV mandrill (SIVmandr), a deletion (D) mutant, and an unnatural palindromic (UP) sequence. The palindrome must be exposed in the context of the complete HIV-2 leader RNA to form the initial loop-loop kissing interaction between two monomers. We have established that the DIS hairpin is a critical determinant for in vitro dimerization of HIV-2 RNA. The palindrome must be exposed in the context of the complete HIV-2 leader RNA to form the initial loop-loop kissing interaction between two monomers. We examined the secondary structure of the HIV-2 DIS region in transcript 1/544, which runs up to the start codon of the gag open reading frame. In vitro transcribed RNA was renatured and subjected to enzymatic and chemical modification. This was followed by primer extension with a radiolabeled DNA primer to identify modified nucleotides and separation of the cDNA products by electrophoresis on sequencing gels. A representative gel is shown in Fig. 7. We used several reagents that specifically react with single stranded nucleotides (RNase T1, RNase One, DMS, and DEPC). The majority of the nucleotides that are predicted to be part of the DIS loop are modified (Fig. 7). Furthermore, nucleotides 403–407 are highly accessible to DMS, DEPC, and RNase T1, whereas the other nucleotides that directly flank the DIS hairpin remain unmodified. However, in the existing structure model for the HIV-2 untranslated leader all 5′-flanking nucleotides in the region 379–408 are shown as single stranded (Fig. 1). We therefore reassessed the structure of the DIS hairpin and propose an elongated stem as drawn in Fig. 7. The same secondary structure of the DIS is predicted by the Mfold 3.0 algorithm (32, 33).

**Fig. 6.** Dimerization of the minimal DIS transcript 397/444 and heterodimer formation. Transcript 397/444 was incubated at optimal dimerization conditions and either heated in the presence of formamide loading buffer and loaded on gel (lane 1) or loaded directly with native loading buffer (lane 2). The positions of the monomer (M1) and dimer (D1) bands are indicated. Two nonspecific faint bands are seen, which were copurifed with transcript 397/444. Lanes 3–5 contain dimer samples of the transcripts 397/444 and 1/444 and the mutant 1/444 GUGCAC, respectively. Monomers (M) and homodimers (D) are marked for the short and long transcript. In lanes 6 and 7, short and long transcripts were mixed before dimer formation. The heterodimer formed between 397/444 and 1/444 is indicated by D12 (lane 6).
G-U base pair into a stronger A-U base pair. A true covariation is present in the top part of the stem in SIV<sub>smm</sub> isolates 6–10, where a U-A base pair is changed into C-G. Destabilization because of a nucleotide substitution is observed in SIV<sub>smm</sub> isolate PBJ1, where a G-C base pair in the middle part of the stem is changed to a G-A mismatch. Most of the sequence differences between virus isolates target single stranded nucleotides, and G to A changes predominate in these purine-rich loop and bulge elements. As mentioned previously, the loop of the DIS hairpin in SIV<sub>stm</sub> differs in that the entire palindrome sequence has been deleted. Interestingly, inspection of the flanking leader sequences revealed the presence of a downstream 6-mer palindrome that is unique for the SIV<sub>stm</sub> sequence. This palindrome is flanked by sequences containing two additional substitutions that are specific for SIV<sub>stm</sub> which allows the formation of a small hairpin structure (Fig. 8; SIV<sub>stm</sub>-specific nucleotide variation is marked by black boxes). Although speculative, this motif may serve as an alternative DIS element in the genome of SIV<sub>stm</sub> to compensate for the deletion of the original palindrome.

