12328–12343 *Nucleic Acids Research, 2022, Vol. 50, No. 21* *Published online 1 December 2022*
https://doi.org/10.1093/nar/gkac1030

Deciphering RNA G-quadruplex function during the early steps of HIV-1 infection

Samir Amrane[1,2,*], Chloé Jaubert[1,2], Amina Bedrat[1,2], Tiffany Rundstadler[3,4], Patricia Recordon-Pinson[1,5], Cindy Aknin[1,5], Aurore Guédin[1,2], Aurore De Rache[1,2], Laura Bartolucci[1,2], Ibra Diene[1,2], Frédéric Lemoine[6,7], Olivier Gascuel[6,8], Geneviève Pratviel[3,4], Jean-Louis Mergny[1,2,9] and Marie-Line Andreola[1,5]

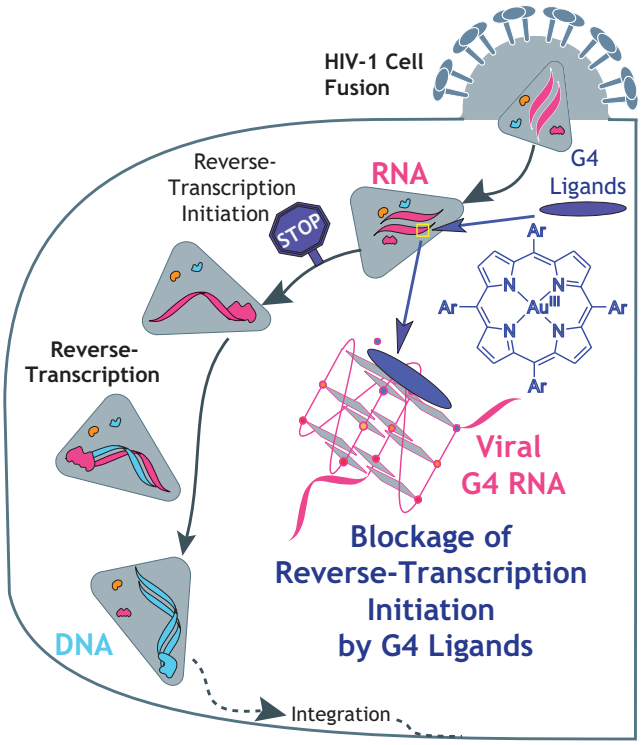
[1]Université de Bordeaux, Bordeaux, France, [2]ARNA Laboratory, INSERM U1212, CNRS UMR 5320, IECB, Bordeaux, France, [3]Université de Toulouse, UPS, INPT, Toulouse, France, [4]Laboratoire de Chimie de Coordination, CNRS UPR 8241, Toulouse, France, [5]MFP laboratory, UMR5234, CNRS, Bordeaux, France, [6]Institut Pasteur, Université de Paris, Unité de Bioinformatique Évolutive, F-75015 Paris, France, [7]Institut Pasteur, Université de Paris, Hub de bioinformatique et biostatistiques, F-75015 Paris, France, [8]Institut de Systématique, Évolution, Biodiversité (ISYEB, UMR 7205 - CNRS, Muséum National d'Histoire Naturelle, SU, EPHE UA), F-75005 Paris, France and [9]Laboratoire d'Optique & Biosciences, Ecole Polytechnique, CNRS, Inserm, Institut Polytechnique de Paris, Palaiseau, France

Received February 09, 2022; Revised September 27, 2022; Editorial Decision October 16, 2022; Accepted October 29, 2022

ABSTRACT

G-quadruplexes (G4s) are four-stranded nucleic acid structures formed by the stacking of G-tetrads. Here we investigated their formation and function during HIV-1 infection. Using bioinformatics and biophysics analyses we first searched for evolutionary conserved G4-forming sequences in HIV-1 genome. We identified 10 G4s with conservation rates higher than those of HIV-1 regulatory sequences such as RRE and TAR. We then used porphyrin-based G4-binders to probe the formation of the G4s during infection of human cells by native HIV-1. The G4-binders efficiently inhibited HIV-1 infectivity, which is attributed to the formation of G4 structures during HIV-1 replication. Using a qRT-PCR approach, we showed that the formation of viral G4s occurs during the first 2 h post-infection and their stabilization by the G4-binders prevents initiation of reverse transcription. We also used a G4-RNA pull-down approach, based on a G4-specific biotinylated probe, to allow the direct detection and identification of viral G4-RNA in infected cells. Most of the detected G4-RNAs contain crucial regulatory elements such as the PPT and cPPT sequences as well as the U3 region. Hence, these G4s would function in the early stages of infection when the viral RNA genome is being processed for the reverse transcription step.

GRAPHICAL ABSTRACT



*To whom correspondence should be addressed. Tel : +33 5 4000 2224; Email: samir.amrane@u-bordeaux.fr
Present address: Aurore De Rache, Department of Chemistry, U. Namur, 61 rue de Bruxelles, B5000 Namur, Belgium.

