

U3 Region in the HIV-1 Genome Adopts a G-Quadruplex Structure in Its RNA and DNA Sequence

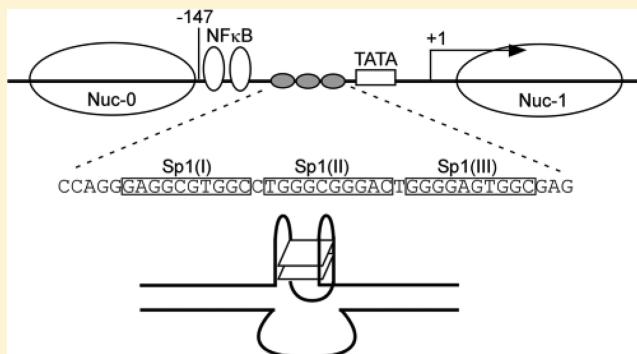
Dorota Piekna-Przybylska,[†] Mark A. Sullivan,[†] Gaurav Sharma,[‡] and Robert A. Bambara*,[†]

[†]Department of Microbiology and Immunology, School of Medicine and Dentistry, University of Rochester, Rochester, New York 14642, United States

[‡]Department of Electrical and Computer Engineering, University of Rochester, Rochester, New York 14627, United States

Supporting Information

ABSTRACT: Genomic regions rich in G residues are prone to adopt G-quadruplex structure. Multiple Sp1-binding motifs arranged in tandem have been suggested to form this structure in promoters of cancer-related genes. Here, we demonstrate that the G-rich proviral DNA sequence of the HIV-1 U3 region, which serves as a promoter of viral transcription, adopts a G-quadruplex structure. The sequence contains three binding elements for transcription factor Sp1, which is involved in the regulation of HIV-1 latency, reactivation, and high-level virus expression. We show that the three Sp1 binding motifs can adopt different forms of G-quadruplex structure and that the Sp1 protein can recognize and bind to its site folded into a G-quadruplex. In addition, a c-kit2 specific antibody, designated hf2, binds to two different G-quadruplexes formed in Sp1 sites. Since U3 is encoded at both viral genomic ends, the G-rich sequence is also present in the RNA genome. We demonstrate that the RNA sequence of U3 forms dimers with characteristics known for intermolecular G-quadruplexes. Together with previous reports showing G-quadruplex dimers in the gag and cPPT regions, these results suggest that integrity of the two viral genomes is maintained through numerous intermolecular G-quadruplexes formed in different RNA genome locations. Reconstituted reverse transcription shows that the potassium-dependent structure formed in U3 RNA facilitates RT template switching, suggesting that the G-quadruplex contributes to recombination in U3.



Recent cellular research revealed that G-quadruplexes formed in promoter regions of cancer-related genes regulate their expression.^{1–11} Formation of this structure is often linked to inhibition of transcription, but stimulation of promoter activity was also demonstrated.^{1,12–19} A G-quadruplex is assembled from two or more G-quartets with compact square structure, in which four guanines from different positions in a G-rich strand are held together by Hoogsteen hydrogen bonding (Figure 1). The G-quadruplexes differ by folding pattern, number of tetrads, size of nontetrad loops, and orientation of the strands in the quadruplex. In addition, whereas most reports show the structure core formed by guanines from G runs (two or more consecutive Gs), unprecedented and bulged G-quadruplexes were also reported with an isolated guanine involved in G-tetrad core formation.^{20,21} The DNA sequence can adopt this non-B configuration when complementary strands are separated in the DNA duplex during transcription and replication. The genomic regions prone to adopt this structure are rich in G residues, and include telomeres and gene promoters. In the case of promoters, multiple Sp1-binding motifs arranged in tandem are often indicated by computational analyses to form G-

quadruplexes, and promoters of cancer-related genes were shown to form this structure in Sp1 binding regions.^{1,12–17}

For RNA sequences, G-quadruplexes were detected in coding and noncoding regions of mRNA, and found to regulate protein synthesis.^{22–25} Formation of this structure in introns was suggested to influence alternative splicing.^{26–29} In HIV-1, sequences prone to adopt G-quadruplex structure are in a region of gag near DIS and in the central part of the genome near the cPPT.^{30–34} In both locations, G-quadruplex formation was associated with dimerization of the homologous templates and increased rate of primer-strand transfers during reverse transcription, suggesting that *in vivo* the structure contributes to dimerization of the viral genomes that promotes recombination.

The ability of the proviral DNA U3 region to adopt G-quadruplexes was recently reported by Perrone and co-workers, who described two parallel-like intramolecular G-quadruplexes, and showed that G-rich sequences of an NF-κB site together with G runs of Sp1 sites are involved in the quadruplex

Received: December 16, 2013

Revised: March 31, 2014

Published: April 7, 2014

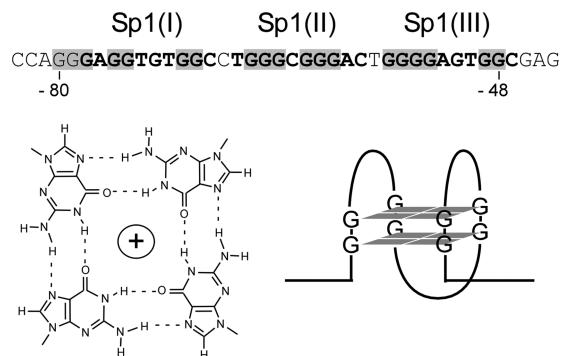


Figure 1. Guanine-rich sequence of the HIV-1 U3 region of the provirus and in the RNA genome might fold into a G-quadruplex. According to QGRS Mapper, runs of G residues (shaded) in three Sp1 binding sites (in bold) in the virus promoter are capable of forming a G-quadruplex. Four guanines are connected through hydrogen bonding to form a single G-quartet stabilized by a monovalent cation (left). Layering of two or more G-quartets forms a G-quadruplex (right).

structure.³⁵ Going beyond this work, we explored the relationship between the formation of G-quadruplex structure in the U3 DNA Sp1 transcription factor binding site region and binding of Sp1. Our native gel analyses, c-kit2 antibody binding analyses, and CD spectra show that the region fully transforms into different forms of intramolecular G-quadruplexes, which likely include a mixture of parallel/antiparallel and/or hybrid configurations that constitute the Sp1 binding site. Additionally, we investigated whether G-quadruplex formation in U3 RNA promotes viral recombination and the implications for a general viral recombination mechanism mediated by periodically spaced genome linkages including those that occur in U3 together with previously reported linkages in the *gag* and cPPT regions.^{30–32,34}

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 NC (55 amino acids) was generously provided by Dr. Robert J. Gorelick (NCI, Frederick, MD). HIV-1 reverse transcriptase (p66/p51 heterodimer) (RT) was purified as described previously.³⁶ The [γ -³²P]ATP was purchased from PerkinElmer Life Sciences. Recombinant Sp1 protein was purchased from Active Motif (Carlsbad, CA). Sp1 polyclonal antibodies and antirabbit IgG, HRP-linked antibodies were purchased from Cell Signaling Technology, Inc. (Danvers, MA). The promoter sequences of viruses are from 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, Inc.) 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 RNA template with the region of three Sp1 binding sites (8960–9051 in the RNA genome) of NL4–3 HIV-1 was made from a DNA template synthesized with the oligomer pair 1/2. (b) For affinity selection analysis, the nontagged RNAs and poly(A) tagged RNAs were made from DNA generated using oligomers 1/3 and 1/4 representing the U3 region (8960–9037) of NL4–3

HIV-1, 5/6 and 5/7 representing the cPPT region (4309–4396) of NL4–3 HIV-1, 8/9 and 8/10 representing the *gag* region (290–403) of NL4–3 HIV-1, and 11/12 and 11/13 representing the *gag* region (303–415) of MAL HIV-1. (c) For the analysis of dimerization in a native gel, the RNA template of the U3 region (8960–9037) in NL4–3 HIV-1 was made from DNA generated using oligomers 1/3. (d) For strand transfer assays, the donor and acceptor templates representing the U3 region (8960–9051 and 8961–9033) of NL4–3 HIV-1 were made from DNA generated using oligomers 1/2 and 14/15, 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 and RNA Oligonucleotides. Sequences of oligonucleotides are in Table S1 (Supporting Information).

Preparation of the 5'-Radiolabeled RNA Template and DNA Primer. DNA oligomers (16, a–e, c-kit, TBA, and TEL) were labeled at the 5' end using T4 polynucleotide kinase (New England BioLab) and [γ -³²P]ATP (6000 Ci/mmol). Preparation of the 5'-radiolabeled RNA template was performed as follows: The gel-cleaned RNA template was treated with shrimp alkaline phosphatase (SAP, Fermentas) at 37 °C for 60 min and then incubated at 65 °C for 25 min to inactivate the enzyme. Following cooling on ice, the reaction mixture was treated with [γ -³²P]ATP (6000 Ci/mmol), 10× PNK buffer, and T4 polynucleotide kinase (New England BioLab). 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 Transcriptase (RT) Progression Assay. The assay was performed in two steps (1) folding/annealing, and (2) primer extension. For folding/annealing, the RNA or DNA templates (2 pmol) were mixed with 5' end labeled primer DNA (2 pmol) in the presence of 0.5 M salt (KCl or LiCl) and 50 mM of Tris HCl (pH 8.0) in a volume of 20 μ L. The mixtures were heated to 95 °C for 5 min and cooled slowly to room temperature. The RNA template with three Sp1 binding sites was synthesized from the PCR product primed with oligomers 1 and 2 (see above). As a DNA template with three Sp1 binding sites, we used oligomer 15. After the annealing/folding step, 2 μ L of the mixture was taken for the RT-catalyzed primer extension reaction carried out in 25 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, reactions were stopped with 1 volume of termination buffer (10 mM EDTA, pH 8.0, 90% formamide (v/v), and 0.1% each xylene cyanole and bromophenol blue). Extension products were resolved on a 6% polyacrylamide–8 M urea gel and analyzed using a PhosphorImager (GE Healthcare). Sizes of DNA products were estimated by using a 5'-radiolabeled 10 bp DNA ladder (Invitrogen).

