

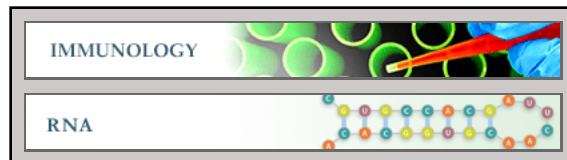
Immunology:

**Mechanism of HIV-1 RNA Dimerization in
the Central Region of the Genome and
Significance for Viral Evolution**

Dorota Piekna-Przybylska, Gaurav Sharma
and Robert A. Bambara

J. Biol. Chem. 2013, 288:24140-24150.

doi: 10.1074/jbc.M113.477265 originally published online July 9, 2013



Access the most updated version of this article at doi: [10.1074/jbc.M113.477265](https://doi.org/10.1074/jbc.M113.477265)

Find articles, minireviews, Reflections and Classics on similar topics on the [JBC Affinity Sites](https://www.jbc.org/).

Alerts:

- [When this article is cited](#)
- [When a correction for this article is posted](#)

[Click here](#) to choose from all of JBC's e-mail alerts

This article cites 47 references, 28 of which can be accessed free at
<http://www.jbc.org/content/288/33/24140.full.html#ref-list-1>

Mechanism of HIV-1 RNA Dimerization in the Central Region of the Genome and Significance for Viral Evolution*

Received for publication, April 15, 2013, and in revised form, July 1, 2013. Published, JBC Papers in Press, July 9, 2013; DOI 10.1074/jbc.M113.477265

Dorota Piekna-Przybylska[‡], Gaurav Sharma[§], and Robert A. Bambara^{‡1}

From the [‡]Department of Microbiology and Immunology, School of Medicine and Dentistry, University of Rochester, Rochester, New York 14642 and [§]Department of Electrical and Computer Engineering, University of Rochester, Rochester, New York 14627

Background: HIV-1 genome dimerization is important for viral survival.

Results: The central region of HIV-1 genome contains a G-rich sequence promoting recombination and capable of forming a dimeric G-quartet.

Conclusion: Parallel alignment and proximity of two HIV-1 genomes associated at the 5' end can be maintained through dimeric G-quartet formed in the central region of viral genome.

Significance: Determining mechanism and factors involved in HIV-1 genome dimerization.

The genome of HIV-1 consists of two identical or nearly identical RNA molecules. The RNA genomes are held in the same, parallel orientation by interactions at the dimer initiation site (DIS). Previous studies showed that in addition to interactions at DIS, sequences located 100 nucleotides downstream from the 5' splice site can dimerize *in vitro* through an intermolecular G-quartet structure. Here we report that the highly conserved G-rich sequence in the middle portion of the HIV-1 genome near the central polypurine tract (cPPT) dimerizes spontaneously under high ionic strength in the absence of protein. The antisense RNA does not dimerize, strongly indicating that RNA dimerization does not exclusively involve A:U and G:C base pairing. The cation-dependent reverse transcriptase pausing profile, CD spectra profile, and cation-dependent association and thermal dissociation characteristics indicate G-quartet structures. Different forms of G-quartets are formed including monomers and, significantly, intermolecular dimers. Our results indicate that RNA genome dimerization and parallel alignment initiated through interactions at DIS may be greatly expanded and stabilized by formation of an intermolecular G-quartet at a distant site near the cPPT. It is likely that formation of G-quartet structure near the cPPT *in vivo* keeps the RNA genomes in proximity over a long range, promoting genetic recombination in numerous hot spots.

HIV-1 efficiently diversifies its population within an infected person, allowing it to evade host immune response and therapeutic treatments. The error-prone HIV-1 reverse transcriptase (RT) contributes significantly to viral diversity. Viral evolution is further accelerated through the process of genetic recombination. The presence of two often non-identical copies

of viral genome in a single virion is a fundamental requirement for genetically significant recombination. One copy of genomic RNA is sufficient for viral replication. The role of the second copy is to allow the RT to switch templates during reverse transcription. This mixes mutations from two genetically different viral genomes, allowing the virus to evolve into more robust forms.

The two genomic RNAs adhere within the cytoplasm; thus, the viral genomes are packaged into the virion as a dimer rather than two RNA monomers (1, 2). The sequences involved in initial viral genome dimerization and encapsidation partially overlap and are localized close to the 5' end (3–5). One of these sequences forms the stem-loop 1 secondary structure with a highly conserved palindromic sequence called the dimer initiation site (DIS).² The first stable interaction is a contact within the loops of two stem-loop 1 hairpins called RNA kissing (6, 7). The interaction is then thought to spread to the stem regions, with their palindromic sequences promoting base pairing; however, this contact is not sufficient by itself to produce dimers *in vivo* that resist extraction and gel electrophoresis (8). Other interactions along the genome seem to be desirable to keep a consistent proximity that would facilitate a generally uniform probability of template switching of the RT throughout reverse transcription, desirable for efficient recombination.

Studies *in vitro* show that in HIV-1 and other retroviruses, certain G-rich genomic sequences can form G-quartet structures (9–13). Four guanine bases can associate through Hoogsteen hydrogen bonding and form a G-tetrad, and two or more guanine tetrads can stack on top of each other to form a G-quartet. The G-tetrads can form by repeated folding of either a single nucleic acid molecule or by association of two or four molecules. The structure has been intensively analyzed in telomeres, and in recent years G-quartets have been shown to regulate activity of promoters, formation of DNA flaps, stability of mRNA, and alternative splicing and translation (14–21). Very little is known about G-quartets in viruses, including

* This work was supported, in whole or in part, by National Institutes of Health Grants GM049573 (to R. A. B.) and GM097334 (to G. S.) and University of Rochester Developmental Center for AIDS Research Grant P30 AI078498 (NIAID). This work was also supported by the University of Rochester School of Medicine and Dentistry.

¹ To whom correspondence should be addressed: Dept. of Microbiology and Immunology, School of Medicine and Dentistry, University of Rochester, 601 Elmwood Ave., Rochester, NY 14642. Tel.: 585-275-2764; Fax: 585-275-6007; E-mail: Robert_bambara@urmc.rochester.edu.

² The abbreviations used are: DIS, dimer initiation site; cPPT, central polypurine tract; NC, nucleocapsid protein; TP, transfer product; DE, donor extension product; nt, nucleotide(s).

HIV-1 and their possible function in the viral life cycle. G-quartet structure formation was found to be essential to maintain the Epstein-Barr virus genome during latency in proliferating cells (22). In HIV-1 and Moloney murine sarcoma virus (MuSV), G-rich sequences in the *gag* region near DIS were shown to form an intermolecular G-quartet (9–13). This structure adheres the two RNA genomes; thus, it represents a different mode of genome dimerization in retroviruses from that of the DIS hairpin interaction. Formation of dimeric G-quartet structure correlates with hot spots of recombination, exhibiting an increased rate of template switching (23, 24).

Here we show that short RNA templates from the central region of the HIV-1 genome, containing the G-rich sequences near the central polypurine tract (cPPT), are able to form both monomer and dimer G-quartet structures. The cPPT is located in the integrase gene and is a region where one of two primers for synthesis of (+) strand DNA is produced during reverse transcription. The G tract in the cPPT and two other G runs located downstream from the cPPT are highly conserved among different HIV-1 isolates and several closely related simian immunodeficiency virus (SIV) species. By application of assays *in vitro* and affinity selection methods, we found that the RNA templates dimerize through the G-rich regions, indicating that in addition to contacts near the 5' end, the central regions of the viral RNA genomes are likely to be maintained in proximity through dimer G-quartet formation. In reconstituted systems of reverse transcription, the formation of G-quartets facilitates RT in switching templates during synthesis of minus strand DNA, suggesting that the structure supports an increased recombination rate.

EXPERIMENTAL PROCEDURES

Materials—DNA oligonucleotides and the HPLC-purified RNA strand used for CD spectra analyses were purchased from Integrated DNA Technologies, Inc. (Coralville, IA). HIV-1 nucleocapsid protein (NC; 55 amino acids) was generously provided by Dr. Robert J. Gorelick. HIV-1 reverse transcriptase (p66/p51 heterodimer) was purified as described previously (25). The [γ - 32 P]ATP was purchased from PerkinElmer Life Sciences. The sequences of viruses were analyzed with the HIV database (www.hiv.lanl.gov).