INTRODUCTION

G-quadruplexes are polymorphic and stable nucleic acid structures

G-quadruplexes (or "G4s") are non-canonical nucleic acid structures formed by guanine-rich DNA or RNA sequences (Figure 1A). G4s are based on the formation of two or more G-quartets; each quartet resulting from the square planar arrangement of four guanines connected by eight hydrogen bonds (Figure 1B). Potassium, and to a lesser extent sodium cation, both present in the cellular environment, stabilize these G-quartets through specific electrostatic interactions with the oxygens of the carbonyl groups of the four guanines. The stacking of two or more G-quartets defines the core of the G4 structure. The guanines of the core can adopt either *syn* or *anti* glycosidic bond angles (Figure 1C). This crucial feature enables a DNA G4 structure to explore a vast space of conformations (1). Indeed, in the case of intramolecular G4s, the four strands can be interconnected by three different types of loops (lateral, diagonal or propeller) of variable lengths and sequences (Figure 1D). Consequently, each of the four strands can either be parallel or antiparallel to its adjacent strands. Depending on this orientation, the four grooves delimited by the strands can be narrow, medium, or wide (Figure 1B). All these structural elements depend on the nucleotide sequence of the G4-forming sequences (G4FSs) and govern the overall shape of the G4 (Figure 1E). The versatility of these parameters is at the root of the tremendous plasticity of these nucleic acid structures. G4 structures are generally stable under physiological conditions, with thermal denaturation temperatures often above 50°C, and sometimes above 90°C (2).

G4s have biological functions in human cells and microbes

G4s were long regarded as mere curiosities; however, as reviewed recently, these structures are now known to be involved in key biological processes (3) such as telomere dynamics (4–6), replication (7), transcription (8) and translation (9,10). In order to infer the biological role of G4s, we and other groups have performed genome-wide bioinformatics studies to identify putative G4FSs in the human genome (11,12). Using the G4Hunter algorithm (11) we identified up to 1.5 million putative G4FSs in the human genome. We also found that roughly half of human promoters have at least one putative G4FS. Several *ex vivo* studies using G4-specific antibodies have confirmed that these structures exist in genomic DNA and cellular RNA (13,14–16). Furthermore, G4s are also conserved in bacteria, fungi, and viruses (17–19) and have been linked to important processes such as antigenic variation, recombination, and latency.

As functional biomolecules, G4s are certainly involved in complex networks of interactions with multiple proteins. Indeed, an increasing number of cellular proteins have been demonstrated to interact with G4s (reviewed here (20)). Depending on the mode of interaction or the function of the G4, the protein will either stabilize the G4 by locking its conformation or it will destabilize it by unfolding it to facilitate a given biological process. For example, the RPA (21) and POT1 (22) proteins present in the telomeric nucleoprotein complex as well as the BLM (23) and PIF1 (24) helicases will tend to unfold the G4 structures present at the telomeres. On the contrary, the proteins SP1 (25), PARP-1 (26) and Nucleolin (27) will stabilize the G4 localized at the promoters and may decrease the expression of the gene as reported in the case of Nucleolin. Some proteins are also known to preferentially bind to G4 RNA. This is the case of the DHX9 (28) protein which destabilizes G4s and the FMRP (29) protein which on the contrary stabilizes the G4 structure. Despite a certain scarcity of high-resolution structures describing G4–protein complexes, several binding modes have already been identified: Stacking at the top of the G-tetrad, groove recognition and loop binding. The stacking interaction is exemplified by the DHX36 helicase (30). The C-terminal domain consists of a G4-binding helix called the DHX36-specific motif (DSM), which sits atop the 5′ end of the G4 with a preference for parallel G4 DNA. The loop recognition mode has been proposed for a synthetic zinc finger protein (ZFP) in which one zinc finger is able to recognize two residues of the G4 loop (31). The groove-binding mode was observed with the equivalent of the human POT1–TPP1 complex that is present in the telomere nucleoprotein found in Oxytricha nova (32). Interestingly, the X-ray structure shows at the same time an interaction with an unfolded G4 and the phosphate backbone of an intact G4. Remarkably, a very recent Cryo-EM structure demonstrated a dual mode of binding by recognizing both grooves and loops. In this case, the G4 is trapped by the POLA1 domain of Tetrahymena telomerase (33). This G4 formed by four telomeric repeats of Tetrahymena binds to a positively charged DNA binding surface of POLA1 mainly through groove recognition.

G4s are promising therapeutic targets

As G4 structures are very distinct from canonical duplexes or single-strands, G4s represent a promising class of pharmacological targets (34), especially against viruses (35). The structural diversity of G4s suggests that a relatively high degree of selectivity can be achieved with small molecules. Rational design of ligands and *in silico* screening have been reported (36, 37), and several thousands G4 ligands are currently described in the G4 ligands database (38) (http://www.g4ldb.com). At least two G4 ligands have been tested in clinical trials as cancer therapies and further applications have been proposed in virology (39). Notably, G4 ligands of various chemical families (bisquinoliniums, naphthalene diimides, acridiniums) are able to inhibit replication of DNA viruses such as herpesviruses and hepatitis B virus as well as RNA viruses such as HIV-1 and hepatitis C virus (40,41). Although this suggests that G4s have important functions in viruses, the G4-based mechanisms of action are not well understood.