Native Gel Analysis of Monomer G-Quadruplexes. A mixture of the 5' end ³²P-labeled (about 300,000 cpm) and unlabeled DNA oligonucleotides (a–e, c-kit, TBA, TEL) at a concentration of 4 μ M and in a final volume of 25 μ L was heated to 95 °C for 3 min, chilled, and incubated for 20 h at room temperature in a buffer containing 10 mM Tris-HCl, pH 7.5, 1 mM EDTA, and 100 mM KCl. Samples were subsequently mixed with one volume of loading buffer (30% glycerol in Tris-EDTA buffer) and applied onto a 15% nondenaturing gel with 0.5× TBE, 10 mM KCl, and 1.5%

glycerol, and run at 4 °C at 6.5 V/cm for 18 h. The gel was dried onto Whatman 3MM Ch paper and analyzed using a PhosphorImager (GE Healthcare).

Circular Dichroism. CD spectra were obtained at 25 °C over a wavelength range of 210–340 nm using an AVIV Circular Dichroism Spectrometer, Model 202. The RNA (oligomers 17 and 23) and DNA (oligomers a–e) samples were at a concentration of 4 or 20 μM, in 10 mM Tris HCl, pH 7.5, 0.3 mM EDTA, and 100 mM KCl. Before analysis, the samples were heated to 90 °C for 10 min and 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 baseline-corrected for signal contributions from the buffer. The data were processed with AVIV Biomedical Inc. software and reported as ellipticity (mdeg) versus wavelength (nm).

Production of Phage Displaying hf2 scFv. The plasmid pIT2 with hf2 sequence was generously provided by Dr. Shankar Balasubramanian.³⁷ The hf2 phagemid was used to produce phages displaying the scFv after infection as previously described³⁸ except that the VCS M13 helper phage (Agilent) was used instead of the KM13 helper. The phages were isolated from the culture supernatants by PEG precipitation and resuspended in 50 mM K₂HPO₄ at pH 7.4 and 100 mM KCl containing 3% BSA.

Phage ELISA. The 10 μM stock solutions of biotinylated oligos b–e (20, 24, 25, and 26), single stranded DNA control (27), and c-kit2 were prepared in 10 mM Tris-HCl at pH 7.4 and 100 mM KCl. The samples were heated to 95 °C for 10 min and annealed over 14 h (o/n) at a rate of 0.1 °C/min down to room temperature in a buffer of 10 mM Tris-HCl at pH 7.4 and 100 mM KCl. In order to form a double stranded DNA, oligos 28 and 29 were mixed in a buffer of 10 mM Tris-HCl at pH 7.4 and 100 mM KCl, then heated at 95 °C for 10 min and cooled to room temperature at a rate of 1 °C/min. For phage ELISA, samples were diluted to 50 nM in 10 mM Tris-HCl at pH 7.4 and 100 mM KCl. The standard ELISA protocol was followed, but 50 mM K₂HPO₄ at pH 7.4 and 100 mM KCl (ELISA buffer) instead of PBS was used to maintain the G-quadruplex conformation. Pierce Streptavidin coated high binding capacity (HBC) strips were coated with biotinylated oligonucleotides for 1 h, then washed three times with ELISA buffer. Wells were blocked (3% BSA in ELISA buffer) for 1 h and then incubated with 50 μL of 2-fold serial dilutions of phages for 1 h. The transductional titer (ampicillin resistance) of the phages was about 7 × 10¹¹ transducing units/mL. After 6 washes (ELISA buffer), wells were incubated with a 1:3,000 dilution of anti-M13-HRP antibody (GE Healthcare) in ELISA buffer + 3% BSA for 1 h. After three washes (ELISA buffer), the ELISA was developed with the substrate TMB. Color development was terminated after 5–10 min with acid, and the absorbance at 450 nm was measured with a plate reader (Tecan).

Pull-Down of Sp1 and Western Blotting. The Sp1 was selected with biotinylated oligonucleotides and streptavidin-coated magnetic beads (Promega). The beads were washed three times with 0.5 mL of 0.5 × SSC buffer and three times with buffer A (25 mM HEPES at pH 7.5, 12.5 mM MgCl₂, 20% v/v glycerol, 0.1% v/v Nonidet P-40, 1 mM dithiothreitol, and 100 mM KCl) containing 3% of BSA. Before the binding of biotinylated DNA to the beads, oligomers c (20) and c-kit (18) were incubated overnight in buffer A to form the G-quadruplex

structure. In order to form dsDNA for protein selection, the pairs of oligomers 18/19 and 20/21 were incubated in buffer A in a ratio of 1:3. Biotinylated DNA (200 μL) samples (1 μM) were incubated with the beads for 30 min at room temperature. Beads were then washed three times with 500 μL of buffer A containing 3% of BSA and blocked with the same buffer for 30 min. All subsequent procedures were performed at 4 °C. Sp1 protein (80 ng from Active Motif) was added to 500 μL of buffer A with 3% of BSA. The mixture was added to the beads and incubated for 20 min, followed by washing six times with 200 μL of buffer A. The beads were resuspended in 20 μL of Laemmli buffer and boiled for 2 min. After removing the beads, the samples were separated on a 4–12% gradient Tris-Glycine polyacrylamide gel (BioRad) and transferred to a PVDF membrane. SP1 was identified by immunoblotting using a rabbit polyclonal antibody diluted 1:5000. A goat antirabbit secondary antibody linked with HRP was used for chemiluminescent detection. For competition binding with Sp1, oligomer c and a nonspecific sequence (oligomer 22) were first incubated under G-quadruplex forming conditions (see above) and mixed with biotinylated dsDNA samples before incubation with buffers containing Sp1.

Affinity Selection with Oligo d(T)₂₅ Magnetic Beads. About 40 pmol of poly(A)-tagged RNA and nontagged RNA templates 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 μL. 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 μL was washed once with binding buffer (20 mM Tris-HCl, pH 7.5, 500 mM KCl, and 1 mM EDTA), resuspended in 180 μL 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, and 1 mM EDTA), each time for 1 min with gentle agitation. In order to elute the RNA, the beads were resuspended in 15 μL of elution buffer (20 mM Tris-HCl, pH 7.5, and 1 mM EDTA), incubated for 3 min in 95 °C, and placed in a magnetic rack to separate the magnetic beads from solution. One volume of loading buffer (10 mM EDTA, pH 8.0, 90% formamide (v/v), and 0.1% each of xylene cyanole and bromophenol blue) was added to the eluted solution, and the products were resolved in 6% polyacrylamide–8 M urea gels. The gels were stained with ethidium bromide.

Cation-Dependent Dimerization and Thermal Dissociation Analysis. Dimerization and melting experiments with RNA dimers of the U3 region were conducted in parallel for each reaction setting. The samples contained a mixture of 5' end ³²P-labeled (about 500,000 cpm) and unlabeled RNA at a concentration of 4 μM and a final volume of 6 μL. To form a dimer, the RNA 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 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 one volume of 2× Tris-EDTA buffer was added. Aliquots of 15 μL were transferred to new tubes and incubated for 8 min at a specific temperature between 30 and 90 °C, then returned to the ice. Samples were mixed with one volume of loading dye (30% glycerol in 1× Tris-EDTA buffer and relevant dyes), and loaded

Table 1. G-Rich Sequences in the Sp1 Binding Region in Different Immunodeficiency Viruses^a

-80	GGG	A GG	TGT	GG	CCT	GGG	C	GGG	ACT	GGGG	AGT	GG	-48	NL43
-83	GGGGG	A GG	TGT	GG	TTT	GGG	CC	GG	AGTT	GGGG	AGT	GG	-48	HIV-1 A4 AM000053
-81	GGGG	A GG	TGT	GG	TTT	GGG	C	GG	AGTT	GGGG	AGT	GG	-48	HIV-1 A1 AB098332
-81	GGGG	A GG	TGT	GG	TTT	GGG	C	GG	AGTC	GGGG	AGT	GG	-48	HIV-1 A1 AF413987
-80	GGG	A GG	CCT	GG	CCT	GGG	C	GGG	ACT	GGGG	AGT	GG	-48	HIV-1 B AB253432
-80	GGG	A GG	CGT	GG	CCT	GGG	C	GGG	ACT	GGGG	AGT	GG	-48	HIV-1 B K03455
-79		GGGGG	TGT	GG	TCT	GGG	C	GGG	ACTA	GGG	AGT	GG	-48	HIV-1 C AX228556
-84	GG	A GG	AGT	GG	TCT	GGG	C	GGG	AC	GGGG	AGT	GG	-53	HIV-1 C AF290027
-78	GG	A GG	AGT	GG	TCT	GGG	C	GGG	ACT	GGG	AGT	GG	-48	HIV-1 C EU293445
-78	GG	A GG	TGT	GG	TCT	GGG	C	GGG	ACT	GGG	AGT	GG	-48	HIV-1 CD AY445524
-80	GGG	A CG	CGTAA	CC	GGG	C	GGG	ACA	GGGG	AGT	GG	-48	HIV-1 D AJ320484	
-66		GG	CGTGA	CCT	GGG	C	GGG	ACT	GGGG	AGT	GG	-48	HIV-1 D DQ054367	
-78	GGG	C GGG		CCAGA	GGG	CCTG		ACT	GGGG	AGT	GG	-48	HIV-1 F1 FJ771006	
-81	GGGG	A GG	CGC	GG	CCT	GGG	A	GGGG	CT	GGGG	AGT	GG	-48	HIV-1 G AB231893
-81	GGGG	A GG	CGC	GG	CCT	GGG	A	GGGG	CT	GGGG	AGT	GG	-48	HIV-1 G AF084936
-81	GGGG	A GG	CGTGA	TCT	GGG	C	GGGG	AT	GGGG	AGT	GG	-48	HIV-1 H AF005496	
-78		GGGG	TGT	GG	TCT	GGG	C	GGGG	TT	GGGG	AGT	GG	-48	HIV-1 H FJ711703
-80	GGG	A GG	TGTCA	CCT	GGG	C	GGGG	TT	GGGG	AGT	GG	-48	HIV-1 U EF029067	
-80	GGG	A GG	TGTTA	CCT	GGG	C	GGGG	TT	GGGG	AGT	GG	-48	HIV-1 U EF029069	
-69	GGG	A GG	TGT				GG	TT	GGGG	AGT	GG	-47	HIV-1 N AJ006022	
-76	GGG	A GGG		ACA	GGGG	C	GG	TTC	GGGG	AGT	GG	-48	HIV-1 O AJ302646	
-79	GG	T GGG	A	GGG	ATAT	GGG	C	GG	TTC	GGGG	AGT	GG	-48	HIV-1 O AY618998
-75	GG	T GG	TC	GGG	A	GG	C	GGGG	TT	GGGG	TGT	GG	-47	SIV(cpz) X521154
-77	GGG	T GG		TTACT	GGG	C	GGG	ACA	GGGG	AGT	GG	-48	SIVcpzTAN1 AF447763	
-78	GGGG	T GGG		TCA	GGG	C	GG	AACAA	GGG	CGT	GG	-48	SIV AF103818	
-78	GGG	A GGG		ACAT	GGG	A	GG	AGCT	GG	T	GGGG	-51	HIV-2 BEN NC 001722	
-80	GGG	A GGG		ACAT	GGG	A	GG	AGCT	GG	T	GGGG	-53	HIV-2 ROD M15390	