Preparation of RNA Templates—RNA molecules were transcribed *in vitro* (Ambion T7-MEGAShortscript kit; Applied Biosystems) from DNA templates amplified by PCR using *Vent* DNA polymerase (New England Biolabs) and two overlapping oligomers with the sequence of the desired region. The following RNA strands were used in our studies. (a) For the reverse transcription assay, the wild-type and mutant RNA template strands with the cPPT region (4309–4396 in the RNA genome) sequences from the NL4-3 HIV-1 were made from DNA templates synthesized with a pair of oligomers, 1/2 and 3/4, respectively. (b) For affinity selection analysis, the non-tagged RNAs and poly(A)-tagged RNAs were made from DNA generated using oligomers 1/2 and 1/5 for the cPPT region (4309–4396) of NL4-3 HIV-1, 6/7 and 6/8 for the *gag* region (290–403) of NL4-3 HIV-1, 9/10 and 9/11 for the *gag* region (303–415) of MAL HIV-1, and 12/13 and 14/15 for 5'-UTR with DIS (1–520 and 183–520) of NL4-3 HIV-1. (c) For analysis of dimerization

in a native gel, the RNA template of the cPPT region (4309–4396) in NL4-3 HIV-1 was made from DNA generated using oligomers 1/2, and for the antisense sequence of the cPPT region, oligomers 16/17 were used. (d) For transfer assays, the donor and acceptor templates for cPPT region (4309–4396) of NL4-3 HIV-1 were made from DNA generated using oligomers 1/5 and 18/2, respectively. After transcription *in vitro*, the RNA templates were purified by polyacrylamide/urea gel electrophoresis and resuspended in water. RNAs were quantitated by UV absorption using a GeneQuant II from Amersham Biosciences.

DNA Oligonucleotides—The forward oligonucleotides (1, 3, 6, 9, 12, 14, 16, 18) for all DNA templates used to generate RNAs were designed to contain the phage T7 RNA polymerase promoter sequence (GGA TCC TAA TAC GAC TCA CTA TAG) immediately upstream of the desired template sequence. The mutations and additions are underlined: **1**, 5'-GGA TCC TAA TAC GAC TCA CTA TAG GGC AGT ATT CAT CCA CAA TTT TAA AAG AAA AGG GGG GAT TGG GGG GTA CAG TG-3'; **2**, 5'-GTC TGT TGC TAT TAT GTC TAC TAT TCT TTC CCC TGC ACT GTA CCC CCC AAT CCC CCC-3'; **3**, 5'-GGA TCC TAA TAC GAC TCA CTA TAG GGC AGT ATT CAT CCA CAA TTT TAA AAG AAA AGG GGG GAT TGT GTG GTA CAG TG-3'; **4**, 5'-GTC TGT TGC TAT TAT GTC TAC TAT TCT TTT CTC TGC ACT GTA CCA CAC AAT CCC CCC-3'; **5**, 5'-TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT GTC TGT TGC TAT TAT GTC TAC TAT TCT TTC CCC TGC ACT GTA CCC CCC AAT CCC CCC-3'; **6**, 5'-GGA TCC TAA TAC GAC TCA CTA TAG GGT GAG TAC GCC AAA AAT TTT GAC TAG CGG AGG CTA GAA GGA GAG AGA TGG GTG CGA GAG CGT CGG TAT TAA GC-3'; **7**, 5'-GCC TTA ACC GAA TTT TTT CCC ATT TAT CTA ATT CTC CCC CGC TTA ATA CCG ACG CTC TCG CAC CCA T-3'; **8**, 5'-TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT GCC TTA ACC GAA TTT TTT CCC ATT TAT CTA ATT CTC CCC CGC TTA ATA CCG ACG CTC TCG CAC CCA T-3'; **9**, 5'-GGA TCC TAA TAC GAC TCA CTA TAG GGT GAG TAC GCC AAT TTT TGA CTA GCG GAG GCT AGA AGG AGA GAG ATG GGT GCG AGA GCG TCA GTA TTA AGC-3'; **10**, 5'-CCC TTA ACC GAA TTT TCT CCC ATG CAT CTA ATT TTC CCC CGC TTA ATA CTG ACG CTC TCG CAC CCA T-3'; **11**, 5'-TTT TTT TTT TTT TTT TTT TTT CCC TTA ACC GAA TTT TCT CCC ATG CAT CTA ATT TTC CCC CGC TTA ATA CTG ACG CTC TCG CAC CCA T-3'; **12**, 5'-GGA TCC TAA TAC GAC TCA CTA TAG GGT CTC TCT GGT TAG ACC AGA TC-3'; **13**, 5'-CCC AGT ATT TGT CTA CAG CCT TCT GAT GTC-3'; **14**, 5'-GGA TCC TAA TAC GAC TCA CTA TAG GCG CCC GAA CAG GGA CTT GAA AG-3'; **15**, 5'-TTT TTT TTT TTT TTT TTT TTT TTT TTT TCC CAG TAT TTG TCT ACA GCC TTC TGA TGT C-3'; **16**, 5'-GGA TCC TAA TAC GAC TCA CTA TAG GGT CAT AAG TAG GTG TTA AAA TTT TCT TTT CCC CCC TAA CCC CCC-3'; **17**, 5'-CAG ACA ACG ATA ATA CAG ATG ATA AGA AAG GGG ACG TGA CAT GGG GGG TTA GGG GGG AAA AGA AAA TT-3'; **18**, 5'-GGA TCC TAA TAC GAC TCA CTA TAG GGA AAA AAA AAA CAG TAT TCA TCC ACA ATT TTA AAA GAA AAG GGG GGA TTG GGG GGT ACA GTG-3'; **19**, 5'-GTC TGT TGC TAT TAT GTC-3'; **20**, 5'-TTT TTT

RNA Dimerization in the Central Region of the HIV-1 Genome

TTT TTT TTT TGT CT-3'; **a**, 5'-AAT CCC CCC TTT TCT T-3'; **b**, 5'-TAC CCC CCA ATC CCC C-3'; **c**, 5'-CTG CAC TGT ACC CCC C-3'; **d**, 5'-TCT TTC CCC TGC ACT G-3'; **e**, 5'-TCT ACT ATT CTT TCC C-3'; **f**, 5'-CTA TTA TGT CTA CTA T-3'; **g**, 5'-GTC TGT TGC TAT TAT G-3'.

RNA Oligonucleotide—The RNA oligonucleotide was 5'-GGG GGG AUU GGG GGG UAC AGU GCA GGG G-3'.

Preparation of the 5'-Radiolabeled RNA Template and the DNA Primer—DNA primers were labeled at the 5' end using T4 polynucleotide kinase (New England Biolabs) and [γ - 32 P]ATP (6000 Ci/mmol). Preparation of 5'-radiolabeled RNA template was performed as follows. The gel-cleaned RNA template was treated with shrimp alkaline phosphatase (Fermentas) at 37 °C for 60 min and then incubated at 65 °C for 25 min to inactivate the enzyme. After cooling on ice, the reaction mixture was treated with [γ - 32 P]ATP (6000 Ci/mmol), 10 \times PNK T4 Polynucleotide Kinase Reaction buffer and T4 polynucleotide kinase (New England Biolabs). After incubation for 1 h at 37 °C, the radiolabeled RNA and DNA primers were separated from unincorporated radionucleotides using a Micro Bio-Spin column (Bio-Rad).

Reverse Transcription Assay—RNA (200 fmol) and the 5' end-labeled DNA primer 19 (300 fmol) were heated to 95 °C for 2 min and annealed by slow cooling to 37 °C in 9 μ l. HIV-1 RT was added to the substrate and incubated for 4 min before the reaction was initiated with MgCl₂ and dNTPs. Reactions were carried out in 12.5 μ l at a final concentration of 50 mM Tris-HCl, pH 8.0, 50 mM KCl or LiCl, 1 mM DTT, 1 mM EDTA, 32 nM HIV-1 RT, 6 mM MgCl₂, and 50 μ M dNTPs. After 30 min of incubation at 37 °C, reverse transcriptions were stopped with 1 volume of termination buffer (10 mM EDTA, pH 8.0, 90% formamide (v/v), and 0.1% each xylene cyanol and bromophenol blue). Extension products were resolved on a 6% polyacrylamide, 8 M urea gel and analyzed using phosphorimaging (GE Healthcare). Sizes of DNA products were estimated by using a 5'-radiolabeled 10-bp DNA ladder (Invitrogen).

Circular Dichroism—CD spectra were obtained at 25 °C over a wavelength range of 220–340 nm using an AVIV circular dichroism spectrometer, model 202. The RNA was at a concentration of 4 μ M in 10 mM Tris HCl, pH 7.5, 0.3 mM EDTA and in the presence of KCl, NaCl, or LiCl at 10 and 50 mM concentrations. Before analysis, the samples were heated to 90 °C for 10 min, gently cooled at a rate of 1 °C/5 min, and incubated at 4 °C overnight. Spectra were recorded using a quartz cell of 1-mm optical path length, with data collected every nanometer at a bandwidth of 1 nm. Each spectrum was recorded three times and base-line-corrected for signal contributions from the buffer. The data were processed with AVIV Biomedical Inc. software and reported as ellipticity (millidegrees) versus wavelength (nm).

Thermal Denaturation of G-quadruplex Structure—Melting curves were obtained by monitoring the change at a wavelength of 265 nm, chosen for having the highest CD value. The temperature was raised from 20 to 90 °C; the heating rate was 2 °C/min. The melting temperature T_m was defined as the temperature of the mid-transition point.