G4s in HIV-1

HIV-1 is a retrovirus that infects cells that express CD4 and one of the chemokine receptors CCR5 or CXCR4 on their surface. The HIV-1 genome is composed of two copies of a single-stranded 9.2 kb-long, plus-strand RNA encoding nine viral proteins (Figure 2A). During infection, the viral

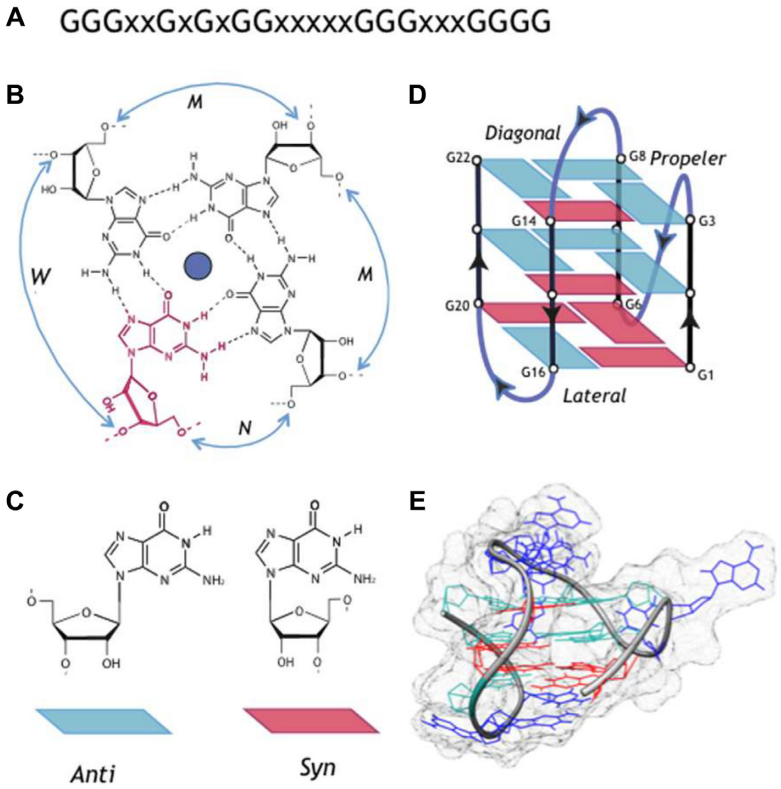


Figure 1. G4 nucleic acid structures. (**A**) Example of a G-quadruplex motif: Repeats of two to four guanines are separated by inter-block sequences composed of 1 to 15 nt (X, any nucleotide). The blocks of more than three Gs can be interrupted by one or two non-G residues. (**B**) Structure of a G-quartet. The guanine in red adopt a *syn* conformation. Those in black are in the *anti* conformation. Different groove sizes are depicted (N, narrow; M, medium; W, wide). The tetrad is stabilized by a central K^+ or Na^+ cation (blue). (**C**) *Syn* and *anti* conformations adopted by Guanines. (**D**) Schematic representation of a hybrid type G4 (PDB = 2LOD) composed of three tetrads with three different loop types: propeller, diagonal and lateral. *Syn* and *anti* conformations appear in red and cyan. (**E**) Surface view of the G4 (PDB = 2LOD). *Syn* and *anti* conformations appear in red and cyan.

RNA is reverse transcribed into a 9.6-kb double-stranded DNA flanked by two regulatory regions called the long terminal repeats (LTRs). The viral DNA is then integrated into the genome of the infected cell. The host cell machinery then transcribes the viral genes and produces viral proteins, leading to the assembly of new viruses. We (42) and others (43,44) have described a G4FS in the U3 region of the LTRs. This region is found at both ends of the proviral DNA genome, whereas it is only present at the 3′ extremity of the RNA genome (Figure 2A). In the 5′-LTR, the U3-G4FS overlaps with the binding sites of SP1 transcription factors and serves as a promoter for the entire viral genome. We have proposed that the formation of a DNA G4 structure in this region may be critical for HIV-1 gene expression (42). Other studies have detected G4FSs in the genes encoding GAG and POL (45) and proposed a role of these structures in the recombination and/or dimerization (46) of the genomes.

In this study, we first screened 2177 complete HIV-1 genomes and identified conserved G4FSs at both the DNA or RNA levels. We detected ten evolutionarily conserved G4FSs. In order to probe the formation of G4 during HIV-1 infection, we treated human cells in culture with a family of porphyrin-based G4 binders. We showed that these compounds inhibited viral infectivity by targeting the viral RNA without affecting enzymatic activity. These molecules acted after the viral entry and prevented the initiation of reverse transcription in a G4-mediated mechanism of action. We finally implemented an RNA pull-down approach based on a G4-specific biotinylated probe and we demonstrated the formation in infected cells of viral RNA G4s. Some of these RNA G4s contain crucial regulatory elements like the PPT and cPPT sequences as well as the U3 region. These data point toward a functional role of G4 structures during the two first hours of HIV-1 replication and pave the way for the development of new anti-HIV therapeutic strategies constructed around these viral G4s.