^aNumbers indicate locations with reference to the transcription start site. G runs are shown in gray boxes.

onto 6% nondenaturing gels run at 4 °C at 7 V/cm for 3–4 h. All gels were dried onto Whatman 3MM Chr paper and analyzed using a PhosphorImager (GE Healthcare).

Strand Transfer Assay. The 5' end labeled DNA primer 16 was heat-annealed to donor RNA by incubation at 95 °C for 5 min and slow cooling to 37 °C. The acceptor template was also present in the mixture. NC at 200% polymer substrate-coating level (100% NC is 7 nt of the polymer substrate per NC molecule) was added and incubated for 3 min. Next, the RT was added to the mixture and incubated for another 4 min to prebind the RT with the substrates, before reactions were initiated with MgCl₂ and dNTPs. Primer, donor, and acceptor strands 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, 50 μM dNTPs, 16 nM donor RNA, 16 nM acceptor RNA, and 32 nM primer. For reactions in the presence of lithium ions, KCl was replaced with 50 mM of 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 cyanole and bromophenol blue). Products were then resolved by 6% polyacrylamide–8 M urea gels and analyzed using a PhosphorImager (GE Healthcare) and ImageQuant software (version 2.1). Sizes of DNA products were estimated by using a 5'-radiolabeled 10 bp DNA ladder (Invitrogen).

RESULTS

G-Rich Sequences Capable of Forming G-Quadruplex Structure Are Present in the U3 Regions of Various HIV Species. Intensive research on G-quadruplexes showed that genomic regions prone to adopt this structure correlate particularly with gene promoters and telomeres, both rich in G residues. More importantly, cellular research revealed that G-quadruplexes in promoter regions of cancer-related genes regulate their expression.^{1–11} Since sequences capable of forming G-quadruplexes were previously found in the HIV-1 genome, we wanted to determine whether this structure might also be formed in the viral promoter.^{30–32,34} Our computational analyses with a software program designed to predict the formation of G-quadruplexes (QGRS Mapper³⁹) revealed that the promoter sequence in HIV-1 NL4–3 U3 between -80 and -48 (+1 refers to transcription start site at the beginning of R),

has a high probability of adopting different forms of G-quadruplex structure (Figure 1). The same region was also predicted to form G-quadruplexes in the genomes of various isolates of HIV-1 (A, B, C, D, F, G, H, U, N, and O), HIV-2, and several SIV species closely related to HIV-1 (Table 1). This region shows a higher variation in primary sequence among different viruses than the protein coding regions. Yet, the region from all of these viruses retains the ability to form G-quadruplex structure as indicated by QGRS Mapper. This suggests that the G-rich part of the promoter is prone to adopt G-quadruplex structure and that the structure regulates viral expression. The G-rich promoter sequence in HIV-1 contains from five to seven G runs and contains three Sp1 binding sites. Some of these G runs are composed of only two guanines, suggesting that the putative G-quadruplex is composed of only two G-quartets. However, single G residues near G runs might also be included in a structure, which is called the bulged G-quadruplex. An extra G tetrad in a core formed with an isolated G residue (not a part of G run) will produce a bulge with non-G residue between tetrads.²¹ Since in retroviruses the U3 is present in both the proviral LTRs, the Sp1 binding sites should be also capable of forming a G-quadruplex at the 3' end of the RNA genome (8996–9028 in RNA HIV-1 NL4–3).

G-Rich Sequences in U3 of HIV-1 RNA and Single Stranded DNA Form G-Quadruplex Structure. The distinctive signature of cation-dependent pauses during RT progression serves as a simple test to verify whether the G-rich region in a template sequence can form G-quadruplex structure.^{33,34,40} In the presence of K⁺ or Na⁺, the G-quadruplex structure is formed in the template, and it pauses RT during cDNA synthesis. In the presence of lithium ions, the structure is not formed, so G-quadruplex pausing does not occur. With lithium, higher salt and template concentration can also induce G-quadruplex formation. The RT progression assay is performed in a low salt (50 mM) and template concentration (16 nM) so that ion type-dependent pausing is clearly evident. Pausing caused by hairpin structures occurs with either ion, allowing it to be readily distinguished.

In order to assess G-quadruplex formation in HIV-1 U3 RNA, a 92-nt long U3 sequence (8960–9051 in the RNA genome) was synthesized by transcription *in vitro*. For analysis of G-quadruplex formation, in the corresponding single stranded DNA we made an oligonucleotide with the sequence

of the region from -116 to -25 in the HIV-1 promoter. HIV-1 RT-directed cDNA synthesis on these RNA and DNA strands was performed in the presence of K^+ or Li^+ , and results in Figure 2 show that cation-dependent pauses of RT were

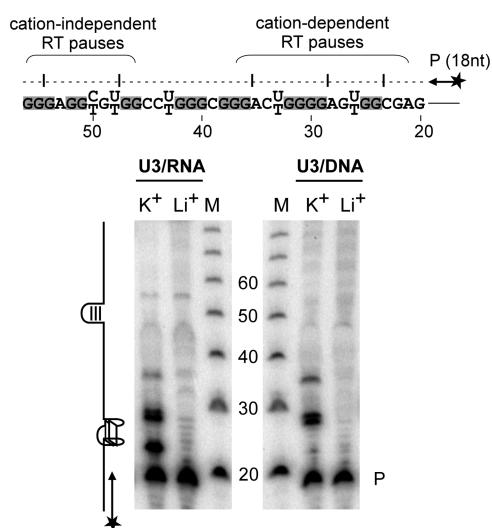


Figure 2. RT progression assay determining G-quartet formation in the Sp1 binding region. Cation-dependent pausing of reverse transcription at the guanine-rich elements in the U3 region was analyzed with RNA and DNA templates. A fragment of the RNA/DNA sequence of the Sp1 binding region is shown (top) with G-rich elements (shaded). Strong pauses of the RT near G-rich elements were observed in the presence of 50 mM of KCl , but not $LiCl$, indicating that these elements are involved in the formation of structure, which is stabilized by potassium ions but destabilized by lithium ions. This is indicative of a G-quadruplex. The cation-independent RT pauses are likely caused by hairpin structures. DNA primer, P; DNA marker, M; K^+ ; $LiCl$, Li^+ .

observed on both templates. Two strong pauses of RT were produced solely in the presence of potassium on the RNA molecule. The first pause is located between marker 20nt and 30nt. In the sequence, this region roughly corresponds to the first G run, which is at position 24. The second strong pause is observed close to marker 30nt and roughly corresponds to the second G run, which starts at 29 indicated on the sequence. A weaker third RT pause is also observed between marker 30nt and 40nt and might correspond to the third G run, which starts at 34 on the sequence. Since these RT pauses are not observed in the presence of lithium ions, they must be caused by a structure formed in the presence of potassium but not lithium. Such a structure is a G-quadruplex. The RT pause profiles for RNA and DNA look similar except for an additional pause site at the first G run encountered in the RNA template. The cation-independent RT pauses occurring in reactions done with either potassium and lithium ions likely result from hairpin structures. Hairpins are known to pause RT, and their formation is dependent on ionic strength but not the identity of the ions.³³

To further confirm the formation of G-quadruplex structure in the HIV-1 U3 region, we analyzed the folding of RNA and DNA sequences by circular dichroism (CD) spectroscopy. CD spectroscopy is widely used to determine G-quadruplex formation in RNA and DNA. The CD spectrum of the sequence folded into a G-quadruplex with a parallel configuration has a positive peak near 263 nm and a negative

peak at 241 nm. Similar peaks appear near 295 and 260 nm, respectively, for antiparallel G-quadruplexes, with an additional peak around 240 nm. The parallel configuration refers to the structure in which the 5'-3' direction of all the strands that form G-quartets is the same. If one or more strands have a 5'-3' direction opposite to the other strands forming G-quartets, the G-quadruplex is said to have adopted an antiparallel topology. However, G-quadruplexes are highly polymorphic, and more complicated forms can be adopted displaying altered CD profiles. CD spectra with two positive peaks around 295 and 263 nm, might indicate that the sequence adopts parallel and antiparallel G-quadruplexes. However, bulged G-quadruplexes and hybrid configurations with strands in parallel and antiparallel in the single G-quadruplex also display a similar profile.^{21,41} Different configurations are seen for G-quadruplexes folded by DNA molecules, but for RNA sequences, all G-quadruplexes described so far exhibit parallel topology.

Our CD structural analysis of a 35-nt long RNA with the sequence of HIV-1 U3 (8995–9029) in 100 mM KCl showed a typical profile of a parallel G-quadruplex, with a positive and negative ellipticity at 263 and 241 nm, respectively (Figure 3).