Affinity Selection with Oligo-d(T)₂₅ Magnetic Beads—About 40 pmol of poly(A)-tagged RNA and non-tagged RNA tem-

plates were mixed (ratio 1:1) in the presence of 50 mM Tris-HCl, pH 8.0, 200 mM KCl, and 1 mM EDTA in a final volume of 20 μ l. The mixtures were heated to 95 °C for 3 min, then chilled on ice and incubated at room temperature for 2 h. Before using oligo-d(T)₂₅ magnetic beads (New England Biolabs), the suspension of 50 μ l was washed once with binding buffer (20 mM Tris-HCl, pH 7.5, 500 mM KCl, 1 mM EDTA), resuspended in 180 μ l of binding buffer, and added to the RNA. The mixture was agitated at room temperature for 10 min, and then placed in a magnetic rack to separate the magnetic beads from solution. The beads were washed once with binding buffer and three times with wash buffer (20 mM Tris-HCl, pH 7.5, 200 mM KCl, 1 mM EDTA), each time for 1 min with gentle agitation. To elute the RNA, the beads were resuspended in 15 μ l of elution buffer (20 mM Tris-HCl, pH 7.5, 1 mM EDTA), incubated for 3 min in 95 °C and placed in a magnetic rack to separate the magnetic beads from solution. 1 volume of loading buffer (10 mM EDTA, pH 8.0, 90% formamide (v/v), and 0.1% each xylene cyanol and bromophenol blue) was added to the eluted solution, and the products were resolved in 4 or 6% polyacrylamide, 8 M urea gels. The gel was stained with ethidium bromide.

Formation of Higher Order Products and Native Gel Analysis—The mixture of the 5' end 32 P-labeled (about 500,000 cpm) and unlabeled RNA at a concentration of 4 μ M and a final volume of 6 μ l was heated to 95 °C for 3 min, chilled, and incubated for 60 min at room temperature in a buffer containing 10 mM Tris-HCl, pH 7.5, 1 M KCl (or 1 M LiCl), 1 mM EDTA. Before electrophoresis, samples were mixed with 1 volume of loading dye (30% glycerol in 1 \times Tris-EDTA and loading pigments) and loaded onto a 6% non-denaturing polyacrylamide gel (0.5 \times Tris borate-EDTA + 10 mM KCl). Electrophoresis was performed in 4 °C at 7V/cm for 3–4 h, and the gel was dried onto Whatman No. 3MM chromatography paper and analyzed using phosphorimaging (GE Healthcare).

Cation-dependent Dimerization and Thermal Dissociation Analysis—Dimerization and melting experiments with RNA dimers of the cPPT region were conducted in parallel at the same molar concentration (4 μ M) for each reaction setting. To form a dimer, the RNA was heat to 95 °C for 3 min, chilled, and incubated for 16 h at room temperature in a buffer containing 10 mM Tris-HCl, pH 7.5, 1 mM EDTA, and one of three different salts (KCl, NaCl, or LiCl), each at 1 M. After incubation, the mixtures were placed on ice, and 1 volume of 1 \times Tris-EDTA buffer was added. Aliquots of 15 μ l were transferred to new tubes and incubated for 8 min at a specific temperature between 40 and 90 °C, then returned to the ice. Samples were mixed with 1 volume of loading dye (30% glycerol in 1 \times Tris-EDTA and loading dyes) and loaded onto 6% non-denaturing gels run at 4 °C. All gels were dried onto Whatman No. 3MM chromatography paper and analyzed using a phosphorimaging (GE Healthcare).

Antisense DNA Oligonucleotide Binding Assay—A 5-fold molar excess of each oligomer (**a-g**) was added to separate dimerization incubations (4 μ M RNA, 10 mM Tris-HCl, pH 7.5, 1 M LiCl, 1 mM EDTA) before the denaturation step. The samples were chilled on the ice and incubated for 16 h at room temperature in a final volume of 7 μ l. Products were resolved

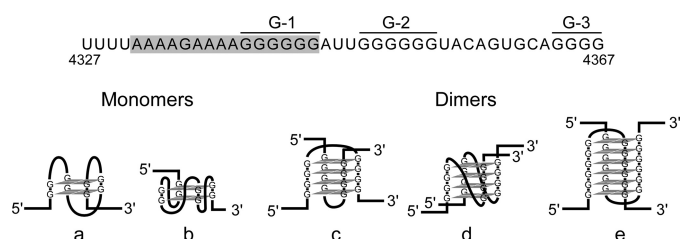


FIGURE 1. The guanine-rich sequence of the HIV-1 cPPT region may fold into a G-quartet. According to QGRS Mapper, a run of G residues in the cPPT (shaded nucleotides) and two other downstream G-rich elements are capable of forming a G-quartet. The structure can be polymorphic and can fold into several different conformations either as a single-molecule quadruplex (monomer) or intermolecular quadruplex (dimer) with parallel (b and d) and antiparallel configuration (a, c, and e). Examples of three different dimers are shown. Two-molecule G-quartets are composed of two sequences that interact in different orientations with only intermolecular hydrogen bonding (c) or with a combination of intra- and intermolecular hydrogen bonding (d and e) (46, 47).

on a 6% non-denaturing gel run at 4 °C in 0.5 × Tris borate-EDTA + 50 mM KCl.

Strand Transfer Assay—The DNA primer 20 was heat-annealed to donor RNA by incubation at 95 °C for 5 min and slow cooling to 37 °C. Acceptor template was also present in the mixture. NC at the 200% polymer substrate level (100% NC is 7 nt/NC polymer substrate molecule) was added and incubated for 3 min. Next, the RT was added to the mixture and incubated for another 4 min to pre-bind the RT with substrates before reactions were initiated with MgCl₂ and dNTPs. Primer, donor, and acceptor were mixed at a ratio of 2:1:1. The final reaction contained 50 mM Tris-HCl, pH 8.0, 50 mM KCl, 1 mM DTT, 1 mM EDTA, 32 nM HIV-1 RT, 6 mM MgCl₂, and 50 μM dNTPs. For reactions in the presence of lithium ions, KCl was replaced with 50 mM LiCl. Reactions were incubated at 37 °C and terminated after 1, 5, 15, and 30 min with 1 volume of termination dye (10 mM EDTA, pH 8.0, 90% formamide (v/v), and 0.1% each xylene cyanol and bromophenol blue). Products were then resolved by 6% polyacrylamide, 8 M urea gels and analyzed using phosphorimaging (GE Healthcare) and ImageQuant (Version 2.1). Sizes of DNA products were estimated by using a 5'-radio-labeled 10-bp DNA ladder (Invitrogen).

RESULTS

G-rich Sequences Capable of G-quartet Formation Are Present in the Middle of the Genome in Various HIV Species—Intensive research on the *gag* region near the dimerization initiation site of different retroviruses showed that formation of an RNA genomic dimer is also supported by interactions through G-quartet structure (9–13). To discover whether there are more distant G-rich sequences potentially capable of adhering the HIV-1 genomes, we used QGRS Mapper, a software program designed to predict formation of this structure (26). Our computational analyses revealed that the central region of the HIV-1 genome near the cPPT can form G-quartet structures with high probability. The sequence contains three G runs, in which the first G-rich element is a part of the cPPT sequence (Fig. 1). Putative G-quartets would also require downstream G-rich elements in HIV-1 that would be part of either an intramolecular structure containing two G-tetrads or a bimolecular structure with up to six G-tetrads. To form the RNA

dimer, only one or two G-rich elements on each RNA molecule would need to participate in the interaction.

Conservation of these elements would attest to their likely function in genome dimerization. The cPPT sequences are conserved among viral species, as expected from their multiple functions. To determine whether cPPT plus flanking sequences are also predicted to participate in G-quartet formation in different viruses, we analyzed genomes of various isolates of HIV-1 (A, B, C, D, F, G, H, U, O), HIV-2 (A, B, G, U), and several SIV species closely related to HIV-1. All analyzed sequences of HIV-1 and SIV have a similar distribution of cPPT-region G-rich elements and are also predicted to form G-quartet structures (Table 1). In the case of HIV-2, two G-rich elements are present and might be involved in formation of a dimer. In summary, our computational analyses of viral sequences suggest that G-rich sequences within and flanking the cPPT are involved in G-quartet formation in all analyzed species.

G-rich Regions Near the HIV-1 cPPT Induce Cation-dependent Pauses of RT—G-quartets are formed preferentially in the presence of specific monovalent cations and can be stabilized by some proteins and chemical agents. The order of cation preference is usually K⁺ > Na⁺ > Cs⁺ > Li⁺. The quadruplex formation rates increase with the salt concentration. Previous studies revealed cation-dependent pauses of RT progression immediately before and within a G-rich RNA template region involved in G-quartet formation (23, 27). In the presence of low concentrations of potassium ions, pause sites caused by hairpin structures and G-quartets were observed; however, in the presence of a low concentration of lithium ions the pauses caused by hairpin structures were not affected, whereas the pauses caused by G runs involved in G-quartet formation were greatly reduced. This cation-dependent pause profile behavior reliably indicates locations where an analyzed sequence can form a G-quartet structure.