MATERIALS AND METHODS

Bioinformatics analysis

G4Hunter was used to detect putative G4-forming sequences in an alignment of 2177 genomes from HIV-1 database (www.hiv.lanl.gov). A phylogenetic tree was inferred using FastME v2.1.5 (options -dF81 -T10) from 2177 HIV-1 sequences. Branch lengths and site-specific evolutionary rates (posterior mean rate) were then estimated using PhyML v3.3.20180621 (options -o lr -u <fastme> -d nt -f m -b 0 -m HKY85 -c 6 -t e -v e -a e –print_site_lnl). Evolutionary rates of sites within motifs were compared to

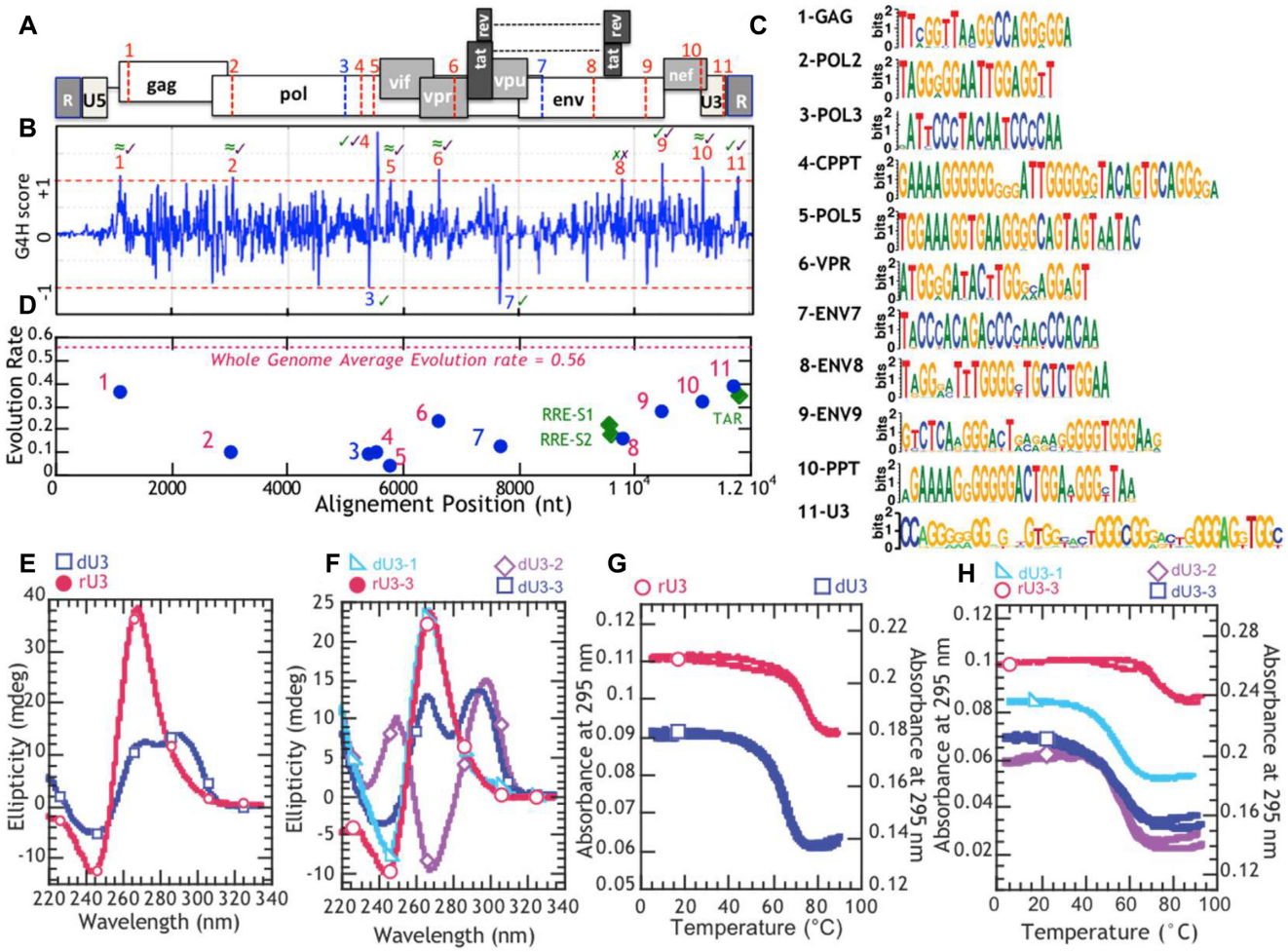


Figure 2. *In silico* detection of the putative G4FSs in HIV-1 and biophysical validation. (**A**) Scheme of HIV-1 genome composed of nine genes. The positions of the putative G4FSs are indicated by red dashed lines for the sequences located on the RNA or DNA plus strands and blue dashed lines indicate those located on the minus strand of the DNA. (**B**) Putative G4FSs detected using G4Hunter using a window of 25 nt. The average of the 2177 scores for each window is then assigned to the first nucleotide of the 25-nt sequence. Regions with scores greater than or equal to 1 (in absolute value) can form a G4, whereas those with scores below 1 (in absolute value) are unlikely to form a G4. The putative G4FSs are numbered. Above the numbers, symbols indicate whether the region forms a stable DNA (green) or RNA (purple) G4 *in vitro* (marked with ✓), an unstable G4 (marked with ≈), or does not form a G4 (marked with ✗). The final sequence alignment has a length of about 12 000 nt exceeding the size of 9.2 kb known for the HIV-1 genome due to the insertion of "gaps" necessary to optimize the sequence alignment. (**C**) Logo representations of the putative G4FSs generated using Weblogo (www.weblogo.com). The heights of the residues are proportional to frequencies in the alignment. (**D**) Average evolutionary rates of each putative G4FS and RRE-S1, RRE-S2, TAR regulatory elements. The average pairwise p-distances are shown in Figure S2. (E–H) Biophysical analyses of the U3-G4FS. The oligonucleotide was dissolved at a concentration of 5–10 μM in a buffer containing 70 mM KCl and 10 mM potassium phosphate at pH 7. (**E**) Circular dichroism profiles of full-length dU3 (blue) and full-length rU3 (pink). (**F**) Circular dichroism profiles of truncated dU3 and truncated rU3. (**G**) UV melting profiles of full-length dU3 and full-length rU3. (**H**) UV melting profiles of truncated dU3 and truncated rU3.

evolutionary rates of sites outside motifs using R v3.4.3. Average pairwise p-distances of sequences within motifs were computed using goalign (47) (https://github.com/evolbioinfo/goalign) v0.3.1 and compared to average pairwise p-distances of 1000 random HIV alignment windows of 30 nt.