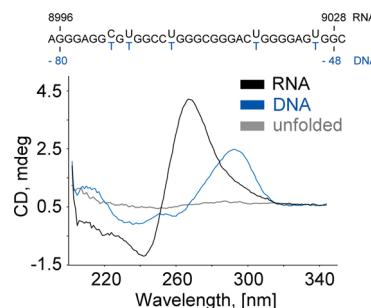


Figure 3. CD spectral analysis of the RNA and single stranded DNA with Sp1 binding sites in HIV-1. CD spectra indicate the formation of the parallel G-quadruplex for the RNA template and an antiparallel or hybrid G-quadruplex for single stranded DNA. For reference, a profile of the G-rich sequence (50% of Gs; GGGGGGAUUGUG UGGUACAGUGCAGAGA), which is unable to adopt G-quadruplex structure, is shown in gray.

However, the CD spectrum of the DNA sequence showed a positive peak around 295 and a negative peak around 260 nm indicating a G-quadruplex with antiparallel topology, although it is possible that the structure is a hybrid with strands in antiparallel and parallel orientations since a peak at 240 nm is not clearly seen.⁴¹ These results confirm that both RNA and single stranded DNA of the HIV-1 U3 region containing the Sp1 binding elements form G-quadruplexes but with different configurations.

Single Stranded DNA Containing the Three HIV-1 Sp1 Sites Adopts Different Monomer G-Quadruplex Configurations. Our cation-dependent RT progression analysis and CD spectra show that a G-quadruplex is formed by the HIV-1 promoter single stranded DNA template. This implies that the DNA sequence adopts a G-quadruplex configuration when complementary strands are separated in the DNA duplex during transcription. The G-rich sequences in promoters of cancer-related genes have been shown to form monomer quadruplex structures.^{2,4,5,8–11} Since the G-quadruplex monomer is expected to have a biological significance in the HIV-1 promoter, we used native gel analysis to determine how efficiently it is formed over dimers and tetramers, which are

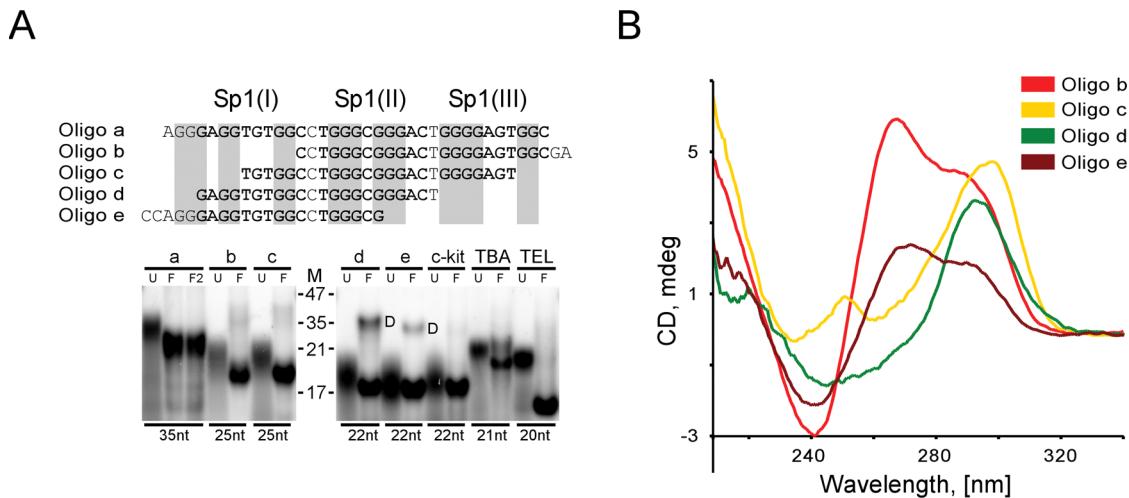


Figure 4. Single stranded DNA of the HIV-1 promoter with Sp1 binding sites folds completely into different G-quadruplex monomers, and each of seven G runs participates in the formation of G-quartets. (A) Native gel analysis showing that the three Sp1 sites adopt monomer G-quadruplexes with different sets of G runs (oligonucleotides b–e). A sequence folded into a monomer G-quadruplex (F) migrates faster than expected for its size. The nonfolded (U) controls of the b, c, d, e, and c-kit promoter also migrate faster and appear as smear bands indicating that 10 mM KCl in the gel induced their folding into G-quadruplex during electrophoresis. The sequences of the c-kit promoter, altered thrombin aptamer (TBA) and telomere (TEL), were confirmed to fold into monomer G-quadruplexes and were used as positive controls. Dimeric G-quadruplexes (D) were observed for oligonucleotides d and e. F, samples incubated in the presence of 100 mM KCl; F2, sample incubated in the presence of 250 mM KCl; and U, samples incubated in the absence of KCl. (B) CD spectra for shorter sequences b–e. Oligonucleotides b and e display a profile known for mixed G-quadruplexes with either parallel or antiparallel configuration and also known for bulged and hybrid G-quadruplexes. CD profiles for oligonucleotides c and d are similar to profiles of G-quadruplexes with antiparallel topology and hybrid or basket-type configuration, respectively.

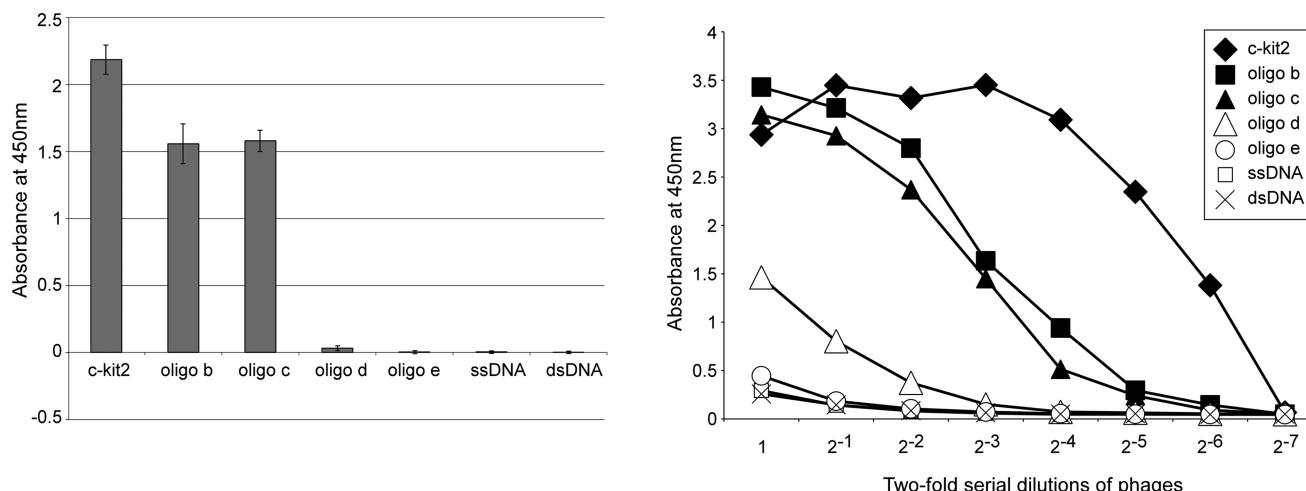


Figure 5. Interactions between phages displaying the hf2 scFv and G-quadruplexes formed in the HIV-1 Sp1 sites. (A) Phage ELISA shows that the c-kit2 G-quadruplex-specific antibody recognizes two of four G-quadruplexes formed by sequences of the HIV-1 Sp1 sites. (B) Binding curves of hf2-DNA interactions with serial dilutions of phages show that bindings of hf2 to oligonucleotides b and c have lower affinities when compared to binding with the c-kit2 G-quadruplex. No significant interactions were observed for oligonucleotide d, and no binding was seen for oligonucleotide e and two controls, single stranded (ssDNA), and double stranded DNA (dsDNA).

intermolecular forms of G-quadruplex that do not require multiple G runs. In this approach, the ^{32}P -labeled strand is first incubated under G-quadruplex folding conditions and then loaded into a native gel to separate monomers, dimers, and tetramers from nonfolded molecules. The gel contains KCl at a concentration of at least 10 mM to maintain the integrity of folded structures. The monomer G-quadruplex appears in the gel as the fastest migrating band, then the nonfolded sequence, whereas dimers and tetramers are slower migrating forms.

Because the Sp1 binding elements of the HIV-1 promoter have seven G runs, to gain information on their individual roles in a G-quadruplex formation, we analyzed several DNA

oligonucleotides covering different sets of four G runs. The sequences of the c-kit promoter (Sp1 site),⁴² altered thrombin aptamer (TBA),⁴³ and telomere (TEL),⁴⁴ had been confirmed to fold into monomer G-quadruplexes and so were used as positive controls. Results in Figure 4A show that the single stranded DNAs with Sp1 binding elements migrate faster than expected when incubated in the G-quadruplex folding condition. Formation of G-quadruplex by shorter sequences (b–e) indicates that each of the seven G runs can be involved in the formation of a G tetrad. The nonfolded samples, which are samples incubated in the absence of salt, do not migrate according to their sizes and appear as smear bands for

oligomers d, e, and c-kit indicating that 10 mM KCl in a gel is sufficient to induce their folding into G-quadruplex during electrophoresis. The 10 mM KCl is sufficient to induce G-quadruplex folding in RNA and single stranded DNA.^{4,34,45} In addition to monomers, oligomers d and e can also form dimers.