The RNA sequence of the cPPT region (4309–4396 in the RNA genome) was synthesized by transcription *in vitro*. Then, reverse transcription by HIV-1 RT was performed using the transcript as a template in the presence of K⁺ or Li⁺ to analyze the cation-dependent pausing profile for this region. Results in Fig. 2A show that cation-dependent pauses of RT were observed in the presence of 50 mM concentrations of salt for the template folded at a concentration of 22 nM. Three RT pause sites in the presence of potassium ions correspond to the first, second, and third runs of G residues in the RNA template and presumably result from formation of different G-quartet structures. The pauses were not observed in the presence of lithium ions. We also performed the reactions in the presence of K⁺, Na⁺, and Li⁺, with folding at a higher salt concentration (200 mM). The mixtures were then diluted to 50 mM concentrations of salt before the addition of RT and initiation of the reaction. The pausing profile of RT showed that the pause at the third G-rich element (G-3) was stronger in the reaction performed in the presence of K⁺ than in Na⁺, which is consistent with known cation preferences for forming G-quartet structure (Fig. 2B). Weak RT pauses at G-rich elements in the presence of Li⁺ indicate that the structure can also be formed in a higher salt concentration, which is also consistent with previous reports showing that G-quartets might be formed efficiently in the

TABLE 1

The G-rich sequences in the cPPT region in different immunodeficiency viruses

Sequence of the cPPT region	Accession number of the virus [name] (subtype)
TTTAAAAAGAAAAGGGGGGATTGGGGGTACAGTGCAGGGG	HIV-1: AM000053 (A4); AB098332 (A1); X04415 [MAL] (A1DK); M19921 [NL4-3] (B); AB253432 (B); K03455 [HXB2] (B); AY228556 (C); AB023804 (C); EU293445 (C); AB097869 (BC); EU448296 (BCU); AY445524 (CD); DQ054367 (D); FJ771006 (F1); AB231893 (G); AF084936 (G); AF005496 (H); FJ711703 (H); AJ006022 (N); EF029067 (U); EF029069 (U)
TTTAAAAAGAAAAGGGGGGATTGGGGGTACAGTGCAGGGG	SIV: X52154 [SIV(cpz)]
TTTAAAAAGAAAAGGGGGGATTGGGGAGTACAGTGCAGGGG	HIV-1: AF290027 (C)
TTTAAAAAGAAAAGGGGGGATTGGGGGTACACTGCAGGAG	HIV-1: AJ320484 (D)
TTTAAAAAGAAAAGGGGGGATTGGGGGTACACTGCAGGAG	SIV: AF447763 [SIVcpzTAN1]
TTTAAAAAGAAAAGGGGGGATTGGGGGTACACTGCAGGAG	HIV-1: AJ302646 (O); AY618998 (O)
TTTAAAAAGAAAAGGGGGGATTGGGGGTACACTGCAGGAG	SIV: AF103818
TTTAAAAAGAAAAGGGGGGATTGGGGGTACACTGCAGGAG	HIV-2: NC_001722 [Ben] (A); U27200 [EHO] (B); AF208027 (G); AY530889 (U); AB499693 (AB)
TTTAAAAAGAAAAGGGGGGATTGGGGGTACACTGCAGGAG	HIV-2: M15390 [ROD] (A)

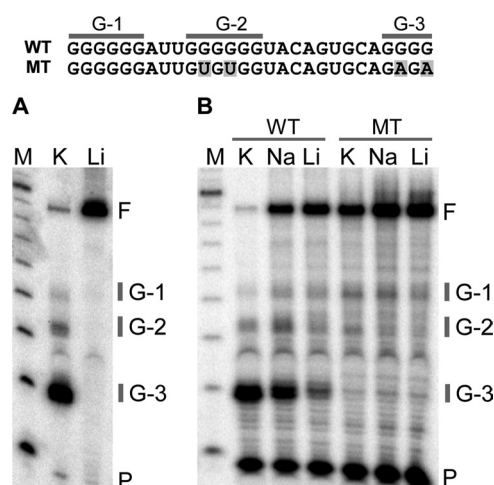


FIGURE 2. Reverse transcription assay determining G-quartet formation in the cPPT region. Cation-dependent pausing of reverse transcription at the guanine-rich elements in the cPPT region was analyzed with wild-type (WT) and mutant (MT) RNA templates. A fragment of the RNA template is shown (top) with G-rich elements (G-1, G-2, G-3) and mutations (nucleotides on gray background) introduced into the second and third G-rich elements. A, strong pauses of the RT at G-rich elements were observed in the presence of 50 mM KCl, but not LiCl, indicating that three runs of G residues are involved in the formation of different types of G-quartet structure. B, shown is a pausing profile of wild-type and mutant RNA templates folded in the presence of 200 mM KCl, NaCl, or LiCl and then diluted to 50 mM salt before initiation of reverse transcription. The profile reveals common G-quartet folding features, consistent with both the expected order of cation preference to form a stable structure and an influence of salt concentration on G-quartet formation. A cation-dependent pausing of RT at G-3 was not observed with mutant RNA templates. F, full-length extension products; P, DNA primer; M, DNA marker; K, KCl; Na, NaCl; Li, LiCl.

presence of lithium ions but are less stable (9, 10). Significantly, for an RNA template with base substitutions in the second and third G-rich elements (MT in Fig. 2), the RT pause corresponding to third G run was eliminated. This indicates that the third G-rich element contributes to formation of a G-quartet, which could be a monomer, dimers (between G-1 and G-3 or between G2 and G-3), or tetramer, all able to pause the RT. The pauses at first and second G-rich elements were clearly visible with the mutant template in the presence of potassium ions but less so in the presence of sodium and lithium, suggesting that both elements could still form G-quartets. However, these pause signals might only result from G-quartets formed between two and four RNA templates, as at least two more G residues are needed for this template to form a monomer G-quartet. In summary, these results demonstrate that G-rich elements in the HIV-1

cPPT region can fold into different forms of G-quartets, including monomers and intermolecular structures.

CD Spectra Analysis Confirms G-quartet Formation by RNA in the cPPT Region of HIV-1—G-quartets have a unique CD spectrum that depends on topology, which might be parallel or antiparallel. The parallel configuration refers to the structure in which the 5'-3' direction of all strands that form the G-tetrad is the same. If one or more strands have a 5'-3' direction opposite to the other strands, the G-quadruplex is said to adopt an antiparallel topology. The CD signature for a parallel configuration has a positive peak at around 265 nm and a negative peak at around 240 nm, whereas antiparallel G-quadruplexes have positive and negative peaks at 295 and 260 nm, respectively (28, 29). For RNA molecules, only a parallel configuration of G-quadruplex structure was observed.

The CD spectra were acquired for 4 μ M RNA in a buffer containing 50 or 10 mM KCl, NaCl, or LiCl. Before analysis, the samples were heated to 90 °C for 10 min, gently cooled at the rate of 1 °C/5 min, and incubated at 4 °C overnight. Results show that the 28-nt-long RNA of the cPPT region (4340–4367) in HIV-1 has a typical CD spectral profile for the parallel G-quadruplex configuration (Fig. 3A). The structure is formed in the presence of 50 mM K⁺, Na⁺, or Li⁺, although the intensity of the maximum peak at 265 nm is reduced for lithium ions. For RNA folded at 10 mM Li⁺, the CD profile almost overlaps with data recorded for RNA incubated in the absence of any salt. Significantly, the positive ellipticity at 265 nm is only slightly reduced for RNA folded in 10 mM KCl, indicating that the sequence easily adopts G-quadruplex configuration even at low salt concentration.

The results from the RT progression assays did not show G-quartet formation in 50 mM LiCl. However, in this method the RNA concentration used for folding was also much lower (22 nM) than for the CD spectra (4 μ M), and the ability of a sequence to form a G-quartet is known to increase with the salt and template concentration.

The thermal stability analysis also indicates that the G-quadruplex structure formed in the cPPT region is very stable in the presence of KCl. Melting curves were obtained by monitoring changes at a wavelength 265 nm at a heating rate of 2 °C/min. Fig. 3B shows normalized CD melting curves for G-quartets folded in the presence of K⁺, Na⁺, and Li⁺. The thermal denaturation for G-quadruplexes formed in 50 mM KCl showed that ~80% of them were still present at 90 °C. The unfolding of the

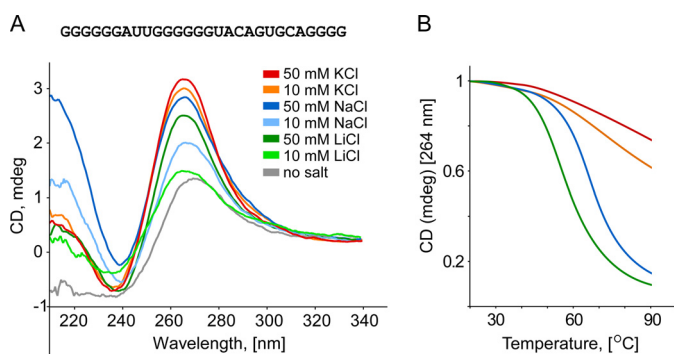


FIGURE 3. CD spectral analysis of the G-rich RNA sequence near the cPPT in HIV-1. A, shown are CD spectra of the G-quadruplex-forming sequence in different salts at concentrations of 10 and 50 mM. The effects of K^+ , Na^+ , and Li^+ on the ellipticity signal are compared with the signal for RNA incubated in the absence of salt. Salt concentrations are 50 mM KCl (red), 10 mM KCl (orange), 50 mM NaCl (dark blue), 10 mM NaCl (light blue), 50 mM LiCl (dark green), and 10 mM LiCl (light green) and no salt (gray). B, shown are normalized CD melting temperature curves for G-quadruplexes formed in 50 mM KCl (red), 50 mM NaCl (blue), 50 mM LiCl (green), and 10 mM KCl (orange).

structure is likely prevented by the presence of potassium ions. Surprisingly, the melting curve for the structures formed in 10 mM KCl showed that only 40% of them were unfolded at the highest temperature, indicating that the G-quadruplex is very stable. The G-quadruplexes formed in 50 mM NaCl or LiCl completely melted with T_m values of 68 and 59 °C, respectively. Because potassium is the dominant monovalent cation inside cells, these results suggest that G-quadruplex configuration could be easily adopted by viral RNA sequence of the cPPT region.