Biophysical analysis

Circular dichroism studies. CD spectra were recorded on a Jasco J-1500 instrument equipped with a Peltier temperature control accessory. Each spectrum was obtained by averaging five scans recorded at a speed of 100 nm/min at 20°C.

Ultraviolet melting experiments. UV-absorbance measurements were performed on a Uvikon XS (Secomam) spectrophotometer thermostated with an external circulating water bath. The absorbance was monitored at 295 nm, and the heating/cooling rate was 0.2°C/min. For each sample, the TDS was obtained by subtracting the absorbance spectrum at 4°C from the one at 95°C as previously described (48).

FRET melting assay. FRET melting experiments were run on a Stratagene Mx3005P real-time PCR equipment in 96-well plates on the DNA sequences listed in Supplementary Table S2. Experiments were performed in 10 mM lithium cacodylate buffer (pH 7.2) and 10 mM KCl and 90 mM LiCl

as previously described (49). The DNA concentration was 0.2 μM. See supplementary methods for a more detailed protocol.

Ligand-induced fluorescence quenching assay. The ligand-induced fluorescence quenching assay was conducted as previously described (50). A Stratagene Mx3005P instrument was used to carry out the assay in 96-well plates. The labeled oligonucleotide was used at 10 nM, and the buffer was 10 mM lithium cacodylate (pH 7.2), 10 mM KCl, 90 mM LiCl in a final volume of 50 μl. The ligand was added over a range concentration from 0 to 10 μM. See supplementary methods for a more detailed protocol.

Viral infectivity assays and RT-qPCR analysis

Virus infection. Infections were performed in a BSL-3 laboratory at the UB'*L3* platform in Bordeaux. Native HIV-LAI was obtained after co-culture between MT4 cells and H9 cells carrying a copy of integrated HIVLAI genome (51). 10^6 MT4 cells were co-cultured with 10^6 H9LAI for 48h then the supernatant was collected and clarified by filtration on 0.22 μM filters. Aliquots were then kept at −135°C. Viral RNAs in supernatant were quantified by qRT-PCR. HeLa P4 cells were infected at a MOI of 1 (determined according to Trono lab: estimation of 1 infectious particle for 1000 particles). HeLa P4 cells (52) used for infection experiments were maintained in DMEM medium (Invitrogen) supplemented with 10 % inactivated fetal calf serum (FCS), 1 mg/ml geneticin (G418, Gibco-BRL). This cell line encodes a Tat-inducible *β-galactosidase* (β-gal) gene driven by the HIV-1 LTR and linked to the expression of the viral Tat protein. HeLa P4 cells were seeded in a 96-well or 24-well plates (containing 10 000 or 100 000 cells per well respectively) 24 h before infection. Serial dilutions of drugs were added at the time of infection.

Infectivity analysis. After 24 h of culture, β-gal activity was quantified by adding 4-MUG mix (50 mM Tris–HCl, pH 8, 100 mM β-mercaptoethanol, 0.05% Triton X-100, 5 mM 4-methylumbelliferyl beta-D-galactoside) to cells. Fluorescence associated with the reaction product was monitored 24 h after adding the 4-MUG mix using a Cytofluor-II plate reader (Applied Biosystems) with excitation/emission filters at 360/460 nm. Cytotoxicity analysis was performed under similar conditions without virus and measured with the CellTiter 96® AQueous One Solution Cell Proliferation Assay System (Promega). Porphyrins are known to be fluorescent which could interfere with this test. However, the gold-porphyrins used here are not fluorescent. There is no interference with the reading of the 4-MUG test.

Nucleic acid extraction. Viral DNA extraction was performed 3, 6 and 24 h after infection in 96-well plates using the High Pure Viral Nucleic Acid Kit (Roche). RNA extraction was performed 2 h after infection in 24-well plates using TRI-Reagent (Molecularsup Research Center, Inc.) after a trypsin treatment to remove any bound viral particles at the surface of the cells. The purity and integrity of the RNAs were determined using capillary electrophoresis on a Bioanalyzer 2100 (Agilent). A 2 μM concentration of ligand allowed an almost complete inhibition of the replication.

