The genome of all retroviruses consists of two homologous RNA molecules noncovalently linked near their 5' end (1-4). The functional role of the dimer linkage structure (DLS) remains unclear. However, it is probably crucial since it exists in all retroviruses. It has been proposed to be involved in the regulation of several key steps in the retroviral cycle, such as encapsidation (5, 6), reverse transcription (12-14), and recombination during reverse transcription (12-14). Therefore, sequences involved in the dimerization of the retroviral genomes are most likely good targets for sense or antisense agents, and their exact localization is the first step toward potential therapeutic action.

Data from the S.D. site, splice donor site; DIS, dimerization initiation site; PCR, polymerase chain reaction.

**Mutational Analysis of the Bipartite Dimer Linkage Structure of Human Immunodeficiency Virus Type 1 Genomic RNA**

(Received for publication, July 19, 1994)

Jean-Christophe Paillart, Roland Marquet‡, Eugene Skripkin§, Bernard Ehresmann, and Chantal Ehresmann

From the Unité Propre de Recherches No. 9002 du Centre National de la Recherche Scientifique, Institut de Biologie Moléculaire et Cellulaire, 15 rue R. Descartes, 67084 Strasbourg-cedex, France

The genome of all retroviruses consists of two homologous RNA molecules associated near their 5' end in a region called the dimer linkage structure. Dimerization of genomic RNA is thought to be important for several functions of the retroviral cycle such as encapsidation, reverse transcription, and translation. In human immunodeficiency virus type 1 (HIV-1), a region downstream of the splice donor site was initially postulated to mediate dimerization. However, we recently showed that the dimerization initiation site is located upstream of the splice donor site and suggested that dimerization may initiate through a loop-loop interaction. Here, we show that single base mutations in the palindromic loop of the dimerization initiation site completely abolish dimerization, while introduction of compensatory mutations restores the process. Furthermore, two single nucleotide mutants that are unable to form homodimers efficiently codimerize, while the wild type RNA and the compensatory mutant, which both form homodimers, are unable to codimerize. These results unambiguously prove the interaction between the palindromic loops of each monomer. By contrast, none of the deletions that we introduced downstream of the splice donor site abolishes dimerization. However, deletions of two purine tracts located in the vicinity of the initiation codon of the gag gene significantly decrease the thermal stability of the HIV-1 RNA dimer.

Retroviral genomes consist of two homologous RNA molecules noncovalently linked near their 5' end (1-4). The functional role of the dimer linkage structure (DLS) remains unclear. However, it is probably crucial since it exists in all retroviruses. It has been proposed to be involved in the regulation of several key steps in the retroviral cycle, such as encapsidation (5, 6), reverse transcription (12-14), and recombination during reverse transcription (12-14). Therefore, sequences involved in the dimerization of the retroviral genomes are most likely good targets for sense or antisense agents, and their exact localization is the first step toward potential therapeutic action.

Dimerization of retroviral RNAs can be induced in vitro by the nucleocapsid protein (7) and by cations (15, 16). In human immunodeficiency type 1 virus (HIV-1), the DLS was first localized within the 100 nucleotides downstream of the splice donor (S.D.) site (7, 15, 16). This proposed DLS is absent from the spliced mRNAs (Fig. 1A). From an earlier study of the in vitro dimerization of several RNA fragments containing this DLS, we concluded that dimerization of HIV-1 genomic RNA involves Watson-Crick base-pairing of their palindromic DIS sequences (21). On the contrary, deletion of the GGAGA consensus tracts in the 3'-DLS does not inhibit dimerization of a HIV RNA that contains the DIS (21). These results rise the possibility that the dimers formed by RNAs containing the 3'-DLS but missing the sequences located upstream of the S.D. site differ from the dimers formed by RNAs containing both the upstream and downstream sequences (20).