Interacting RNA Strands with the HIV-1 cPPT Region Sequence Are Selected by an Affinity Isolation Method—One, two, or four nucleic acid molecules can form a G-quadruplex. To test whether G-quartets that we detected are formed as intermolecular structures, we developed an affinity selection method, with which the interactions between nucleic acids can be tested by selection of interacting partners with one tagged by a poly(A) sequence. Magnetic beads conjugated with oligo(dT)₂₅ were used for affinity selection. The method has been used to select mRNAs and was modified here by using buffers suitable for G-quartet formation. The interacting partners are distinguished in a denaturing gel stained with ethidium bromide. We tested the specificity of this approach using RNA strands with the HIV-1 DIS synthesized by transcription *in vitro*. The poly(A)-tagged RNA with the sequence 183–520 of the HIV-1 RNA genome (with the DIS at position 257–262) could select another RNA having the DIS (1–520) but devoid of poly(A). Non-tagged DIS RNA could not be selected by magnetic beads in the absence of the poly(A)-tagged partner, demonstrating that the observed interactions are not the result of nonspecific binding to the magnetic beads (Fig. 4A).

To determine whether RNA molecules with the sequence of the HIV-1 cPPT region interact, the RNA strands corresponding to positions 4309–4396 of the HIV-1 NL4-3 RNA genome were synthesized with poly(A) tails and were co-incubated with equivalent RNA strands but devoid of poly(A) sequence. Because the *gag* regions near DIS from HIV-1 MAL and NL4-3 were already shown to form dimers through G-quartet structure, we used RNA strands of these regions as

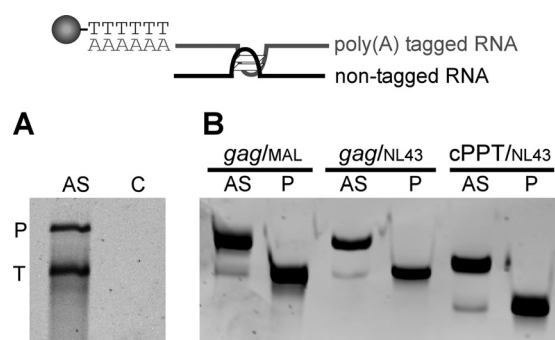


FIGURE 4. Affinity selection of RNAs interacting through intermolecular G-quartet structure. Poly(A)-tagged RNA templates were elongated at the 3' end with an A₂₅ polymer to select them with oligo(dT)₂₅ magnetic beads. To determine whether RNA molecules can interact, the poly(A)-tagged RNA was mixed with corresponding non-tagged RNA and incubated under dimerization conditions for 2 h. The oligo(dT)₂₅ magnetic beads were added to select the poly(A)-tagged RNAs, and after three rounds of washes, the samples were eluted. The selected poly(A) RNAs and interacting partners were analyzed in a denaturing gel stained with ethidium bromide. A, a control experiment shows that the non-tagged RNA with DIS (RNA genomic sequence 1–520) can be selected with oligo(dT)₂₅ magnetic beads only in the presence of poly(A)-tagged RNA with DIS (183–520). No RNA was selected in the absence of a poly(A)-tagged partner (line C). B, the RNAs of the *gag* region (RNA genomic sequence 303–415 of the MAL isolate and 290–403 of the NL4-3) and cPPT region (4309–4396 of the NL4-3) could be isolated with the corresponding poly(A)-tagged RNA partners after affinity selection with oligo(dT)₂₅ magnetic beads. AS, affinity selection; T, poly(A)-tagged RNA; P, non-tagged RNA partner; C, control.

positive controls for dimerization (9, 10, 12). Our RNA strands having the relevant *gag* sequences (303–415 in HIV-1 MAL, 290–403 in HIV-1 NL4-3) did not include the DIS. The RNA partners were combined and incubated to allow possible dimerization and subsequently used for affinity selection with magnetic beads. As shown in Fig. 4B, the non-tagged RNA strands of the *gag* region of the two different HIV-1 species (MAL and NL4-3) and non-tagged RNA strands of the cPPT region were co-selected with corresponding poly(A)-tagged RNAs. This indicates that the G-rich cPPT region of the HIV-1 RNA genome is a likely additional point of contact between the two RNA genomes that augments dimerization initiated through DIS.

Dimerization of RNA Strands from the HIV-1 cPPT Region—The complexes of interacting RNA partners from the cPPT region selected by magnetic beads might have represented a mixture of dimers and tetramers in which G-quartets are formed between two and four nucleic acid molecules, respectively. To determine what types of complexes are formed between these RNA molecules, we used a native gel analysis assay in which RNA templates of the cPPT region were radio-labeled, incubated in the presence of 1 M KCl or LiCl, and then resolved in a native polyacrylamide gel. As a control, we used RNA strands with the antisense sequence.

An 88-nt HIV-1 fragment of the cPPT region (4309–4396) including the G-rich segments self-associated to form mostly a dimeric complex and a lesser amount of tetrameric complex, both with reduced electrophoretic mobility. In contrast, the antisense RNA template remained unfolded, which strongly suggests that strand interaction resulted from structures that cannot exclusively depend on A:U and G:C base pairs (Fig. 5). Surprisingly, the higher order structures were formed more efficiently in the presence of LiCl than in the presence of KCl, a

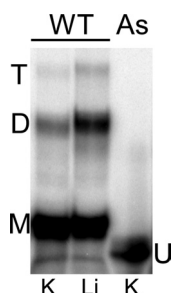


FIGURE 5. Dimerization of the HIV-1 cPPT region. The RNA templates of sense (WT) and antisense (As) strands were incubated under dimerization conditions for 2 h in the presence of 1 M KCl or LiCl. Formation of monomers (M), dimers (D), and tetramers (T) was analyzed on a non-denaturing gel run at 4 °C in the presence of 10 mM KCl and 0.5 × Tris borate-EDTA. Li, lithium; K, potassium; U, unfolded RNA.

result not observed at the same concentration of RNA in the CD spectral analysis. However, the concentration of salt in both experiments was also significantly different (1 M for RNA dimerization and 50 mM for CD spectra).

The higher rate of G-quartet dimer formation in the presence of LiCl is consistent with results obtained for G-quartet dimers formed from *gag* region sequences in which the yield of dimers correlated inversely with the size of monovalent cation (*i.e.* $\text{Li}^+ > \text{Na}^+ > \text{K}^+$) (9, 10). However, although the RNA G-quartet dimers of the *gag* region fold at a slower rate in the presence of potassium ions, the complexes are much more stable than those folded in the presence of lithium ions. Our results indicate that this is also the case with the G-rich sequences of the cPPT region. A thermal dissociation experiment (Fig. 6) revealed that although only 12% dimerization occurred in the presence of potassium ions, the dimers remained stable when incubated for 8 min at temperatures between 40 and 90 °C. Formation of dimers was more efficient in the presence of sodium ions, with ~34% of the strands forming complexes, and they also remained stable during the incubation at higher temperatures. However, in the presence of lithium ions, although up to 49% of the strands formed dimers, they exhibited substantial breakdown after incubation at 50 °C and were completely disrupted after incubation at 90 °C. These results demonstrate that complexes formed by the RNA template of the cPPT region display characteristics known for intermolecular RNA G-quartet structures.

To confirm that the G-rich sequences contributed to dimerization of the templates, we attempted to form dimers with the G-rich sequences in the additional presence of different antisense 16-nt-long DNA oligonucleotides that were expected to prevent G-quartet formation by interacting with complementary regions of sense RNA template (Fig. 7). Because dimers were more efficiently formed in the presence of lithium ions, we used this salt for this experiment. Inclusion of a 5-fold molar excess of each DNA oligomer confirmed that dimer formation is affected by DNA oligomers that bind the G-rich sequences. In particular, G-rich element G-1 within the cPPT and G-2 were indicated to be critical for RNA templates dimerization. These results establish that dimer formation within the cPPT region depends on G residues and involves two of the most conserved G-rich elements.