RT-PCR and qPCR. Serial dilutions of pNL4-3 and GAPDH plasmids were used to derive a standard curve. For total DNA quantification, qPCR reactions were performed using the http://GoTaq/qPCR/Master/Mix (Promega) to amplify the *gag* viral gene (Gag_F: 5′-GCCTATTGCACCAGGCCAGAT-3′ and Gag R2: 5′-GTGAAGCTTGCTCGGCTCTTAGA-3′) and the *GAPDH* cellular gene (GAPDH+: 5′-GGAAGGTGAAGGTCGGAGTCAACGG-3′ and GAPDH-: 5′-TCCTGGAAGATGGTGATGGGATTTC-3′). For RNA quantification, RT-qPCR reactions were performed using the http://GoTaq/1-Step/RT-qPCR/System (Promega) to amplify the viral LTR (LTRNec001: 5′-GCCTCAATAAAGCTTGCCT-3′ and LTR131: 5′-GGCGCCACTGCTAGAGAT-3′) and the *GAPDH* cellular gene. Amplifications of strong-stop DNA, first jump DNA, and second jump DNA were performed using the following primers and probes: R (5′-GTTAGACCAGATCTGAGCCTGG-3′), OH65 (5′-GCTAGAGATTTTCCACACTG-3′), OH64 (5′-TAACTAGGGAACCCACTGC-3′), U3 (5′-AGCAGCTGCTTTTTGCCTGTACT-3′), U5 (5′-ACACAACAGACGGGCACACAC-3′), and postPBS (5′-CCTGCGTCGAGAGAGCTCCTCTGG-3′). For Taqman experiments, OHC64 in the R region and postPBS were used with 5′-FAM/3′-TAMRA probe (5′-ACACAACAGACGGGCACACACTA-3′). See supplementary methods for a more detailed protocol.

RNA G4 pull-down. We adapted the protocol developed by Yang *et al*. for transcriptomic profiling of RNA G-quadruplexes (G4-RNA) using the G4RP-seq (53). Cells infection was performed for 2 hours. After washing in DEPC-PBS, cells were fixed for 5 min at room temperature with 1% formaldehyde in fixing buffer (50 mM HEPES–KOH pH 7.5, 100 mM NaCl, 1 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0 in DEPC-H_2O), then quenching of the cross-linking was obtained by addition of 125 mM glycine. After centrifugation the pellet was washed in DEPC-PBS and suspended in 400 μl G4Rp buffer (150 mM KCl, 25 mM Tris pH 7.4, 5 mM EDTA, 0.5 mM DTT, 0.5% NP40, Roche RNase inhibitor). Sonication was performed with a Vibracell (Bioblock scientific). After centrifugation, 5% of the supernatant was kept as input. AuMA-biotin was added at a concentration of 40 μM and incubated overnight at 4°C on a rotator. Streptavidin beads (Promega) were incubated for 2 h at 4°C on a rotator. After three washes with G4RP buffer followed by two washes with DEPC-PBS, beads were suspended in 100 μl of DEPC-PBS. Samples were incubated for 1 h at 70°C to reverse the crosslink, before purification of RNAs with TRIZOL. Amplification of RNAs and qRT-PCR was performed as described in RT-PCR and qPCR method (see previous paragraph). The sizes of the RNA fragments produced during the sonication process were analysed by capillary electrophoresis (Figure S26). The majority of the fragments were between 100 and 200 nucleotides in size and a minority of RNA fragments was up to 700 nucleotides in size. The primers were designed to amplify products with sizes comprised between 118 and 219 nucleotides and displaying only one G4FS (Table S3). Sam-

ples were analyzed by migration on an agarose gel. See supplementary methods for a more detailed protocol.

RESULTS

In silico screening reveals evolutionarily conserved putative G4-forming sequences in the HIV-1 genome

RNA viruses have compact genomes and conserved secondary structures have important functions during the viral life cycle as illustrated by the TAR and RRE structural elements of HIV-1 (54). RNA viruses evolve extremely rapidly. Indeed, different subspecies often diverge by 10–30% at the sequence level (55). We reasoned that functional G4s should be conserved. In order to identify evolutionary conserved G4s in the HIV-1 genome, we used the G4Hunter algorithm (11) to search for putative G4FSs in an alignment of 2177 complete HIV-1 genomes from the M (subtypes A, B, C, D, F, G) and O groups. We detected 11 putative G4FSs: nine located on the plus-stranded RNA and DNA genomes and two on the minus-strand DNA (Figure 2A, B and Table 1). These putative G4FSs are well conserved (Figure 2C). In order to precisely assess the level of conservation in comparison with the rest of HIV-1 genome, we estimated the evolution rate at each position of the alignment using FastME (56) and PhyML (57) (Figure 2D). A phylogenetic tree was inferred from the 2177 HIV-1 genomes. As "speed of evolution" is allowed to vary across the sites of the alignments during tree inference, branch lengths and site-specific evolutionary rates (posterior mean rates) could be estimated. Moreover, the average pairwise p-distances (i.e. the number of mutations normalized by the sequence length) between the sequences of each putative G4FS were computed as well as the average pairwise p-distance between the sequences of 1000 randomly selected HIV-1 alignment windows of 30 nucleotides (nt) (Figures S1 and S2). Interestingly, the calculated mutation rate was significantly lower in putative G4FSs than in the rest of the alignment, and the average pairwise distances of the putative G4FS was lower than the average pairwise p-distances of random HIV-1 alignment windows; these data are indicative of a high conservation of these putative G4FSs. Moreover, the average evolution rate of the 11 putative G4FSs was significantly slower than the average evolution rate of the entire HIV-1 genome (Figure 2D). Most of the putative G4FSs had evolution rates similar or even lower than those of the TAR and RRE structural elements with evolution rates ranging from 0.16 to 0.39 for sequences 1, 6 and 8–11 and between 0.4 and 0.13 for sequences 2–5 and 7.