In this paper, we studied the relative contributions of the DIS and the 3'-DLS in the HIV-1 RNA dimerization process by extensive site-directed mutagenesis. In a previous report, we proposed that the initiation step of HIV-1 RNA dimerization is the annealing of the two monomers through Watson-Crick base-pairing of their palindromic DIS sequences (Fig. 1B) (21). Here, we unambiguously demonstrated a direct interaction between the DIS present on each monomer during dimerization. We also introduced a number of mutations in the 3'-DLS. None of the deletions in this region was found to inhibit dimerization. However, we identified two purine-rich sequences involved in the stabilization of the RNA dimer.

**EXPERIMENTAL PROCEDURES**

Plasmid Construction and RNA Synthesis—Plasmid construction and restriction enzyme digestions were conducted according to...
FIG. 1. A, scheme of the 5' terminal nucleotides of HIV-1 RNA. The main RNA elements are represented: TAR, the trans-activation responsive element; PBS, the replication primer tRNA<sup>ax2</sup>-binding site; DIS, the dimerization initiation site; E/3'-DLS, the elements necessary for the in vivo encapsidation of the unspliced viral RNA and the previously proposed dimer linkage structure, and AUG, the initiator codon for synthesis of gag and gag-pol polyproteins. B, secondary structure model of the DIS (nucleotides 256-294 of HIV-1 (Mal isolate) in monomeric and dimeric forms as proposed by Skripkin et al. (21). The nucleotides proposed to interact during the initiation of dimerization are encircled. Postulated structures of the monomer (top) and of the dimer (bottom).

standard procedures (22). Plasmid pJCB was obtained by inverse PCR on the plasmid pHIVG-4 (15). The PCR primers were designed to eliminate the 615-4005 HIV-1 region of pHIVG-4 and allow a self-ligation of the PCR products after restriction by Sool. The resulting pJCB plasmid also contains a new PvuII restriction site at the 3' end of the HIV-1 sequence. Plasmids containing the mutations described in Fig. 2 were obtained by inverse PCR on the plasmid pJCB. The PCR products were gel purified, phosphorylated, and used to transform JM109 Escherichia coli cells after ligation.

All plasmids were digested with PvuII before transcription to obtain RNAs corresponding to nucleotides 1-615 of the Mal isolate of HIV-1 (+1 being the first nucleotide of the genomic RNA) (23). Alternatively, pUCB was linearized with RsaI in order to obtain RNA 1-311. RNA synthesis and purification was as described previously (15). Labeled RNAs were synthesized using [a<sup>32</sup>P]ATP (Amersham) (50 nCi of DNA template); alternatively, RNAs were labeled at their 3' end with [a<sup>32</sup>P]pCp and RNA ligase from phage T4.

In Vitro Dimerization of HIV-1 RNAs—In a typical experiment, wild type or mutant unlabeled RNA was diluted in Milli-Q (Millipore) water at 400 nM final concentration together with the corresponding radioactive RNA (3-5 nCi, 0.01-0.04 pg), heated for 2 min at 90 °C, and chilled for 2 min on ice. Dimerization was initiated by addition of 2 μl of 5-fold concentrated buffer (final concentration: 50 mM sodium cacodylate, pH 7.5, 300 mM KCl, and 5 mM MgCl<sub>2</sub>), the samples were incubated at the appropriate temperature, then cooled for 2 min on ice. The dimerization yield of wild type and mutant RNAs was compared after a 30-min incubation at 37 °C. In the kinetic experiments, incubation was at 37 °C for 30 s to 90 min. For the RNA dimer stability experiments, the samples were incubated for 30 min at 30 °C, then the temperature was gradually increased by 7 °C steps. After a 5-min incubation at the appropriate temperature, an aliquot was loaded on gel after addition of glycerol (20% final concentration).

Samples were analyzed on 1.1% agarose gels. Electrophoresis were performed in 45 mM Tris borate, pH 8.3, 0.1 mM MgCl<sub>2</sub>. Agarose gels were fixed for 10 min in 10% trichloroacetic acid and dried for 40 min under vacuum at room temperature. Bands corresponding to the monomer and dimeric species were quantified using a BAS 2000 Bio-Imager (Fuji).