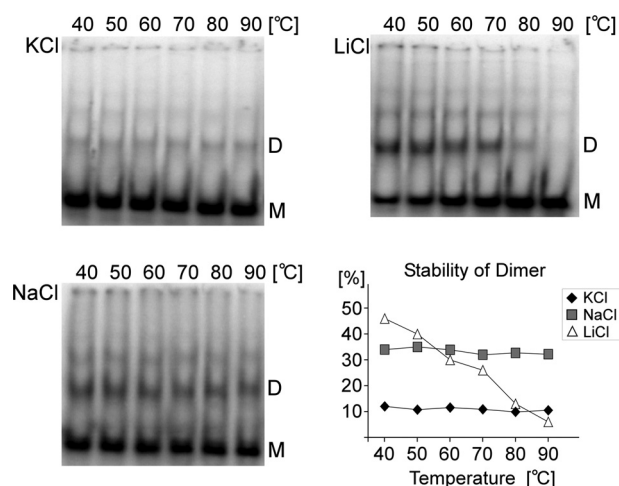


FIGURE 6. Cation-dependent association and thermal stability of the RNA dimer formed in the cPPT region. The RNA dimers were formed in parallel at a concentration of 4 μM in buffers containing 1 M KCl, NaCl, or LiCl. 1 volume of 1 × Tris-EDTA buffer was added, and 15-μl aliquots were incubated at the indicated temperatures for 8 min. Thermal stabilities were measured by analyzing samples in non-denaturing gels run at 4 °C. M, monomer; D, dimers.

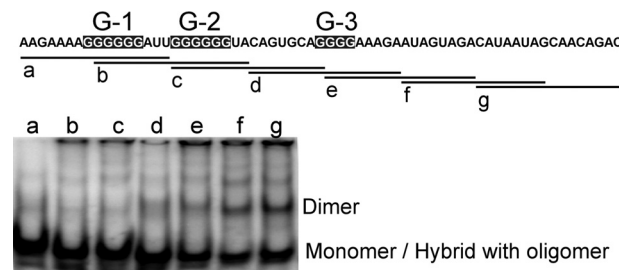


FIGURE 7. Antisense oligonucleotide binding assay. Dimerization of the cPPT region was performed in the presence of a 5 M excess of different antisense DNA oligomers (a–g) binding to specific cPPT region sequences.

G-quartet Structure Near the cPPT Promotes Template Switching by the RT—We previously showed that a major recombination hot spot in the HIV-1 *gag* region near DIS correlates with sequences rich in G residues that can form G-quartet structures (23, 24, 30). We showed enhanced strand transfer resulting from G-quartet dimer formation that holds the templates in close proximity and/or G-quartet monomer or dimer formation that increases frequency of RT RNase H by pausing the RT (Fig. 8A).

We have now acquired evidence that the G-quartet structure formed in the cPPT region also promotes template switching by the RT. We were initially encouraged in this expectation because the distribution of recombination breakpoints across multiple HIV-1 genomes revealed several concentrations, one of which was within a 200-nt-long sequence spanning from near the cPPT to the 3' end of the *pol* gene that contains runs of G residues (31). To test the basis of the observed template switching, we constructed a reconstituted system to determine whether G-quartet structure induces the RT to transfer the synthesis of minus-strand DNA from one RNA template to another. This system consisted of HIV-1 RT, HIV-1 NC, a primer (DNA oligonucleotide), and two RNA templates representing two copies of the HIV-1 RNA genome denoted as donor and acceptor for primer strand transfer (Fig. 8B). The reaction was initiated from a ^{32}P -labeled oligo(dT) DNA primer

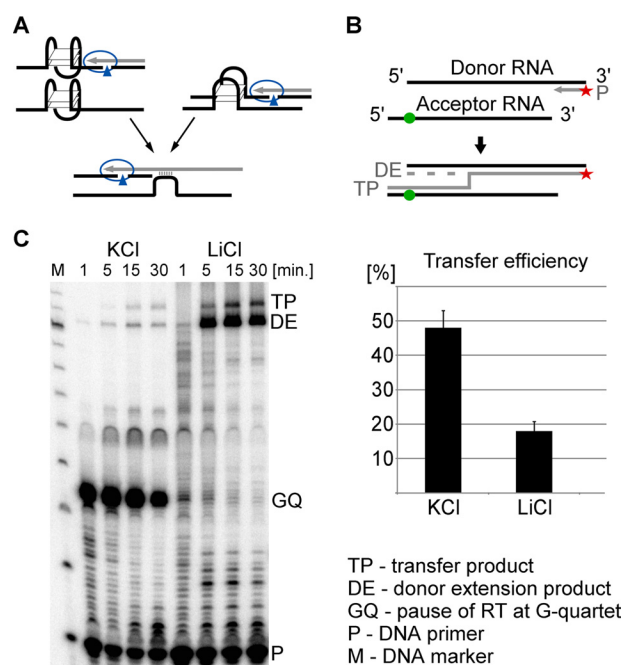


FIGURE 8. G-quartet formation in the cPPT region facilitates RT template switching during reverse transcription. A, G-quartet formation promoted minus-strand DNA transfer (gray line) through pausing of the RT (blue oval), which increased RNA cleavages (blue triangle) and allowed interaction with the second RNA template. The transfers were also promoted through an RNA template proximity effect resulting from template dimerization through a G-quartet dimer. B, shown is a reconstituted system to analyze the influence of G-rich elements on strand transfer during HIV-1 minus-strand DNA synthesis *in vitro*. Donor and acceptor RNA templates represent two copies of the viral RNA genome in which reverse transcription is initiated from a ^{32}P -labeled DNA primer (P) annealed to donor RNA. The transfer reactions onto acceptor RNA will result in synthesis of TP distinguished in a denaturing gel from the DE by its greater length. The acceptor RNA does not share a homology (green dot) with two nucleotides at the 5' end of the donor RNA. These alterations prevent end transfers of donor extension products; thus, transfer products should only originate internally. C, shown is a time course of strand transfer reactions performed in the presence of potassium and lithium ions. Samples were collected at 1, 5, 15, and 30 min after the reaction was initiated. Formation of a G-quartet in the RNA templates paused the RT during minus-strand DNA synthesis and influenced the yield of the final products. The transfer efficiency was 2.6-fold higher in reactions in which the templates could form a G-quartet than for reactions with substrates that could not form the structure.

annealed to the donor RNA. The donor and acceptor templates shared a homology spanning an 86-nt sequence of cPPT region (4311–4396). The donor RNA template (4309–4396) was elongated with a poly(A) sequence at its 3' end to serve as a primer binding site. The acceptor RNA (4311–4396) was elongated at the 5' end with GGAAAAAAAAA, so that transfer products could be separated and distinguished on a denaturing gel from DNA synthesized on the donor RNA. Moreover, the acceptor template does not share homology with two nucleotides at the 5' end of the donor RNA (4309–4310, green dot in Fig. 8B) to prevent transfers from the end of the donor template. Thus, all transfers to the acceptor would only originate from internal regions of the donor template, as they do *in vivo* during synthesis over the cPPT region. When the primer switched from the donor onto the acceptor, the completion of the synthesis on the acceptor yielded a 115-nt transfer product (TP). However, if the primer completed its synthesis on the donor without transfer, it yielded a 104-nt donor extension product

(DE). To determine whether formation of G-quartets influence transfer efficiency, we performed strand transfer assays in the presence of a low concentration of either K^+ or Li^+ (Fig. 8C). The transfer efficiency of reactions was calculated by comparison of values of donor extension products and transfer products using the formula: transfer efficiency = $100 \times \text{TP}/(\text{TP} + \text{DE})$. We also note that monovalent cations added to the reaction do not significantly affect the enzymatic activity of RT (32, 33).

In these strand transfer reactions, the major cation-dependent pause site of RT synthesis was clearly visible at the first G-rich element (GQ) in the presence of K^+ but not Li^+ . The strong pause evidently caused dissociation of the RT, as fewer final products were made when the pause was prominent. Significantly, the strong pause induced by potassium, presumably resulting from G-quartet formation, resulted in a 2.6-fold increase in transfer efficiency when compared with the reaction in the presence of lithium ions, wherein the structure was not formed. The strong pausing of the RT most likely promoted cleavages in the RNA donor template catalyzed by the RT RNase H, as previously demonstrated (23, 34). It is also probable that G-quartet induced template dimerization, promoted transfer by template proximity. Thus, this result strongly indicates that the high recombination rate in the cPPT region observed *in vivo* is associated with formation of G-quartet structure.