Ten putative G4-forming sequences adopt stable intramolecular and bimolecular DNA or RNA G4 structures *in vitro*

To verify whether candidate sequences form G4s, we synthesized DNA and RNA corresponding to the eleven and nine putative G4FSs, respectively (Table 1), and analyzed them using several biophysical approaches. The experimental data for sequence 11 (U3-pG4FS) are presented in Figure 2E–H and data for the other 10 putative G4FSs are summarized in Table 1 with experimental results shown in supplementary information (Figures S3–S16).

Conformational analysis by circular dichroism spectroscopy (Figure 2E, F and Table 1) revealed that the rU3 RNA sequence adopts a parallel fold, whereas its dU3 DNA counterpart adopts a hybrid-type G4. The dU3-1 and dU3-2 DNA sequences, which are oligonucleotides corresponding to shorter version of the U3 region, have anti-parallel and parallel CD signatures, respectively. These data are in agreement with previous NMR studies (42,44, 58). dU3-3 has a hybrid-type G4 profile similar to the full-length dU3 sequence. The RNA version rU3-3 also has characteristics of a parallel signature identical to rU3. Analyses of the molecularities of the oligonucleotides by size-exclusion chromatography (SEC) confirmed intramolecular folding for most sequences (Table 1, Figure S17–S20). Analyses of the stabilities using thermal denaturation experiments allowed the determination of the melting temperatures (T_ms) (Figure 2G, H) which ranged from 56°C to 63°C for DNA and 72°C to 75°C for RNA. Hence, the DNA versions of the U3-G4FS are able to form stable intramolecular G4 structures of parallel, antiparallel, or hybrid conformations while the RNAs are all parallel. The thermal differential spectra (TDS) and isothermal difference spectra (IDS) were also recorded and confirmed the formation of G4s for all U3 sequences except rU3-2 (Figures S3–S16 and Table 1).

Overall, among the nine putative G4FSs in the RNA genome, all but rENV8 were able to form stable intramolecular and/or bimolecular G4 structures with T_ms ranging from 45°C to 85°C. In the case of the 11 DNA sequences, only five (dU3, dENV7, dENV9, dCPPT, dPOL3) formed intramolecular and/or bimolecular G4 structures with T_ms higher than 35°C. The remaining DNA sequences either did not form G4 structures (dENV8) or formed unstable multimeric G4s (dGAG, dPOL2, dPOL5, dVPR, dPPT) with T_ms lower than 30°C, which are unlikely to be formed under physiological conditions (Figure 2B). These bioinformatic and biophysical analyses suggest that there are eight stable G4FSs in the HIV-1 RNA genome, whereas the viral DNA counterpart seems to be less prone to G4 formation, with only three G4FS located on the plus strand and two on the minus strand.

Porphyrin-based G4-ligands bind to the DNA and RNA U3-G4FS *in vitro* with high affinity and specificity

In an effort to determine whether viral G4FSs actually form G4 structures during HIV-1 infection, we investigated the effect of specific G4-stabilizing ligands on the virus life cycle. We previously designed metalated porphyrins with bulky and positively charged substituents at the four *meso*-positions (59–62) (Figure 3A). These compounds bind to the G4 structures by stacking on top of the G-quartets. In contrary to the first generation of porphyrins like TMPyP4 (called here H2T) these derivatives have higher affinity for G4 structures than for double-stranded DNA (63–65). Using the FRET melting assay (49), we studied the binding properties of six porphyrins toward three different DNA G4 conformations derived from the U3-G4FS (Table S2): dU3-1 (parallel), dU3-2 (antiparallel) and dU3-3 (hybrid). Data are shown for the fluorescently labeled dU3-1 in Figure 3B–D and for the other two conformations in Figure S21. The addition of 2.5 molar equivalents of TMPyP4 (H2T) or met-