The CD spectral analysis confirms the formation of G-quadruplexes for sequences b, c, and e (Figure 4B); however, the profiles differ for each sequence indicating that they exhibit different configurations. The CD spectra for oligomers b and e have two peaks around 295 and 260 nm, which might represent a mixture of parallel and antiparallel G-quadruplexes.⁴⁶ However, such profiles are also described for hybrid (3 + 1) and bulged G-quadruplexes.^{21,41} The profile for oligomer d has a strong maximum around 295 nm and a minimum around 250 nm, and such a profile is generally not observed for G-quadruplex structures, although few reports attributed these spectral characteristics to hybrid conformations containing a mixture of both parallel and antiparallel strand orientations and basket-type G-quadruplexes.^{47,48} Other techniques, like NMR spectroscopy would need to be used to determine what structure is formed by oligomer d. The CD spectrum for oligomer c is similar to the profile describing G-quadruplexes with antiparallel configuration. These results indicate that the Sp1 binding region can easily and fully adopt different monomer G-quadruplexes. The ability of the sequence to adopt different G-quadruplex configurations was previously reported and might reflect that G-quadruplexes with different topology could have different functions.⁴⁹

hf2 Antibody Recognizes Two G-Quadruplexes Formed in the Sp1 Binding Region of the HIV-1 Promoter.

Engineered proteins, such as single-chain antibodies and the Gq1 zinc finger protein, have been previously generated as molecular probes to study G-quadruplex structures.^{37,50} Using phage display technology, Fernando and co-workers generated an hf2 antibody with high binding affinity to the G-quadruplex formed by c-kit2 found in the promoter of the *c-kit* proto-oncogene.³⁷ Recent studies determined that this antibody also has binding affinity to G-quadruplexes formed in other genomic regions.⁴⁶ In order to determine whether hf2 would bind to structures formed in the Sp1 binding sites of HIV-1, we produced phages displaying hf2 antibodies on their surface and performed a phage ELISA against structures formed by oligomers b–e. Results showed that the hf2 displaying phage binds to G-quadruplexes formed by two HIV-1 sequences, the oligomers b and c (Figure 5A). Serial dilutions of phages showed that the highest affinity of hf2 antibody was seen for the G-quadruplex formed by the c-kit2 sequence, as expected (Figure 5B). A lower affinity was observed for structures formed by sequences of oligomers b and c, indicating that these G-quadruplexes must contain some structural features similar to those of the G-quadruplex formed by the c-kit2 sequence. A very weak binding affinity was detected for the structure formed by oligomer d, whereas no binding was detected for oligomer e and two controls, single stranded and double stranded DNA. As an additional control, we performed a phage ELISA for phages displaying unrelated antibodies. No binding was observed for all analyzed sequences (data not shown). In summary, our results show that the hf2 antibody is able to recognize and bind to two G-quadruplex structures of the Sp1 binding region in HIV-1 promoter. This directly proves that oligomers b and c adopt G-quadruplex configurations. The lack of interactions between antibody hf2 and oligomers d and e suggest that both sequences do not form G-quadruplexes. However, although the

hf2 antibody displayed ability to recognize different G-quadruplexes, likely they do not recognize all configurations. Thus, other methods would have to be used to provide direct evidence that oligomers d and e adopt this structure.

Sp1 Protein Binds to a G-Quadruplex Formed in the HIV-1 Promoter.

Recent studies showed that the Sp1 binding sites of the c-kit promoter fold into a G-quadruplex that is recognized and bound by Sp1 protein.⁴² The G-rich sequence of the c-kit Sp1 site and distribution of G runs are different from Sp1 sites in HIV-1. In order to determine whether Sp1 can also bind to an HIV-1 sequence folded into a G-quadruplex, we used a previously described affinity selection approach.⁴² In this method, the pull-down of protein is performed with the 3'-biotinylated oligonucleotide with the sequence folded into G-quadruplex and immobilized to streptavidin-coated magnetic beads. The Sp1 protein used for affinity selection is at a concentration of 1.98 nM.

As indicated above, the three Sp1 sites have the ability to adopt different G-quadruplex configurations, in which different sets of G runs participate in the formation of a G-quartet. However, for some configurations, the folding might involve G runs of two Sp1 binding sites leaving one site unfolded. To ensure that only one Sp1 site is available for protein binding in a pull-down of Sp1 and that it is folded, we used as bait a 21-nt sequence of oligomer c with the Sp1 site II surrounded by two G runs from Sp1 sites I and III (Figure 4). The native gel analysis shows that this sequence transforms completely into a monomer G-quadruplex (Figure 4A), and CD spectra indicate that the G-quadruplex is likely in an antiparallel configuration (Figure 4B).

In assessing Sp1 binding to a G-quadruplex in the HIV-1 promoter, we used, as a positive control, an oligonucleotide with the sequence of the c-kit promoter for which the binding of Sp1 to its site folded as a G-quadruplex had been confirmed.⁴² As a negative control, selection of Sp1 protein was done with beads not coupled to any DNA. As seen in the Western blots in Figure 6A, the Sp1 protein was pulled down with the HIV-1 Sp1 binding site II folded into a G-quadruplex with the same efficiency as with a G-quadruplex of the c-kit promoter. When the dsDNA of this sequence was used for Sp1 selection, the interaction of protein with the DNA was disrupted by the presence of the 3 and 6 molar excess of

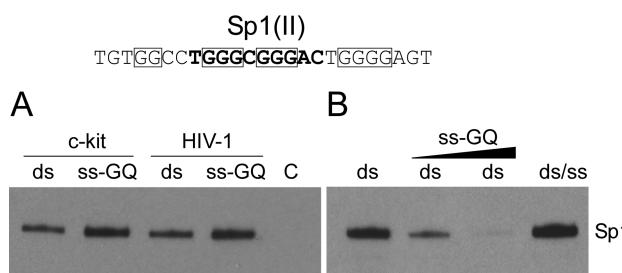


Figure 6. Sp1 binds to a G-quadruplex in the HIV-1 promoter. (A) Sp1 is selected by a G-quadruplex (ss-GQ) with the sequence of the Sp1 site II of the HIV-1 promoter (top) with the same efficiency as that selected by a G-quadruplex with the sequence of the c-kit promoter. Sp1 is also selected with the same efficiency by these sequences in dsDNA form. (B) Three and six molar excess of oligomer c with Sp1 site II (top) folded into a G-quadruplex competes efficiently with dsDNA sequence for binding with Sp1. Sp1 binding to the dsDNA is not affected in the presence of 6 molar excess of a nonspecific 18-nt sequence (T)₁₅CTA.

oligomer c with the Sp1 site II folded into a G-quadruplex (Figure 6B). This indicates that the Sp1 site in a G-quadruplex form competes efficiently with dsDNA for binding with Sp1. Importantly, the interaction of Sp1 with dsDNA was not affected by the presence of a 6 molar excess of the nonspecific sequence ($T_{15}CTA$; ds/ss line in Figure 6B). In summary, these results confirm that Sp1 protein can recognize and bind to its binding element in the HIV-1 promoter folded into a G-quadruplex configuration.

Dimerization of RNA Strands in the HIV-1 U3 Region.

Since the two copies of the HIV-1 RNA genomes are held together at DIS, the ability of the RNA strands to fold into a G-quadruplex configuration raises the possibility that multiple G-quadruplex structures formed between the two viral RNA genomes support additional dimer contacts during reverse transcription. In order to determine whether an intermolecular G-quadruplex in the U3 region can be formed between two RNA strands, we first used our previously developed affinity selection approach to test if two homologous sequences of U3 can interact.³⁴ In this method, the RNA strands tagged with poly(A) tail are incubated with nontagged RNA molecules, and then the mixture is subjected to affinity selection with magnetic beads conjugated with oligo-d(T)₂₅. The interaction between nucleic acids is measured by analyzing selected RNA molecules. The interacting partners are distinguished by their size in a denaturing gel stained with ethidium bromide. The quantity of selected nontagged template is expected to be lower since interactions will also occur between two tagged RNAs and two nontagged RNAs. In addition, the formation of an intermolecular G-quadruplex between two templates likely competes with the formation of intramolecular G-quadruplexes in both templates.

The RNA strands corresponding to positions 8960–9037 of the HIV-1 NL4–3 RNA genome were synthesized with poly(A) tails and were then coincubated with equivalent RNA strands but devoid of a poly(A) sequence. Because the gag regions near DIS from HIV-1 MAL and NL4–3, and the cPPT region in NL4–3 were shown to form G-quadruplex dimers, we used RNA strands of these regions as positive controls.^{30–32,34} The RNA strands with gag sequences (303–415 in HIV-1 MAL and 290–403 in HIV-1 NL4–3) did not include the DIS region. The RNA partners were combined and incubated to allow dimerization, and subsequently used for affinity selection with magnetic beads. As shown in Figure 7, the nontagged RNA strands of the U3 region were coselected with corresponding poly(A) tagged RNAs, similar to results with the gag and cPPT regions with which dimeric G-quadruplex formation was previously reported. The slower migration rate for tagged RNA of U3 region results from a longer poly(A) tag used for this template. Nontagged RNA with the U3 region sequence was not selected by magnetic beads in the absence of the poly(A) tagged partner, demonstrating that observed interactions are not resulting from nonspecific binding to the magnetic beads.