DISCUSSION

Electron microscopic evidence suggests that the two genomic RNAs in HIV-1, murine leukemia virus, Rous sarcoma virus, avian reticuloendotheliosis virus, endogenous feline type C virus RD-114, and woolly monkey sarcoma virus are linked but only within the region close to their 5' ends (35–39). However, the denaturing conditions used in electron microscopy likely disrupt most of the RNA secondary and tertiary structures. Under milder conditions additional points of contact were observed between the two RNA genomes of Rous sarcoma virus, suggesting that interactions at the 5' end could be the most stable point of a multiple contact association (40). It was also shown that nicked viral RNA of avian leukosis virus migrates as a dimer, suggesting that additional interactions in other parts of the genome maintain integrity of damaged RNA (41). Moreover, a significant amount of HIV-1 viruses with dimerized genomes were formed even when the DIS was made non-functional with dimerization-disruptive mutations, although the viruses were less infectious (8, 42). Intensive analysis of the HIV-1 *gag* region RNA-RNA interactions revealed that in addition to dimerization at DIS, RNA templates with the *gag* sequence adhere by formation of intermolecular G-quartets folded from G-rich sequences located ~100 nt 3' of DIS (9–13). The initial interactions through DIS and parallel alignment of the RNA strands raises the local RNA concentration to a point at which additional interactions are favored at other locations of the genome through intermolecular G-quartets. One of those places appears to be the central region of the HIV-1 genome near the cPPT, where three runs of G residues are present. The G-rich elements in this region are conserved among different HIV-1 isolates and can also be found in closely related SIV strains. In the case of HIV-2, there

are two G-rich elements, G-1 and G-2, in this region, sufficient to form a dimer, and our results indicate that they are critical for forming an intermolecular G-quartet. With an affinity selection approach and native gel analysis we show that two RNA molecules with the cPPT region sequence can self-associate, and the dimeric complex has similar features and properties resembling those described for RNA dimers formed in the *gag* region through an intermolecular G-quartet. The G-2 and G-3 elements in the HIV-1 cPPT region were previously shown to form G-quartet structure in DNA, and it was suggested that the elements are involved in formation of the cPPT DNA flap (43).

Structurally, G-quartets can be polymorphic and adopt several different forms depending on their sequence and concentration and the milieu of monovalent cations. Our results also show this diversity, and whereas G-quartet dimers are evident at a high concentration of RNA and salt, monomers are almost solely formed in the presence of potassium ions at a low salt and RNA template concentration. Of course, our results cannot fully simulate the structural arrangements of the cPPT region in a virion. However, they reliably demonstrate that the sequences having G-rich elements in the middle portion of the RNA genome are capable of forming a G-quartet dimer and are likely a recombination-relevant position of interaction between the two viral genomes.

Our results also confirm that G-quartets formed near the HIV-1 cPPT can effectively pause synthesis by the RT and induce DNA primer strand transfer from a donor to an acceptor RNA template, suggesting that a G-quartet formed near the cPPT facilitates RT template switching. In fact, an analysis of 271 sequences of the HIV-1 group M subtype indicates that the sequence near the 3' end of the *pol* gene, containing the cPPT, is a common site of recombinant breakpoints (31). Very likely, the increased rate of recombination in this region derives from the presence of G-quartets.

Because, in our experiments monomers were formed more efficiently than dimers, the stimulation of transfers *in vitro* probably did not primarily result from a proximity effect of dimerization. However, dimerization of templates through DIS and G-quartets in *gag* was previously shown to stimulate strand transfer (24, 44). Moreover, in the context of whole genomes, the opportunity for multiple contact points should favor dimeric over monomeric G-quartet formation. Consequently, we expect that dimerization is also a substantial contributor to recombination in the cPPT region. Significantly, the two genomic RNAs adhering by formation of an intermolecular G-quartet near the cPPT would have a global effect on recombination as, together with the *gag* adherence point and possibly others, the adhering would maintain the proximity of the viral genomes over a long distance. This would contribute to increased recombination rates in many other locations.

Previous studies showed that *pol* gene sequences and others located downstream display recombination hot spots (30, 31, 45). Because reverse transcription proceeds from the 3' to the 5' end of the genomic RNA, the dimerization near cPPT would facilitate recombination in downstream (3' on the template) regions only, as all interactions between genomes would be disrupted by synthesis of minus-strand DNA. Investigation of the distribution of recombination breakpoints across multiple

HIV-1 genomes revealed the presence of a recombination hot spot lying ~1 kb downstream of the cPPT, within the 600-nt-long sequence window of the first exon of *tat*, *vpu*, *rev* and the beginning of *env* (31). In fact, this region together with the second exon of *tat* and *rev* exhibits the highest rate of recombination throughout the whole genome. We propose that global genome spatial alignment, mediated by multiple sites of dimerization, augments the recombination rate at this and other major hot spots. We also suspect that local dimerization of genomic RNA in different regions plays a role in distinguishing RNA genomes from messenger RNA molecules.

The ability of viral RNA genomes with binding-disruptive mutations in DIS to form some dimers indicates that the mechanism of genome dimerization also involves other regions (8, 42). Our results show that the sense RNA template of the cPPT region has an ability to spontaneously dimerize through G-quartet formation. Thus, we suggest that alignment of HIV-1 RNA *in vivo* is also mediated by formation of G-tetrads between viral genomes. Because self-association of RNA templates through the G-quartet structure occurs at a slower rate than through DIS, G-dimers would appear after interaction at DIS and parallel alignment of two viral genomes, which would bring homologous regions into proximity and initiate formation of intermolecular G-quartets in *gag*, near cPPT, and likely in other locations. This is in agreement with an observation that genomic RNA dimerization undergoes maturation process, as gRNA dimers of newly released viral particles are less stable than genomic RNA dimers of 48-h-old virions (8).

Formation of G-quartet dimers between homologous sequences throughout retroviral genomes could be a spontaneous reaction, similar to formation of double-stranded DNA or formation of helices involving hybridization of complementary sequences. Any sequence with two runs of two or more G residues could potentially be a region where the RNA genomes associate, although the distance between two G runs and the sequence context will be significant factors determining the probability of the formation of G-quartet structure. For example, the G residues might be occluded in stable hairpins so that formation of G tetrads cannot occur. Thus, more advanced analyzes with application of computational approaches are needed to predict genomic regions more prone to form G-quartet dimers between associated retroviral genomes. Moreover, the sequence requirements for formation of G-quartets are yet not fully understood, as the structure might also involve some adenines, as was proposed for the *gag* region (9, 10). The cPPT sequence also has a long track of A residues that might participate in formation of a quadruplex. The pausing of RT during reverse transcription occurs primarily at G residues, but that does not rule out involvement of A residues, particularly those after G residues, forming a purine run.

In summary, there is still little known about the mechanisms and factors involved in dimerization of viral genomes despite the fact that retroviral RNA dimerization is a key element in virus propagation and survival. Thus, viral genome dimerization might be a good target for inhibitors of replication of HIV-1 and other retroviruses. However, understanding of this process is prerequisite for development of suitable therapeutic approaches.

Acknowledgment—We are grateful to Dr. Robert J. Gorelick of the National Cancer Institute for NC used in these studies.