DISCUSSION

We produced a set of transcripts encoding different regions of the untranslated leader of the HIV-2 RNA genome that have been implicated in RNA dimerization. Whereas dimer formation of transcripts that include the DIS hairpin structure (e.g. 1/444) was strongly induced in the presence of MgCl<sub>2</sub>, this was not the case for transcripts encompassing the TAR hairpin (1/124) or transcripts encoding both the TAR hairpin and the PBS region (1/379). A very low level of dimers is formed by these transcripts. This may have been the focus of attention in other studies, in which dimerization was performed at much higher RNA concentrations. With the DIS-containing transcripts, we obtained dimerization levels of at least 20% in a buffer with 5 mM MgCl<sub>2</sub> and 40 mM NaCl and with an RNA concentration as low as 35 nM. In a previous study, it was suggested that the TAR hairpin domain is the determinant for dimerization of HIV-2 RNA (9). However, both the high RNA and NaCl concentrations (300 mM compared with 40 mM in the present study) promote nonspecific RNA interactions, probably accounting for the observed “TAR dimers”. Such nonspecific RNA interactions at high concentrations of monovalent cations were also observed in this study. For instance, incubation of RNA in TEN buffer (100 mM NaCl) induces a low level of nonspecific dimerization of all HIV-2 transcripts (Fig. 3, lanes 1, 5, 9, 13, 17, and 21). Similar monovalent cation-triggered dimerization effects have been described in studies with HIV-1 RNA that reported increased dimer formation at high KCl concentrations (16, 34).

It has been reported recently that a palindromic sequence within the PBS is the determinant for in vitro dimerization of HIV-2 RNA (30). The authors of that study suggested that the HIV-2 leader RNA dimerizes without the DIS hairpin, which is not consistent with our observations. Compared with our assay conditions, the concentrations of both the RNA transcripts and the monovalent cations are highly elevated in the study of Jossinet et al. (30), which is likely to induce nonspecific RNA interactions. We think that the assay conditions used in the other studies favor the formation of DIS-independent RNA interactions, which correspond to the background levels measured in our assay system. Obviously, it remains possible that the TAR and PBS domains act as secondary dimerization sites (26), but our data demonstrate that the DIS motif is critical for the initial RNA-RNA contact and dimerization.

We have confirmed that the DIS hairpin is necessary for dimer formation by annealing of a series of antisense DNA oligonucleotides to the 1/444 transcript. Only the oligonucleo-
tides that target the DIS region are inhibitory to dimerization. We observed no effects with oligonucleotides that target either the TAR hairpin or the palindrome within the PBS. In addition, annealing of the natural tRNA<sub>Lys-3</sub> primer onto the PBS did not inhibit dimer formation (results not shown).

The minimal 397/444 transcript, which encodes the DIS hairpin and some flanking nucleotides, dimerizes effectively even at a low RNA concentration. In a heterodimer assay, we measured that the minimal DIS domain 397/444 dimerizes with the same efficiency as the full-length 1/444 transcript. This challenges the results of Jossinet et al. (30), who measured efficient dimer formation with the DIS hairpin of the HIV-1 genome but not with the isolated HIV-2 DIS hairpin. We think that this HIV-2 DIS dimerization defect is due to the use of a PCR primer with a sequence that only partially corresponds with the actual HIV-2 leader sequence. Mfold analysis of the RNA containing this non-HIV-2 sequence predicts a secondary structure in which the DIS palindrome sequence is base paired and therefore unable to participate in a loop-loop kissing interaction (result not shown).

Having established that the HIV-2 DIS hairpin is sufficient for dimer formation, we tested the influence of the palindrome sequence composition. The results with the substitution and deletion mutants prove that the presence of a 6-mer palindrome within the DIS loop is critical for <i>in vitro</i> dimerization, which is consistent with several HIV-1 studies (14, 17, 35, 36). The dimerization efficiency of the palindrome variants GCGGCG, GUGCAC, and GUCGAC is quite similar to that of the wild type GUUACC, suggesting that there is no strict palindrome sequence requirement for dimer formation. However, not all palindromic sequences support RNA dimerization. We identified two poorly dimerizing palindromes, UGUACA and GUUAAC, which share the property that their sequences contain four A/U nucleotides. Because A-U base pairs are weaker than G-C base pairs, we propose that the palindrome must contain two or less A/Us to promote efficient dimerization through base pairing. In support of this rule, the mutant palindrome CUUAAG was also shown to be inactive in dimerization in the context of HIV-1 RNA (36). The importance of a high G/C content within the DIS palindrome has also been established in studies that compared the HIV-1 isolates LAI (GGCGCG) and MAL (GUUACC), of which the latter was less efficient in dimer formation (15, 16, 34). In SELEX experiments that selected RNA molecules that dimerize efficiently, the vast majority of the recovered sequences were also G/C-rich (37). This further supports the importance of a relatively strong base pairing interaction for DIS-mediated dimerization.