All experiments were conducted at least twice (usually three or four times), and the error ranges given in the tables correspond to two standard deviations. In the case of thermal denaturation experiments, the T<sub>m</sub> value of the wild type RNA is slightly higher that the earlier published value (20). The origin of this difference is unknown. In order to ensure that all T<sub>m</sub> data published in the present paper are directly comparable, the denaturation experiments were performed at the same time with three mutants and the wild type 1-615 RNA, as internal control, and all samples were loaded on the same agarose gel.

RESULTS

We used extensive mutagenesis to study the relative contributions of the DIS and the 3'-DLS in the dimerization process of HIV-1 genomic RNA. For each mutant RNA, we analyzed the dimerization yield and, when significant dimerization was ob-
erved, the dimerization kinetics and the thermal stability of the RNA dimer. All mutations were studied in the same context, i.e. HIV-1 RNA 1-615, in order to avoid the artifacts that can occur when RNAs with different 3' or 5' ends are used (20, 21). Also, since HIV-1 genomic RNA contains both the DIS and the 3'-DLS, RNA 1-615 is a better model of the genomic RNA than RNA fragments restricted to the 3'-DLS.

**Mutations in the 3'-DLS**—We showed that the sequences located downstream of the S.D. site increases the thermal stability of the HIV-1 RNA dimer (Table I) (20). Interactions downstream of the S.D. site may involve the consensus RGGARA tracts that are found in the putative dimerization region of most retroviruses (15, 16). First, we constructed mutants to test the role of these consensus purine tracts located downstream of the S.D. site (Fig. 2, mutants DLSΔ338-349, DLSΔ377-385, DLSΔ396-404, and DLSΔ419-433).

Deletion of the first, second, or fourth purine tract has only minor effects on the dimerization yield, while deletion of the third purine tract (DLSΔ396-404) more severely decreases the dimerization of the corresponding RNA (Table I). However, dimerization is far from being abolished as observed with the DIS mutants (see below). Analysis of the dimerization kinetics and the thermal stability of the RNA dimers indicates little influence of the deletion of the second, the third, or the fourth consensus purine tract. The only effect is a 30-40% decrease of the dimerization kinetics when the third or fourth purine tracts are deleted, while the thermal stability of the corresponding dimers are unaffected (Fig. 3 and Table I). By contrast, deletion of the first purine tract (DLSΔ338-349) reduces the melting temperature of this mutant RNA dimer by 6 °C, and the kinetic constant is decreased by 23% compared to the wild type RNA (Fig. 4 and Table I). These results indicate that the first consensus purine tract may be involved in dimerization of HIV-1 genomic RNA. However, the rather small effects produced by the deletion of the purine tracts are not sufficient to definitively prove the direct involvement of the consensus tracts in the dimerization process.

Another possibility is that some of the purine tracts play indeed an important stabilizing role in the dimer, but since the tracts are redundant, the missing tract could be replaced by the remaining ones. To test this redundancy hypothesis, we simultaneously deleted the four consensus purine tracts (DLSΔ4CONSSENS). As already observed (21), this mutant significantly dimerizes (Table I). Indeed, the dimerization yield of this mutant is similar to that of mutant DLSΔ419-433. However, the thermal stability of the dimer and the dimerization kinetics of mutant DLSΔ4CONSSENS are more affected than those of any mutant containing a single purine tract deletion (Figs. 3 and 4, and Table I). The kinetic constant is decreased by 50%, and the thermal stability of the dimer by 12 °C (Figs. 3 and 4). Indeed, the thermal stability of the dimer formed by the mutant DLSΔ4CONSSENS RNA is the same as the thermal stability of the dimer formed by an RNA lacking the entire 3'-DLS (RNA 1-311WT, Table I). Thus, it seems that the consensus purine tracts are functionally redundant and that their simultaneous deletion completely abolishes the stabilization of the dimer provided by the sequences located downstream of the S.D. site.