In order to determine whether RNA interactions in U3 have characteristics of intermolecular G-quadruplexes, the interactions were also investigated by native gel analysis. Previous studies showed that the ability of a test sequence to form a dimer through intermolecular G-quadruplex increased with template and salt concentration; however, the yield of RNA dimers correlated inversely with the size of monovalent cation (i.e., $Li^+ > Na^+ > K^+$).^{30,31,34} Thus, G-quadruplex RNA dimers can form more efficiently in the presence of a high

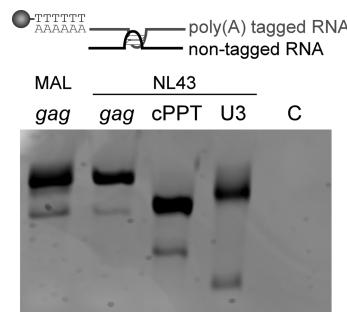


Figure 7. Affinity selection of HIV-1 RNAs enriched in G residues. Poly(A) tagged RNA templates are sequences elongated at the 3' end with a poly(A) tag in order to select them with oligo(dT)₂₅ magnetic beads. After incubation and affinity selection, the samples were analyzed in a denaturing gel stained with ethidium bromide. The nontagged RNAs (faster migrating species) of the gag region (RNA genomic sequence 303–415 of the MAL isolate and 290–403 of the NL4-3), cPPT region (4309–4396 of NL4-3), and U3 region (8960–9037 of NL4-3) were selected with corresponding poly(A) tagged RNA partners. No U3 region RNA was selected in the absence of a poly(A) tagged partner (lane C).

concentration of LiCl, than KCl, although complexes are less stable. To compare the cation-dependent association and thermal stability profiles of complexes formed by the G-rich U3 region, we used radiolabeled RNA templates and analyzed complexes in a native gel. A 78-nt HIV-1 fragment of the U3 region (8960–9037) at a concentration of 4 μ M in buffer containing 1 M of KCl, NaCl, or LiCl was heated to 95 °C, chilled on ice, and subsequently incubated for 1 h at room temperature. A 1 M concentration of salt was used to provide optimal conditions for the G-quadruplex folding in the presence of lithium ions. The mixtures were then exposed for 8 min to different temperatures, and the stability of the complexes was analyzed in a native gel. The results in Figure 8A show that dimer complexes were formed more efficiently in the presence of LiCl and NaCl, than in the presence of KCl. This observation is consistent with results obtained in previous studies for regions of gag and cPPT in HIV-1.^{30,31,34} With higher temperature, all complexes were less stable, but the rate of dissociation was higher for complexes formed in the presence of lithium ions. For example, about 77% of complexes formed in the presence of potassium ions were still present at 70 °C, whereas about 1/3 of complexes formed in the presence of sodium and about a half of those formed in the presence of lithium ions had dissociated at this temperature. The low level of RNA G-quadruplex dimers formed in the presence of potassium ions resulted from a slower rate of folding, and up to 50% yield of complexes could be achieved after 24 h of incubation at lower salt concentration (0.2 or 0.5 M) (Figure 8B). These results demonstrate that dimeric complexes formed by the RNA template of the U3 region display characteristics expected for intermolecular RNA G-quadruplex structures.

G-Quadruplex in RNA Facilitates RT in Switching Templates during Reverse Transcription. The folding of an RNA sequence into a structure that can pause RT is linked to an increased rate of viral recombination. Our previous studies showed that under conditions that encourage the formation of G-quadruplex structure, the rate of template switching during reverse transcription increased for the G-rich regions in gag and near the cPPT, suggesting that the structure increases the efficiency of recombination.^{33,34} These observations are in agreement with studies *in vivo* on the distribution of

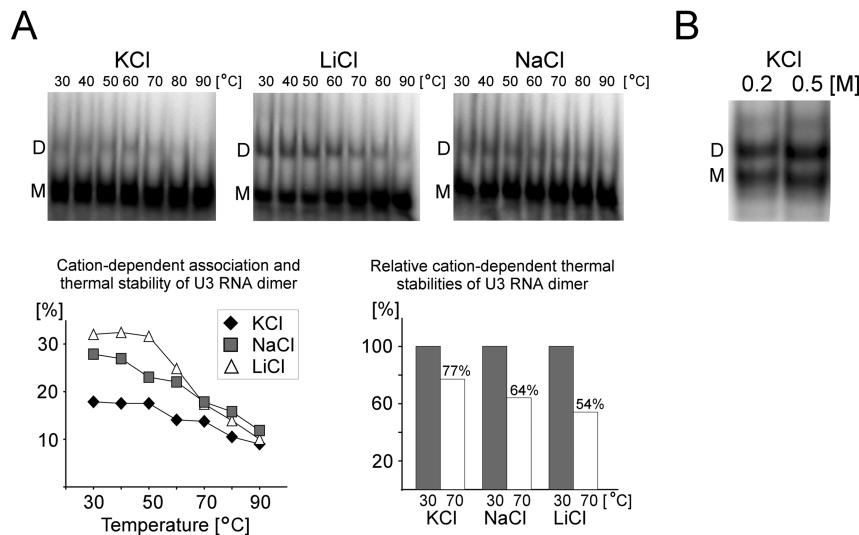


Figure 8. Cation-dependent association and thermal stability of the RNA dimer formed by a sequence with the Sp1 sites of HIV-1 U3. (A) RNA dimers were allowed to form for 1 h at a template concentration of 4 μ M in buffers containing 1 M KCl, NaCl, or LiCl. One volume of Tris-EDTA buffer was added, and 15- μ L aliquots were incubated at the indicated temperatures for 8 min. Thermal stabilities of RNA dimers were measured by analyzing samples in nondenaturing gels run at 4 °C. Higher yield of RNA dimers formed in the presence of 1 M LiCl and their lower thermal stability are characteristics of complexes formed through intermolecular G-quadruplexes. (B) The yield of RNA dimers increases after overnight incubation in the presence of potassium ions and at a lower salt concentration (0.2 and 0.5 M).

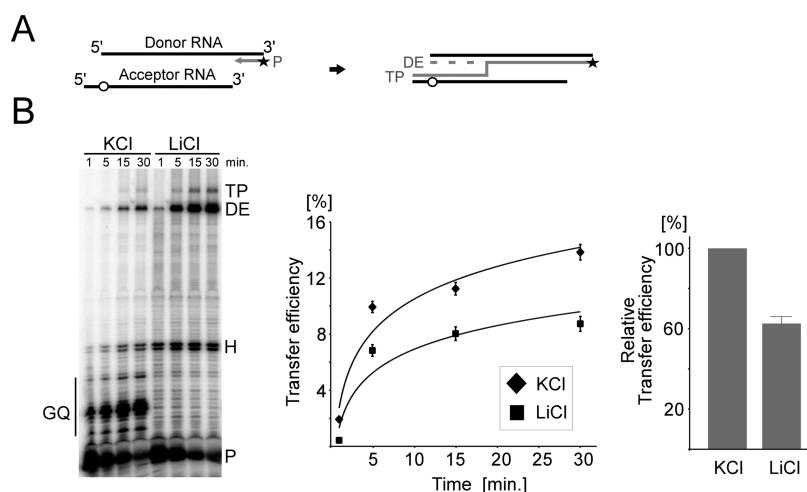


Figure 9. Formation of the structure stabilized by potassium ions in the HIV-1 U3 region facilitates RT template switching during reverse transcription. (A) Reconstituted system to analyze the influence of G-rich elements on strand transfer during HIV-1 minus strand DNA synthesis *in vitro*. Donor and acceptor RNA templates represent two copies of the viral RNA genome; in which reverse transcription is initiated from a 32 P-labeled DNA primer annealed to the donor RNA. The acceptor RNA does not share a homology (circle) with two nucleotides at the 5' end of the donor RNA. TP, transfer product; DE, donor extension product; and P, DNA primer. (B) 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 potassium-dependent structure, anticipated to be a G-quadruplex, in the RNA template paused the RT during minus strand DNA synthesis and influenced the yield of the final products. The transfer efficiency decreased about 37% in reactions with lithium ions, presumably because the templates could not form a G-quadruplex.

recombination breakpoints in the HIV-1 genome, which show a higher recombination rate in these two regions.^{51–53} The U3 region of the RNA genome is also a site of frequent recombination events, which generally occur upstream of the Sp1 binding sites.⁵⁴ Significantly, studies revealed that efficient crossovers in U3 rely on the presence of the 150-nt long sequence at the 3' end of U3, containing the G-rich elements of the Sp1 binding sites.⁵⁵

In order to evaluate the influence of G-quadruplex formation on strand transfer efficiency (the recombination reaction), we

used a reconstituted system consisting of HIV-1 RT, HIV-1 nucleocapsid protein (NC), a primer (DNA oligonucleotide), and two RNA templates representing the two copies of the HIV-1 RNA genome (Figure 9A). To initiate the reaction, a 32 P-labeled DNA primer was annealed to the RNA template (HIV-1 genome sequence 8960–9054), denoted here as donor RNA. The second RNA template, denoted as acceptor RNA, had the HIV-1 sequence 8961–9033 and shared homology with the donor RNA over 72-nt of the Sp1 binding region. The 5' end of the acceptor RNA was elongated with GGAAAA-

AAAAAA so that transfer products (TP) could be separated and distinguished on a denaturing gel from DNA fully extended on the donor RNA (DE). End transfer of the DE was prevented because our acceptor RNA did not share homology with two nucleotides at the 5' end of the donor RNA (a circle in Figure 9A). Thus, all transfers to the acceptor RNA originated only from internal regions of the donor template, as they do *in vivo* during reverse transcription over the U3 region.

The strand transfer assays were performed in the presence of a low concentration of either K⁺ or Li⁺. The monovalent cations added to the reaction do not significantly affect the enzymatic activity of RT.^{56,57} The major cation-dependent pause sites of RT synthesis were clearly visible within the G-rich region in the presence of K⁺ but not Li⁺ (Figure 9B). A nonion-dependent RT pause, likely resulting from a hairpin structure, was visible. The strong pauses in potassium, presumed to result from G-quadruplex formation, evidently caused dissociation of the RT since fewer final products were made when compared with that in reactions in lithium, where no RT pauses were seen in the G-rich region. The transfer efficiency of reactions was calculated by comparison of values of donor extension products and transfer products using the following formula: transfer efficiency = 100 × TP/(TP + DE). As expected, the elimination of ion-dependent RT pauses by lithium caused a drop in transfer efficiency of about 37%. Since ion-dependent RT pauses are associated with G-quadruplex formation, the effect of increased template switching likely resulted from RT encountering G-quadruplex. This is consistent with a general observation that a structure capable of pausing the RT also facilitates transfer reactions; however, the effect observed here is less striking than for two other G-quadruplex forming sequences in a region of *gag* and cPPT.^{33,34} In summary, these results show that G-quadruplex(es) formed at the 3' end of the viral RNA genome are likely contributors to the increased recombination rate in the U3 region.