REFERENCES

- Moore, M. D., Fu, W., Nikolaitchik, O., Chen, J., Ptak, R. G., and Hu, W. S. (2007) Dimer initiation signal of human immunodeficiency virus type 1. Its role in partner selection during RNA copackaging and its effects on recombination. *J. Virol.* **81**, 4002–4011
- Moore, M. D., Nikolaitchik, O. A., Chen, J., Hammarskjöld, M. L., Rekosh, D., and Hu, W. S. (2009) Probing the HIV-1 genomic RNA trafficking pathway and dimerization by genetic recombination and single virion analyses. *PLoS Pathog.* **5**, e1000627
- Clever, J. L., and Parslow, T. G. (1997) Mutant HIV-1 genomes with defects in RNA dimerization or encapsidation. *J. Virol.* **71**, 3407–3414
- Laughrea, M., Jetté, L., Mak, J., Kleiman, L., Liang, C., and Wainberg, M. A. (1997) Mutations in the kissing-loop hairpin of HIV-1 reduce (...) genomic RNA packaging and dimerization. *J. Virol.* **71**, 3397–3406
- McBride, M. S., and Panganiban, A. T. (1996) The human immunodeficiency virus type 1 encapsidation site is a multipartite RNA element composed of functional hairpin structures. *J. Virol.* **70**, 2963–2973
- Skripkin, E., Paillart, J. C., Marquet, R., Ehresmann, B., and Ehresmann, C. (1994) Identification of the primary site of the human immunodeficiency virus type 1 RNA dimerization *in vitro*. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 4945–4949
- Paillart, J. C., Skripkin, E., Ehresmann, B., Ehresmann, C., and Marquet, R. (1996) A loop-loop “kissing” complex is the essential part of the dimer linkage of genomic HIV-1 RNA. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 5572–5577
- Song, R., Kafaie, J., Yang, L., and Laughrea, M. (2007) HIV-1 viral RNA is selected in the form of monomers that dimerize in a three-step protease-dependent process. The DIS of stem-loop 1 initiates viral RNA dimerization. *J. Mol. Biol.* **371**, 1084–1098
- Marquet, R., Baudin, F., Gabus, C., Darlix, J. L., Mougél, M., Ehresmann, C., and Ehresmann, B. (1991) Dimerization of human immunodeficiency virus (type 1) RNA. Stimulation by cations and possible mechanism. *Nucleic Acids Res.* **19**, 2349–2357
- Sundquist, W. I., and Heaphy, S. (1993) Evidence for interstrand quadruplex formation in the dimerization of human immunodeficiency virus 1 genomic RNA. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 3393–3397
- Awang, B. G., and Sen, D. (1993) Mode of dimerization of HIV-1 genomic RNA. *Biochemistry* **32**, 11453–11457
- Marquet, R., Paillart, J. C., Skripkin, E., Ehresmann, C., and Ehresmann, B. (1994) Dimerization of human immunodeficiency virus type 1 RNA involves sequences located upstream of the splice donor site. *Nucleic Acids Res.* **22**, 145–151
- Ly, H., Nierlich, D. P., Olsen, J. C., and Kaplan, A. H. (1999) Moloney murine sarcoma virus genomic RNAs dimerize via a two-step process. A concentration-dependent kissing-loop interaction is driven by initial contact between consecutive guanines. *J. Virol.* **73**, 7255–7261
- Siddiqui-Jain, A., Grand, C. L., Bearss, D. J., and Hurley, L. H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 11593–11598
- Sun, D., Guo, K., and Shin, Y. J. (2011) Evidence of the formation of G-quadruplex structures in the promoter region of the human vascular endothelial growth factor gene. *Nucleic Acids Res.* **39**, 1256–1265
- Kumari, S., Bugaut, A., Huppert, J. L., and Balasubramanian, S. (2007) An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. *Nat. Chem. Biol.* **3**, 218–221
- Marcel, V., Tran, P. L., Sagne, C., Martel-Planche, G., Vaslin, L., Teulade-Fichou, M. P., Hall, J., Mergny, J. L., Hainaut, P., and Van Dyck, E. (2011) G-quadruplex structures in TP53 intron 3. Role in alternative splicing and in production of p53 mRNA isoforms. *Carcinogenesis* **32**, 271–278
- Gomez, D., Lemarteleur, T., Lacroix, L., Mailliet, P., Mergny, J. L., and Riou, J. F. (2004) Telomerase down-regulation induced by the G-quadruplex ligand 12459 in A549 cells is mediated by hTERT RNA alternative splicing. *Nucleic Acids Res.* **32**, 371–379
- Didiot, M. C., Tian, Z., Schaeffer, C., Subramanian, M., Mandel, J. L., and Moine, H. (2008) The G-quartet containing FMRP binding site in FMR1 mRNA is a potent exonic splicing enhancer. *Nucleic Acids Res.* **36**, 4902–4912
- Fisette, J. F., Montagna, D. R., Mihailescu, M. R., and Wolfe, M. S. (2012) A G-rich element forms a G-quadruplex and regulates BACE1 mRNA alternative splicing. *J. Neurochem.* **121**, 763–773
- Brown, V., Jin, P., Ceman, S., Darnell, J. C., O'Donnell, W. T., Tenenbaum, S. A., Jin, X., Feng, Y., Wilkinson, K. D., Keene, J. D., Darnell, R. B., and Warren, S. T. (2001) Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome. *Cell* **107**, 477–487
- Norseen, J., Johnson, F. B., and Lieberman, P. M. (2009) Role for G-quadruplex RNA binding by Epstein-Barr virus nuclear antigen 1 in DNA replication and metaphase chromosome attachment. *J. Virol.* **83**, 10336–10346
- Shen, W., Gao, L., Balakrishnan, M., and Bambara, R. A. (2009) A recombination hot spot in HIV-1 contains guanosine runs that can form a G-quartet structure and promote strand transfer *in vitro*. *J. Biol. Chem.* **284**, 33883–33893
- Shen, W., Gorelick, R. J., and Bambara, R. A. (2011) HIV-1 nucleocapsid protein increases strand transfer recombination by promoting dimeric G-quartet formation. *J. Biol. Chem.* **286**, 29838–29847
- Pandey, V. N., Kaushik, N., Rege, N., Sarafian, S. G., Yadav, P. N., and Modak, M. J. (1996) Role of methionine 184 of human immunodeficiency virus type-1 reverse transcriptase in the polymerase function and fidelity of DNA synthesis. *Biochemistry* **35**, 2168–2179
- Kikin, O., D'Antonio, L., and Bagga, P. S. (2006) QGRS Mapper. A web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res.* **34**, W676–W682
- Schaeffer, C., Bardoni, B., Mandel, J. L., Ehresmann, B., Ehresmann, C., and Moine, H. (2001) The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif. *EMBO J.* **20**, 4803–4813
- Smargiasso, N., Rosu, F., Hsia, W., Colson, P., Baker, E. S., Bowers, M. T., De Pauw, E., and Gabelica, V. (2008) G-quadruplex DNA assemblies. Loop length, cation identity, and multimer formation. *J. Am. Chem. Soc.* **130**, 10208–10216
- Paramasivan, S., Rujan, I., and Bolton, P. H. (2007) Circular dichroism of quadruplex DNAs. Applications to structure, cation effects, and ligand binding. *Methods* **43**, 324–331
- Dykes, C., Balakrishnan, M., Planelles, V., Zhu, Y., Bambara, R. A., and Demeter, L. M. (2004) Identification of a preferred region for recombination and mutation in HIV-1 gag. *Virology* **326**, 262–279
- Fan, J., Negroni, M., and Robertson, D. L. (2007) The distribution of HIV-1 recombination breakpoints. *Infect. Genet. Evol.* **7**, 717–723
- Filler, A. G., and Lever, A. M. (1997) Effects of cation substitutions on reverse transcriptase and on human immunodeficiency virus production. *AIDS Res. Hum. Retroviruses* **13**, 291–299
- Yong, W. H., Wyman, S., and Levy, J. A. (1990) Optimal conditions for synthesizing complementary DNA in the HIV-1 endogenous reverse transcriptase reaction. *AIDS* **4**, 199–206
- Chen, Y., Balakrishnan, M., Roques, B. P., and Bambara, R. A. (2003) Steps of the acceptor invasion mechanism for HIV-1 minus strand strong stop transfer. *J. Biol. Chem.* **278**, 38368–38375
- Höglund, S., Ohagen, A., Goncalves, J., Panganiban, A. T., and Gabuzda, D. (1997) Ultrastructure of HIV-1 genomic RNA. *Virology* **233**, 271–279
- Bender, W., Chien, Y. H., Chattopadhyay, S., Vogt, P. K., Gardner, M. B., and Davidson, N. (1978) High molecular weight RNAs of AKR, NZB, and wild mouse viruses and avian reticuloendotheliosis virus all have similar dimer structures. *J. Virol.* **25**, 888–896
- Bender, W., and Davidson, N. (1976) Mapping of poly(A) sequences in the electron microscope reveals unusual structure of type C oncornavirus RNA molecules. *Cell* **7**, 595–607
- Kung, H. J., Hu, S., Bender, W., Bailey, J. M., Davidson, N., Nicolson, M. O., and McAllister, R. M. (1976) RD-114, baboon, and woolly monkey viral RNA's compared in size and structure. *Cell* **7**, 609–620
- Murti, K. G., Bondurant, M., and Tereba, A. (1981) Secondary structural

RNA Dimerization in the Central Region of the HIV-1 Genome

- features in the 70S RNAs of Moloney murine leukemia and Rous sarcoma viruses as observed by electron microscopy. *J. Virol.* **37**, 411–419
40. Mangel, W. F., Delius, H., and Duesberg, P. H. (1974) Structure and molecular weight of the 60–70 S RNA and the 30–40 S RNA of the Rous sarcoma virus. *Proc. Natl. Acad. Sci. U.S.A.* **71**, 4541–4545
41. Ortiz-Conde, B. A., and Hughes, S. H. (1999) Studies of the genomic RNA of leukosis viruses. Implications for RNA dimerization. *J. Virol.* **73**, 7165–7174
42. Shen, N., Jetté, L., Liang, C., Wainberg, M. A., and Laughrea, M. (2000) Impact of human immunodeficiency virus type 1 RNA dimerization on viral infectivity and of stem-loop B on RNA dimerization and reverse transcription and dissociation of dimerization from packaging. *J. Virol.* **74**, 5729–5735
43. Lyonais, S., Hounsou, C., Teulade-Fichou, M. P., Jeusset, J., Le Cam, E., and Mirambeau, G. (2002) G-quartets assembly within a G-rich DNA flap. A possible event at the center of the HIV-1 genome. *Nucleic Acids Res.* **30**, 5276–5283
44. Balakrishnan, M., Roques, B. P., Fay, P. J., and Bambara, R. A. (2003) Template dimerization promotes an acceptor invasion-induced transfer mechanism during human immunodeficiency virus type 1 minus-strand synthesis. *J. Virol.* **77**, 4710–4721
45. Zhuang, J., Jetzt, A. E., Sun, G., Yu, H., Klarmann, G., Ron, Y., Preston, B. D., and Dougherty, J. P. (2002) Human immunodeficiency virus type 1 recombination. Rate, fidelity, and putative hot spots. *J. Virol.* **76**, 11273–11282
46. Williamson, J. R. (1994) G-quartet structures in telomeric DNA. *Annu. Rev. Biophys. Biomol. Struct.* **23**, 703–730
47. Phan, A. T., and Patel, D. J. (2003) Two-repeat human telomeric d(TAGGGTTAGGGT) sequence forms interconverting parallel and antiparallel G-quadruplexes in solution. Distinct topologies, thermodynamic properties, and folding/unfolding kinetics. *J. Am. Chem. Soc.* **125**, 15021–15027