If, as proposed in early studies (15, 17, 19, 23), dimerization of HIV-1 RNA involves purine quartets, some non-consensus G-rich sequences may also be implicated in the dimerization process. Potential interaction sites are the sequences GGA-GG335, GGG355, and GAGAG362. The GGAGGG335 sequence is located at the apex of a stem-loop structure (5, 24) and may therefore be easily accessible for tertiary interactions. This stem-loop structure was proposed to be an important element of the packaging signal of HIV-1 RNA (25). Given the possible link between dimerization and encapsidation (6-11), we constructed mutants containing either a deletion of the complete stem-loop (Fig. 2, mutant DLSΔ326-339) or a substitution of the GAGGG335 loop sequence (Fig. 2, mutant DLSΔ331-334). These mutations affect the dimerization yield but have no significant effect on the thermal stability of the dimer nor on the dimerization kinetics (Table I).

Sequences GGG355 and GAGAG362 are located at the bottom and at the apex of a small stem-loop immediately following the initiation codon of the gag gene, respectively (5). The negative effect of RNA dimerization on the translation of the gag gene (5) would be easily explained if these sequences are involved in the dimerization process. We either deleted the entire stem-loop (Fig. 2, mutant DLSΔ351-366) or substituted the GAGA sequence (mutant DLSΔ358-361). The thermal stability of both mutant RNAs is significantly decreased (Fig. 4B and Table I), suggesting that the GAGAG362 sequence may play a direct role in the stability of the RNA dimer. The dimerization yield of both mutants is strongly decreased, but the effect is more pronounced with the substitution than with the deletion. Opposite effects are observed on the kinetic constant of dimerization: the kinetics is decreased by the deletion but increased in this substitution mutant (Table I).

Other elements potentially involved in the dimerization of HIV-1 RNA are a small stem-loop (nucleotides 305-316) encompassing the splice donor site (3' of nucleotide 305) (5, 26), and an U tract located immediately downstream of this stem loop (nucleotides 319-332) that was cross-linked to the nucleocapsid protein (7). Deletions in these regions drastically reduced packaging of the genomic HIV-1 RNA into the viral particles (27-29). Although both deletions decrease the dimerization yield (Table I, mutants DLSΔ306-317 and DLSΔ319-323), they do not significantly affect the thermal stability of the RNA dimer. Surprisingly, the dimerization rate is 2-fold enhanced by deletion of the splice donor site stem-loop.

**Mutations in the DIS**—The initial step of dimerization of HIV-1 RNA was proposed to be the annealing of the central perfect 12-mer DIS, which is present in a loop on each RNA monomer, through Watson-Crick base pairing, followed by the subsequent melting and annealing of the monomer stems (21). In the dimeric structure, the two DIS sequences form three helices separated by internal loops (Fig. 1B). Mutants DLSΔ265-287, DLSΔ279-287, DISS274-276, DISA276, DISC275, DISG278, and DISCOMPENS were designed to further test the involvement of the DIS sequence in the dimerization process of HIV-1 RNA.

---

**Table I**

<table>
<thead>
<tr>
<th>RNA</th>
<th>Dimerization yield (%</th>
<th>Thermal stability °C</th>
<th>k_monomer/k_dimer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-615 wt</td>
<td>69</td>
<td>52.1</td>
<td>1.00</td>
</tr>
<tr>
<td>1-611 wt</td>
<td>41</td>
<td>42</td>
<td>1.35</td>
</tr>
<tr>
<td>DLSΔ306-317</td>
<td>44</td>
<td>53</td>
<td>2.27</td>
</tr>
<tr>
<td>DLSΔ319-323</td>
<td>52</td>
<td>52</td>
<td>1.03</td>
</tr>
<tr>
<td>DLSΔ326-339</td>
<td>38</td>
<td>55</td>
<td>1.25</td>
</tr>
<tr>
<td>DLSΔ338-349</td>
<td>65</td>
<td>46</td>
<td>0.77</td>
</tr>
<tr>
<td>DLSΔ351-366</td>
<td>50</td>
<td>46</td>
<td>0.64</td>
</tr>
<tr>
<td>DLSΔ377-385</td>
<td>61</td>
<td>52</td>
<td>1.18</td>
</tr>
<tr>
<td>DLSΔ396-404</td>
<td>43</td>
<td>50</td>
<td>0.67</td>
</tr>
<tr>
<td>DLSΔ419-433</td>
<td>57</td>
<td>49</td>
<td>0.61</td>
</tr>
<tr>
<td>DLSΔ4CONSSENS</td>
<td>59</td>
<td>40</td>
<td>0.50</td>
</tr>
<tr>
<td>DLSΔ331-334</td>
<td>41</td>
<td>52</td>
<td>0.83</td>
</tr>
<tr>
<td>DLSΔ358-361</td>
<td>37</td>
<td>41</td>
<td>1.39</td>
</tr>
</tbody>
</table>