■ DISCUSSION

Many promoters have multiple runs of G-residues, including those with several G-rich Sp1 transcription factor binding motifs in tandem. Using genome-wide computational analysis, it has been suggested that such tandem Sp1 binding sites form G-quadruplex structures.^{58,59} Direct evidence that Sp1 binding elements form G-quadruplexes that regulate gene expression has already been derived from studies of the *c-myc* promoter.^{1,12,15–17} The HIV-1 promoter has three Sp1 binding elements, and our analysis of the region in U3 revealed that the sequence readily adopts different G-quadruplex configurations within both its RNA and single stranded DNA forms. These results predict that during viral infection the G-quadruplex structure forms in both the RNA genome and in the promoter of the HIV-1 provirus, and presumably has important biological functions in both environments. In fact, recently Perrone and co-workers published results showing that HIV-1 promoter activity is impaired in the presence of a G-quadruplex binding ligand, indicating that indeed the structures have a role in regulating virus expression.³⁵ The group also demonstrated that different G-quadruplexes are formed in the U3 region of DNA and that some of them involve G runs of Sp1 sites together with a G run in the NF-κB site, which was not a subject of our studies. However, these and our results together show that the entire G-rich transcription factor binding region in the HIV-1 proximal promoter adopts different forms of G-quadruplex structure.

The variation of genomic sequence among various viral species is higher for the U3 region having Sp1 binding sites than for the protein coding regions. However, according to our computational analysis, the sequence variants all retain the ability of the Sp1 binding sites to form G-quadruplex structure. This observation additionally supports the conclusion that G-quadruplex structure is an important element of the HIV-1 promoter region that is maintained by a strong evolutionary pressure.

Because the G-quadruplex structure forms in the Sp1 transcription factor binding region, it is likely that it participates in the regulation of promoter function. Recent studies of the c-kit promoter showed that Sp1 protein can bind to its sequence folded into a G-quadruplex.⁴² The sequence and G run distribution in HIV-1 Sp1 sites differ from those of the Sp1 binding region in the c-kit promoter. However, our results also show Sp1-G-quadruplex interactions, indicating that Sp1 could regulate HIV-1 promoter activity through binding dsDNA or G-quadruplex structure and that these binding characteristics might be linked with different functions of Sp1. Moreover, we found that the G-rich HIV-1 sequence with three Sp1 binding sites can adopt alternative configuration G-quadruplexes with different interacting sets of G runs and that antibodies specific to the c-kit2 G-quadruplex recognize two of these forms. Such diversity suggests that the G-rich Sp1 binding region uses its ability to adopt several non-B DNA configurations as a complex switching mechanism involving subtle Sp1 interaction differences to fine-tune transcriptional output.

Why might such complex alternative protein-DNA interactions be necessary? Sp1 is a strong activator of HIV-1 expression, but Sp1 activity is also associated with maintaining virus latency.^{60,61} Importantly, the Sp1 binding elements are within a region of the HIV-1 promoter that remains nucleosome-free despite hypermethylation of two CpG islands that flank the HIV-1 transcription start site.^{62,63} As a result, the region remains accessible to transcription regulators, any of which might act through G-quadruplex structure. G-quadruplex binding agents and proteins were shown to suppress cellular promoter activity by stabilizing a G-quadruplex formed in the Sp1 binding region, but unwinding of the structure reactivated gene expression.^{1,12–17} Interestingly, stimulation of cellular promoter activity through G-quadruplex structure was recently also demonstrated.^{18,19} However, recently published results by Perrone and co-workers indicate that G-quadruplexes formed in HIV-1 promoter have a rather inhibitory effect on viral transcription.³⁵

Formation of G-quadruplexes in the HIV-1 RNA genome has previously been demonstrated in *gag* near the RNA 5' end and recently in the central region of the genome near the cPPT.^{30–32,34} Significantly, both sequences can form dimer complexes through intermolecular G-quadruplex structure, indicating that G-quartets might be formed that link the two RNA genomes. Our results show that short RNA molecules with HIV-1 sequences containing the Sp1 binding sites can also dimerize and that their interactions have characteristics of the intermolecular G-quadruplex. This suggests that the U3 region is an additional point of contact in a multiply linked genome dimer and together with G-rich sequences in *gag* and cPPT helps to maintain interaction of the HIV-1 genomes along their whole sequence. Apart from dimerization through the DIS region, additional interactions through intermolecular G-quadruplexes would explain the ability of viral RNA genomes to maintain a dimeric configuration when binding-disruptive

mutations are introduced in DIS.^{64,65} Moreover, additional contacts would also be necessary to keep the two 9kb-long RNA sequences in proximity for efficient reverse transcription and the observed widely distributed hot spots for recombination. The ability of the G-rich sequence in U3 to form a dimer also means that a contact point at the 3' end provides equal accessibility of either of the 3' ends of the copackaged RNA genomes for minus strand DNA transfer. Indeed, studies showed that transfers of the minus strand DNA primer occur with the same frequency to each 3' genomic end.⁵⁴

Consistent with these expectations, our results demonstrate that the G-quadruplex formed by the Sp1 binding sites is likely a structural element contributing to the increased recombination rate in U3. Detailed measurements of strand transfers at the ends of HIV-1 showed that 11 of 86 analyzed clones (12%) underwent homologous recombination in the U3 region, and all crossovers occurred upstream of the G-rich region of the Sp1 binding sites.⁵⁴ Other results showed that recombination in U3 dropped significantly in tests conducted with a template construct missing a 150-nt long sequence containing the putative dimerization site at the 3' end, suggesting the existence of sequence elements in this region that are crucial for efficient homologous recombination in this region.⁵⁵

Our analysis of minus strand transfer using a reconstituted system shows that the transfer is cation-dependent and that transfer efficiency decreases in the presence of lithium ions, which destabilize G-quadruplex structure, although the effect is less dramatic than observed for G-rich regions of *gag* and cPPT. This suggests that G-quadruplex(es) formed in the region of Sp1 binding sites are one of the factors causing increased recombination in U3. The correlation between potential G-quadruplex formation and hot spots for recombination was also found in the *gag* and cPPT regions and is in an agreement with previous analyses *in vitro* showing that G-quadruplexes facilitate RT template switching during reverse transcription.^{33,34,66}

In summary, our current studies, combined with previous work, show that G-rich sequences in the HIV-1 genome are capable of forming G-quadruplexes in both RNA and DNA forms of the genome. The distribution of recombination hot spots correlates with the sites where G-quadruplexes are formed, and reconstituted systems confirm that these structures facilitate strand transfers. The concept of G-quadruplexes regulating the activity of the HIV-1 promoter is new, and determining how and when G-quadruplexes regulate HIV-1 transcription would enhance our understanding of HIV-1 latency and reactivation, which might help to identify a new molecular target for therapeutic reactivation of virus replication.

ASSOCIATED CONTENT

Supporting Information

DNA and RNA oligonucleotides. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*Department of Microbiology and Immunology, School of Medicine and Dentistry, University of Rochester, 601 Elmwood Ave., Rochester, NY 14642, USA. Phone: +1 585 275-2764. Fax: +1 585 275-6007. E-mail: robert_bambara@urmc.rochester.edu.

Funding

This work was supported by the National Institutes of Health (NIH) [GM049573 to R.A.B. and GM097334 to G.S.]. We received additional support from the University of Rochester Developmental Center for AIDS Research grant P30 AI078498 (National Institutes of Health/National Institute of Allergy and Infectious Diseases) and the University of Rochester School of Medicine and Dentistry. Funding for open access charge: National Institutes of Health.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We are grateful to Dr. Robert J. Gorelick, of the National Cancer Institute, for the NC used in these studies. We thank Dr. Shankar Balasubramanian for sharing with us the pIT2 plasmid construct with hf2 sequence.

ABBREVIATIONS

DIS, dimer initiation site; cPPT, central polypurine tract; NC, nucleocapsid protein; TP, transfer product; DE, donor extension product; nt, nucleotide(s)

REFERENCES

- (1) 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.
- (2) Phan, A. T., Modi, Y. S., and Patel, D. J. (2004) Propeller-type parallel-stranded G-quadruplexes in the human c-myc promoter. *J. Am. Chem. Soc.* 126, 8710–8716.
- (3) Rankin, S., Reszka, A. P., Huppert, J., Zloh, M., Parkinson, G. N., Todd, A. K., Ladame, S., Balasubramanian, S., and Neidle, S. (2005) Putative DNA quadruplex formation within the human c-kit oncogene. *J. Am. Chem. Soc.* 127, 10584–10589.
- (4) Sun, D., Guo, K., Rusche, J. J., and Hurley, L. H. (2005) Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human VEGF gene by the presence of potassium and G-quadruplex-interacting agents. *Nucleic Acids Res.* 33, 6070–6080.
- (5) De Armond, R., Wood, S., Sun, D., Hurley, L. H., and Ebbinghaus, S. W. (2005) Evidence for the presence of a guanine quadruplex forming region within a polypurine tract of the hypoxia inducible factor 1alpha promoter. *Biochemistry* 44, 16341–16350.
- (6) Ambrus, A., Chen, D., Dai, J., Jones, R. A., and Yang, D. (2005) Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization. *Biochemistry* 44, 2048–2058.
- (7) Phan, A. T., Kuryavyi, V., Gaw, H. Y., and Patel, D. J. (2005) Small-molecule interaction with a five-guanine-tract G-quadruplex structure from the human MYC promoter. *Nat. Chem. Biol.* 1, 167–173.
- (8) Dai, J., Dexheimer, T. S., Chen, D., Carver, M., Ambrus, A., Jones, R. A., and Yang, D. (2006) An intramolecular G-quadruplex structure with mixed parallel/antiparallel G-strands formed in the human BCL-2 promoter region in solution. *J. Am. Chem. Soc.* 128, 1096–1098.
- (9) Cogoi, S., and Xodo, L. E. (2006) G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Res.* 34, 2536–2549.
- (10) Fernando, H., Reszka, A. P., Huppert, J., Ladame, S., Rankin, S., Venkitaraman, A. R., Neidle, S., and Balasubramanian, S. (2006) A conserved quadruplex motif located in a transcription activation site of the human c-kit oncogene. *Biochemistry* 45, 7854–7860.
- (11) Xu, Y., and Sugiyama, H. (2006) Formation of the G-quadruplex and i-motif structures in retinoblastoma susceptibility genes (Rb). *Nucleic Acids Res.* 34, 949–954.