It should also be noted that the sequence and/or structural context of any palindrome is likely to influence its dimerization ability. For instance, whereas the unnatural palindrome GCCGAC is highly efficient in the HIV-2 DIS, it has been shown to function poorly in an HIV-1 background (38). Likewise, the two inactive palindromes UGUACA and GUUAAC with four A/U nucleotides are found in the DIS of naturally occurring virus isolates of SIV<sub>syk</sub> and SIV<sub>mandr</sub>, respectively (26). Assuming that these DIS elements function in RNA dimerization in their natural context, it will be of interest to map putative enhancer elements that allow the use of such A/U-rich palindromes in these viruses. An alternative way of enhancing the base pairing capacity is by extension of the palindrome. For instance, a 10-mer palindrome was suggested to control dimerization of the foamy virus RNA genome (39).

The heterodimerization assay in Fig. 6 shows that no heterodimers are formed between a transcript with the wild type palindrome GGUACC and a transcript with the GUGCAC palindrome. Obviously, the loss of four base pairs abolishes the kissing-loop complex that is held together by six base pairs in the wild type RNA (Fig. 9). However, in the context of the extended or tight dimer, the loss of these central base pairs is not expected to be detrimental. Computer-assisted RNA folding indicates that a large internal loop is formed. The number of base pairs and free energy of the dimer interactions are indicated. All palindromes are marked by boxes. All free energy calculations were made using the Mfold algorithm on model hairpins that mimic the dimer interaction (kissing-loop and extended duplex) as closely as possible. A, secondary structure model of the kissing-loop complex formed between HIV-2 ROD wild type DIS hairpins and the extended duplex. B, schematic kissing-loop complex and extended dimer of an A/U-rich palindrome mutant. C, heterodimer formation is blocked at the level of the kissing-loop complex, although a reasonably stable extended duplex with a large internal loop can be folded.
The DIS palindrome must be exposed in the context of the full-length HIV-2 leader RNA to initiate dimerization. Using enzymatic and chemical probing, we verified this in the context of transcript 1/544, which encodes the entire leader RNA to the start codon of the gag open reading frame (Fig. 7). The majority of the nucleotides within the DIS loop can be modified by single strand-specific reagents, indicating that they are not base paired. Similar results have been described for the single strand-specific RNAses T1 and T2 (40). The structure probing also suggests an extension of the DIS stem by five additional base pairs. Extension of the DIS stem confers increased stability, which may restrict the transition from kissing-loop complex to tight dimer. However, the extended hairpin structure contains a large internal loop that modulates the overall stability. This loop has a high A/G content, a feature that is shared by the leader RNA. It has been reported that interactions between adenosines and content as a common feature in all single stranded regions (29). For HIV-1, it has recently been shown that the DIS palindrome in vivo is masked within the complete leader RNA by a long leader RNA.

It has been reported that interactions between adenosines and helix minor grooves lead to the formation of compact globular structures within the single stranded regions of group I ribozymes (42). A similar phenomenon may occur in the HIV-2 leader RNA.