*This T<sub>m</sub> is slightly higher than previously determined (20). See "Experimental Procedures" for details.

---

---
Fig. 2. HIV-1 RNA mutants synthesized in vitro. A, localization of the mutations in the 3' DLS region (311-440). B, localization of the deleted (-) or substituted (underlined) sequences in the DIS region (260-290). These mutations are also indicated on the monomer model of the DIS.
A-U base pairs are replaced by G-U pairs (Fig. 2, DISG278), and when two compensatory mutants RNA displays no significant differences neither in the thermal stability of the dimer, as was proposed in Fig. 2B.

**DISCUSSION**

Our results clearly show the importance of the DIS in the dimerization process of HIV-1 RNA, in vitro, and confirm the mechanism of dimerization initiation that we recently proposed (Fig. 1B) (21). Not only deletions, but also single base substitutions totally abolish or strongly reduce dimerization of 1–615 RNA. This is consistent with our previous results, since we initially identify the DIS by chemical interference experiments (21). In this technique, at most one nucleotide is statistically modified per RNA molecule, so that the DIS could be found only if one modification in the DIS significantly decreases the dimerization yield.

The fact that a compensatory mutation in the palindromic sequence of the DIS restores dimerization is in total agreement with the dimerization scheme that we previously proposed (21). Furthermore, we now unambiguously demonstrated the interaction of the central palindromic sequence of the DIS within the dimer by testing the formation of heterodimers with mutant RNAs. These results demonstrate that we can favor the formation of either homo- or heterodimers. We are presently investigating whether the DIS is also the major determinant of HIV-1 genomic RNA dimerization ex vivo. If it is the case, mutations of the DIS will be valuable tools to either favor, or hinder, copackaging of mutated RNAs in cells simultaneously transfected or infected with several retroviral vectors. Thus, mutations in the DIS may be useful in controlling recombinant between these vectors.

Effects of the mutations in the 3’-DLS are not so dramatic than those observed with the DIS mutants: no deletion in the 3’-DLS abolishes dimerization of 1–615 RNA. Some of the mutations in the 3’-DLS have no detectable effects (DLSA377–386). For most of the mutants (DLSA396–404, DLSA326–359, DLSA331–354, DLSA306–317, and DLSA319–329), the dimerization yield, and in some cases the dimerization rate, are affected, but the thermal stability of the RNA dimer is not significantly affected. These results probably reflect an indirect conformational effect of the mutation, rather than a direct involvement of the mutated sequences, in the dimerization process. Chemical and enzymatic probing of the mutant RNA dimers indicates only limited structural perturbations in the vicinity of the mutated sites. The monomeric form of these mutants DZSA276 and DISA279–287 almost totally inhibit dimerization (Fig. 3B).  