- (12) Seenisamy, J., Rezler, E. M., Powell, T. J., Tye, D., Gokhale, V., Joshi, C. S., Siddiqui-Jain, A., and Hurley, L. H. (2004) The dynamic character of the G-quadruplex element in the c-MYC promoter and modification by TMPyP4. *J. Am. Chem. Soc.* 126, 8702–8709.
- (13) Sun, D., Liu, W. J., Guo, K., Rusche, J. J., Ebbinghaus, S., Gokhale, V., and Hurley, L. H. (2008) The proximal promoter region of the human vascular endothelial growth factor gene has a G-quadruplex structure that can be targeted by G-quadruplex-interacting agents. *Mol. Cancer Ther.* 7, 880–889.
- (14) Gonzalez, V., Guo, K., Hurley, L., and Sun, D. (2009) Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein. *J. Biol. Chem.* 284, 23622–23635.
- (15) Thakur, R. K., Kumar, P., Halder, K., Verma, A., Kar, A., Parent, J. L., Basundra, R., Kumar, A., and Chowdhury, S. (2009) Metastases suppressor NM23-H2 interaction with G-quadruplex DNA within c-MYC promoter nuclease hypersensitive element induces c-MYC expression. *Nucleic Acids Res.* 37, 172–183.
- (16) Gonzalez, V., and Hurley, L. H. (2010) The C-terminus of nucleolin promotes the formation of the c-MYC G-quadruplex and inhibits c-MYC promoter activity. *Biochemistry* 49, 9706–9714.
- (17) Brown, R. V., Danford, F. L., Gokhale, V., Hurley, L. H., and Brooks, T. A. (2011) Demonstration that drug-targeted down-regulation of MYC in non-Hodgkins lymphoma is directly mediated through the promoter G-quadruplex. *J. Biol. Chem.* 286, 41018–41027.
- (18) Baral, A., Kumar, P., Halder, R., Mani, P., Yadav, V. K., Singh, A., Das, S. K., and Chowdhury, S. (2012) Quadruplex-single nucleotide polymorphisms (Quad-SNP) influence gene expression difference among individuals. *Nucleic Acids Res.* 40, 3800–3811.
- (19) Wei, P. C., Wang, Z. F., Lo, W. T., Su, M. I., Shew, J. Y., Chang, T. C., and Lee, W. H. (2013) A cis-element with mixed G-quadruplex structure of NPGPx promoter is essential for nucleolin-mediated transactivation on non-targeting siRNA stress. *Nucleic Acids Res.* 41, 1533–1543.
- (20) Phan, A. T., Kuryavyi, V., Burge, S., Neidle, S., and Patel, D. J. (2007) Structure of an unprecedented G-quadruplex scaffold in the human c-kit promoter. *J. Am. Chem. Soc.* 129, 4386–4392.
- (21) Mukundan, V. T., and Phan, A. T. (2013) Bulges in g-quadruplexes: broadening the definition of g-quadruplex-forming sequences. *J. Am. Chem. Soc.* 135, 5017–5028.
- (22) 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.
- (23) 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.
- (24) Beaudoin, J. D., and Perreault, J. P. (2010) 5'-UTR G-quadruplex structures acting as translational repressors. *Nucleic Acids Res.* 38, 7022–7036.
- (25) Beaudoin, J. D., and Perreault, J. P. (2013) Exploring mRNA 3'-UTR G-quadruplexes: evidence of roles in both alternative polyadenylation and mRNA shortening. *Nucleic Acids Res.* 41, 5898–5911.
- (26) 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.
- (27) Gomez, D., Lemarteleur, T., Lacroix, L., Mailliet, P., Mergny, J. L., and Riou, J. F. (2004) Telomerase downregulation induced by the G-quadruplex ligand 12459 in AS49 cells is mediated by hTERT RNA alternative splicing. *Nucleic Acids Res.* 32, 371–379.
- (28) 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.
- (29) 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.
- (30) Marquet, R., Baudin, F., Gabus, C., Darlix, J. L., Mougel, 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.
- (31) 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.
- (32) Marquet, R., Paillard, 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.
- (33) 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.
- (34) Piecka-Przybylska, D., Sharma, G., and Bambara, R. A. (2013) Mechanism of HIV-1 RNA dimerization in the central region of the genome, and significance for viral evolution. *J. Biol. Chem.* 288, 24140–24150.
- (35) Perrone, R., Nadai, M., Frasson, I., Poe, J. A., Butovskaya, E., Smithgall, T. E., Palumbo, M., Palu, G., and Richter, S. N. (2013) A dynamic G-quadruplex region regulates the HIV-1 long terminal repeat promoter. *J. Med. Chem.* 56, 6521–6530.
- (36) Pandey, V. N., Kaushik, N., Rege, N., Sarafianos, 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.
- (37) Fernando, H., Rodriguez, R., and Balasubramanian, S. (2008) Selective recognition of a DNA G-quadruplex by an engineered antibody. *Biochemistry* 47, 9365–9371.
- (38) de Wildt, R. M., Mundy, C. R., Gorick, B. D., and Tomlinson, I. M. (2000) Antibody arrays for high-throughput screening of antibody-antigen interactions. *Nat. Biotechnol.* 18, 989–994.
- (39) 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, W676–682.
- (40) 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.
- (41) Vorlickova, M., Kejnovska, I., Sagi, J., Renciuk, D., Bednarova, K., Motlova, J., and Kypr, J. (2012) Circular dichroism and guanine quadruplexes. *Methods* 57, 64–75.
- (42) Raiber, E. A., Kranaster, R., Lam, E., Nikan, M., and Balasubramanian, S. (2012) A non-canonical DNA structure is a binding motif for the transcription factor SP1 in vitro. *Nucleic Acids Res.* 40, 1499–1508.
- (43) Fialova, M., Kypr, J., and Vorlickova, M. (2006) The thrombin binding aptamer GGTTGGTGTGGTTGG forms a bimolecular guanine tetraplex. *Biochem. Biophys. Res. Commun.* 344, 50–54.
- (44) Zaug, A. J., Podell, E. R., and Cech, T. R. (2005) Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro. *Proc. Natl. Acad. Sci. U.S.A.* 102, 10864–10869.
- (45) Arora, A., and Suess, B. (2011) An RNA G-quadruplex in the 3' UTR of the proto-oncogene PIM1 represses translation. *RNA Biol.* 8, 802–805.
- (46) Lam, E. Y., Beraldi, D., Tannahill, D., and Balasubramanian, S. (2013) G-quadruplex structures are stable and detectable in human genomic DNA. *Nat. Commun.* 4, 1796.
- (47) Smargiasso, N., Gabelica, V., Damblon, C., Rosu, F., De Pauw, E., Teulade-Fichou, M. P., Rowe, J. A., and Claessens, A. (2009) Putative DNA G-quadruplex formation within the promoters of Plasmodium falciparum var genes. *BMC Genomics* 10, 362.
- (48) Ambrus, A., Chen, D., Dai, J., Bialis, T., Jones, R. A., and Yang, D. (2006) Human telomeric sequence forms a hybrid-type intra-

- molecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Res.* 34, 2723–2735.
- (49) Qin, Y., Rezler, E. M., Gokhale, V., Sun, D., and Hurley, L. H. (2007) Characterization of the G-quadruplexes in the duplex nuclease hypersensitive element of the PDGF-A promoter and modulation of PDGF-A promoter activity by TMPyP4. *Nucleic Acids Res.* 35, 7698–7713.
- (50) Isalan, M., Patel, S. D., Balasubramanian, S., and Choo, Y. (2001) Selection of zinc fingers that bind single-stranded telomeric DNA in the G-quadruplex conformation. *Biochemistry* 40, 830–836.
- (51) 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.
- (52) Dykes, C., Balakrishnan, M., Planell, 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.
- (53) Fan, J., Negroni, M., and Robertson, D. L. (2007) The distribution of HIV-1 recombination breakpoints. *Infect. Genet. Evol.* 7, 717–723.
- (54) Yu, H., Jetzt, A. E., Ron, Y., Preston, B. D., and Dougherty, J. P. (1998) The nature of human immunodeficiency virus type 1 strand transfers. *J. Biol. Chem.* 273, 28384–28391.
- (55) Moumen, A., Polomack, L., Roques, B., Buc, H., and Negroni, M. (2001) The HIV-1 repeated sequence R as a robust hot-spot for copy-choice recombination. *Nucleic Acids Res.* 29, 3814–3821.
- (56) 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.
- (57) 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.
- (58) Eddy, J., and Maizels, N. (2008) Conserved elements with potential to form polymorphic G-quadruplex structures in the first intron of human genes. *Nucleic Acids Res.* 36, 1321–1333.
- (59) Todd, A. K., and Neidle, S. (2008) The relationship of potential G-quadruplex sequences in cis-upstream regions of the human genome to SP1-binding elements. *Nucleic Acids Res.* 36, 2700–2704.
- (60) Jiang, G., Espeseth, A., Hazuda, D. J., and Margolis, D. M. (2007) c-Myc and Sp1 contribute to proviral latency by recruiting histone deacetylase 1 to the human immunodeficiency virus type 1 promoter. *J. Virol.* 81, 10914–10923.
- (61) Marban, C., Suzanne, S., Dequiedt, F., de Walque, S., Redel, L., Van Lint, C., Aunis, D., and Rohr, O. (2007) Recruitment of chromatin-modifying enzymes by CTIP2 promotes HIV-1 transcriptional silencing. *EMBO J.* 26, 412–423.
- (62) Verdin, E., Paras, P., Jr., and Van Lint, C. (1993) Chromatin disruption in the promoter of human immunodeficiency virus type 1 during transcriptional activation. *EMBO J.* 12, 3249–3259.
- (63) Kauder, S. E., Bosque, A., Lindqvist, A., Planell, V., and Verdin, E. (2009) Epigenetic regulation of HIV-1 latency by cytosine methylation. *PLoS Pathog.* 5, e1000495.
- (64) 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.
- (65) Shen, N., Jette, 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.
- (66) 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.