For HIV-1, it has recently been shown that the DIS palindrome is masked within the complete leader RNA by a long distance base pairing interaction with the upstream poly(A) region (43). This may serve as a mechanism to prevent premature genome dimerization in the virion-producing cell. Two alternative conformations of the leader RNA were proposed, which either support or restrict dimerization. We did not observe a similar inhibition of dimerization in HIV-2 RNA. DIS-containing transcripts dimerize readily, independent of the presence of the upstream poly(A) region. Furthermore, antisense oligonucleotides that anneal to the poly(A) region do not enhance dimerization, as was found for HIV-1 RNA. Because we assume that HIV-2 RNA dimerization must also be regulated in vivo, we are currently investigating whether the HIV-2 RNA genome contains alternative DIS silencer elements.

Acknowledgments—We thank Wim van Est for artwork.

REFERENCES
1. Coffin, J. M. (1984) in RNA Tumor Viruses (Weiss, R., Teich, N., Varmus, H., and Coffin, J., eds) Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
2. Greatorex, J. S., and Lever, A. M. L. (1995) J. Gen. Virol. 76, 3467–3474
3. Huthoff, H., and Berkhout, B. (1998) J. Virol. 72, 2877–2882
4. Huthoff, H., Ohaen, A., Goncalves, J., Panganiban, A. T., and Gabius, H. J. (1997) Virol. 233, 271–279
5. Bender, W., and Davidson, N. (1976) Cell 7, 595–607
6. Darlix, J.-L., Gabius, H., Nugeyere, M. T., Clavel, F., and Barre-Sinoussi, F. (1990) J. Mol. Biol. 216, 689–699
7. Gotte, M., Li, X., and Wainberg, M. A. (1999) Arch. Biochem. Biophys. 365, 199–210
8. Awang, G., and Sen, D. (1993) Biochemistry 32, 11453–11457
9. Berkhout, B., Oude Essink, B. B., and Schoneveld, I. (1993) FASEB J. 7, 185–187
10. Bieth, E., Gabius, H., and Darlix, J.-L. (1990) Nucleic Acids Res. 18, 119–127
11. Clever, J., Sassetti, C., and Parslow, T. G. (1995) J. Virol. 69, 2101–2109
12. Clever, J. L., Wong, M. L., and Parslow, T. G. (1996) J. Virol. 70, 5902–5908
13. Greatorex, J. S., Laisse, V., Doherty, E. A., and Berkhout, B. (1996) Nucleic Acids Res. 24, 2919–2923
14. Haddrick, M., Lear, A. L., Cann, A. J., and Heaphy, S. (1996) J. Mol. Biol. 259, 55–68
15. Laughaire, M., and Jette, L. (1994) Biochem. 33, 13464–13474
16. Marquet, R., Baudin, F., Gabius, H., Darlix, J.-L., Mougel, M., Ehresmann, C., and Ehresmann, B. (1991) Nucleic Acids Res. 19, 2349–2357
17. Paillart, J. C., Marquet, R., Skripkin, E., Ehresmann, B., and Ehresmann, C. (1999) J. Biol. Chem. 274, 27486–27493
18. Skripkin, E., Paillart, J. C., Marquet, R., Ehresmann, B., and Ehresmann, C. (1994) Proc. Natl. Acad. Sci. U. S. A. 91, 4945–4949
19. Sourdiquet, W. I., and Heaphy, S. (1993) Proc. Natl. Acad. Sci. U. S. A. 90, 3383–3387
20. Takaku, H., Yamasaki, T., and Yashinaka, Y. (1993) J. Virol. 67, 1830–1839
21. Clever, J. L., and Parslow, T. G. (1997) J. Virol. 71, 3407–3414
22. Muriaux, D., Girard, P.-M., Bonnet-Mathoniere, B., and Paolotti, J. (1995) J. Biol. Chem. 270, 19290–19299
23. Lauhaire, M., and Jette, L. (1996) Biochemistry 35, 1589–1598
24. Muriaux, D., De Roquigny, H., Roques, B., and Paolotti, J. (1996) J. Biol. Chem. 271, 33688–33692
25. Doherty, E. A., Batey, R. T., Masquida, B., and Doumdjian, J. A. (2001) Nat. Struct. Biol. 8, 339–343

H. Huthoff and B. Berkhout, unpublished results.