**Fig. 3. Analysis of the dimerization kinetics.** The experimental dimerization data of RNAs (WT) 1–615, DLS A306–317, DLS A3419–433, and DLS A4CONSEN were introduced into equation $1/M_d = 1/M_0^2 + 2 M_0 k_{on} t$ corresponding to the second order “conformation model” (20).
**Fig. 4. Thermal stabilities of dimers of RNAs 1-615 (wild type) and mutants.** A, WT RNA 1–615 (lanes 1–7) and RNA DLS Δ4CONSEN (lanes 8–14) as example of experimental data. Melting experiments were conducted as described under “Experimental Procedures.” Lanes 1 and 8, 30 °C; lanes 2 and 9, 37 °C; lanes 3 and 10, 44 °C; lanes 4 and 11, 52 °C; lanes 5 and 12, 58 °C; lanes 6 and 13, 65 °C; and lanes 7 and 14, 72 °C. B, thermal stability curves of RNAs 1–615 (WT), DLS Δ338–349, DLS Δ351–366, DLS S358, and DLS Δ4CONSEN. For the sake of comparison, the dimerization yields were normalized according to the dimerization yield at 30 °C.

RNAs is structurally heterogeneous, as already observed for the wild type RNA monomer (5). We previously showed that the incomplete dimerization of HIV-1 RNA fragments is due to the presence of monomers that have a conformation incompatible with dimerization, rather than to the limited stability of the dimers (20). Thus, variations of the dimerization yield and kinetics among the mutants RNAs probably also reflect differences in the monomer structures.

On the contrary, important changes in the thermal stability of the dimers probably indicate that the mutated sequences are directly involved in the dimerization process. Only two tested regions downstream of the S.D. site are directly involved in the thermal stability of HIV-1 RNA dimer: one is located immediately upstream of the initiation codon of the gag gene, and the second one is located at the apex of the short stem-loop immediately 3' of the gag gene start. Interestingly, both sequences...
A, RNA 1-311WI' coincubated in dimer buffer with either lane 3 or dimer; the results obtained with mutant DISCOMPENS indicated the involvement of these sequences in G quartets is incompatible with the cation dependence of RNA 1-615 dimerization (20) and remains an open question. These alternating G and A sequences may also form other unusual structures such as parallel duplexes (31). Of course, they could also stabilize the dimer by classical Watson-Crick interactions.

Deletion of the second consensus purine tract, which was proposed to be involved in the dimerization of short RNAs restricted to the 3'-DLS (17, 18), has no effect neither on the dimerization of RNA 1-615, which also contains the DIS, nor on the thermal stability of the corresponding dimer. This result further stresses the importance of working with RNAs containing the whole 5' region of HIV-1 genomic RNA.

Although the DIS appears more important than the 3'-DLS in promoting RNA dimerization, the interactions involving sequences downstream of the S.D. site may be crucial for the functions linked to dimerization. Since these interactions involve sequences located immediately upstream and downstream of the initiation codon of the gag gene, they may easily explain the negative effect of the RNA dimerization on the translation of this gene that was previously reported (5). In addition, deletions downstream of the S.D. site drastically affect RNA packaging (27-29). Recombination and complete reverse transcription require strand transfer of reverse transcriptase (12, 14). This process should be greatly enhanced by an overall parallel orientation of the two genomic RNA molecules. Indeed, this orientation could be obtained by the intermediate of purine-mediated non-canonical interactions.

Finally, the stabilization of the dimer by the sequences downstream of the S.D. site that we observed in vitro may be involved in the maturation step observed in vivo. It has been recently shown that murine leukemia virus and HIV-1 RNA dimers undergo a protease-dependent maturation, following the release of the viral particles (30, 32). Although our data are not directly comparable with those of Fu et al. (30), one can notice that the Tm of the immature (less than 50°C) and mature (55°C) HIV-1 RNA dimers are quite close from the values that we determined for wild type RNAs 1-311 and 1-615, respectively.

Since we showed that the interactions in the 3'-DLS cannot take place within the context of a large RNA in the absence of a functional DIS, the DIS is probably a best target than the 3'-DLS for antiretroviral drugs such as sense or antisense agents.

Acknowledgments—C. Isel is acknowledged for fruitful discussions. D. Mignot is acknowledged for skillful technical assistance.

REFERENCES