Table 1. Analysis of the putative G4-forming sequences

N°	Name	Start[a]	Evolution rate[b]	Start[c]	Sequence 5′→ 3′	CD	IDS/TDS	T_m (°C)	SEC	G4FS?
1	dGAG	1126	0.36	388	GGTTAAGGCCAGGGGGA	Other	(+)	<15	mono/bi/h	YES
	rGAG				GGUUAAGGCCAGGGGGA	Parallel	(+)	38/48	mono/bi	YES
2	dPOL2	3043	0.1	1936	GGGGGAATTGGAGG	hybrid	(+)	26/44	mono/bi	YES
	rPOL2				GGGGGAAUUGGAGG	Parallel	(+)	56/61	mono/bi	YES
3*	dPOL3	5413	0.09	4205	GGGGATTGTAGGGGATC	hybrid	(+)	47/68	bi	YES
4	dCPPT-1	5540	0.1	4329	AAGGGGGGATTGGGGGGTACAGTGCAGGGGAA	Parallel	(+)	44/48	mono/bi/h	YES
	rCPPT-2				GGGGGGAUUGGGGGGUACAGUGCAGGGG	Parallel	(+)	70/78	mono/bi	YES
5	dPOL5	5770	0.04	4499	GGAAAGGTGAAGGGG	Hybrid	(+)	<30	mono/bi/h	YES
	rPOL5				GGAAAGGUGAAGGGG	Parallel	(+)	43/50	bi	YES
6	dVPR	6610	0.24	5245	GGGGATACTTGGGCAGGAGTGG	Anti-P	(+)	<30	mono/bi	YES
	rVPR				GGGGAUACUUGGGCAGGAGUGG	Parallel	(+)	43/46	mono	YES
7*	dENV7	7661	0.13	5982	TGTGGGTTGGGGTCTGTGGG	Parallel	(+)	32/55	bi	YES
8	dENV8	9784	0.16	7535	AGGAATTTGGGGCTGCTCTGGA	Other	(–)	NO	mono	NO
	rENV8				AGGAAUUUGGGGCUGCUCUGGA	Other	(–)	50	nd	NO
9	dENV9	10 471	0.18	8111	GGGACTACAGAGGGGGGTGGG	Parallel	(+)	59/81	mono/bi	YES
	rENV9				GGGACUACAGAGGGGGGUGGG	Parallel	(+)	69/75	bi	YES
10	dPPT-1	11 178	0.32	8611	AAGGGGGGACTGGAAGGGCT	hybrid	(+)	<35	mono/bi/h	YES
	rPPT-2				GGGGGGACUGGAAGGGCU	Parallel	(+)	54/62	bi	YES
11	dU3	11 823	0.38	9005	GGGAGGCGTGGCCTGGGCGGGACTGGGGAGTGG	Hybrid	(+)	63	mono	YES
	rU3				GGGAGGCGUGGCCUGGGCGGGACUGGGGAGUGG	Parallel	(+)	74	mono/bi	YES
	dU3-1				GGGCGGGACTGGGGAGTGGC	Parallel	(+)	56	mono/bi	YES
	rU3-1				GGGCGGGACUGGGGAGUGGC	Parallel	(+)	75	mono	YES
	dU3-2				TGGCCTGGGCGGGACTGGG	Anti-P	(+)	56	mono	YES
	rU3-2				UGGCCUGGGCGGGACUGGG	Other	(–)	nd	mono	NO
	dU3-3				AGGGAGGCGTGGCCTGGGCGGG	Hybrid	(+)	55	mono	YES
	rU3-3				AGGGAGGCGUGGCCUGGGCGGG	Parallel	(+)	72	mono/bi	YES

[a]Starting of the sequence in the alignment of 2177 genomes. [b]Evolution rate is calculated for the 2177 sequences [c]Starting of the sequence in NC-001802 genome isolate. *The G4-forming sequence is only present in the (–) strand at the DNA level. CD: topology deduced from circular dichroism spectra; Anti-P stands for antiparallel. IDS/TDS: isothermal or thermal differential spectra. (+) indicates that a negative peak at 295 nm is present, and (–) means there is no peak at 295 nm. T_m: melting temperature, the uncertainty of the value is ±1°C; two values indicate that there was hysteresis; T_m deduced from both cooling and heating experiments are provided. SEC: size exclusion chromatography; mono indicates monomolecular, bi indicates bimolecular, and h indicates higher order structures. G4FS?: YES indicates the formation of a G4, NO indicates that the sequence does not form a G4.

alated porphyrins AuT, AuMA, or AuPG considerably stabilized the G4 structure with increases in T_m of 25°C or more. The complexes formed by AuMA or AuPG were not melted completely at 90°C (Figure 3C). The non-metalated H2MA and H2LA derivatives only stabilized the G4 structure by 9.9°C and 5.6°C, respectively, likely as result of a lower affinity for G4 structures (Figure 3D). All six ligands stabilized the parallel conformation to a greater extent than the hybrid or anti-parallel conformations (Figure 3E, F). These results suggest that this family of ligands prefers to bind to the parallel G4 structures. Binding specificity for G4 was further assessed using a competition assay in which the thermal stabilization was measured in the presence of a duplex (ds26) competitor (Dx in Figure 3B–D). AuT and TMPyP4 (H2T) proved to be poorly selective because the stabilization almost completely disappeared in the presence of 50 molar equivalents of the non-fluorescent competitor duplex. In contrast, AuMA, H2MA, AuLA and H2LA appeared to be selective for G4 as stabilization was only modestly affected by the presence of the double-stranded competitor. In the case of the highly stable U3-G4 RNAs (T_m > 72°C), we could not use the FRET melting assay to study the binding of the ligands. Therefore we used a ligand-induced fluorescence quenching assay (50), to analyze the interaction of the porphyrins with the 5′ or 3′ fluorescently-labeled rU3-1-cy5 RNA (Figures 3G–I, S22, and Table S2). This approach allowed us to determine the dissociation constants (K_d) of each porphyrin derivative with either the 5′ or the 3′ quartet of the G4 structure. The fluorescence quenching of the Cy5 only occurs when the ligand is bound closely to the dye (50). For instance, in the case of the 3′ quartet, the dissociation constants ranged from 40 nM for AuMA to 1 μM for H2LA. Similar results were obtained for the second binding site at the 5′ extremity. This analysis allowed us to rank the ligands according to their binding affinities and specificities (Table 2).

These data led to several conclusions. First, the metalation of the porphyrin plays an important role on the affinity. The insertion of a gold(III) ion in AuMA increased the ΔT_m by almost 20°C relative to H2MA and improved the affinity by a factor of 5−10 depending on the 5′ or 3′ quartet. The incorporation of gold(III) adds an extra positive charge and induces electron deficiency in the aromatic cycle of the porphyrin (60). Second, the substituents in *meso*-position influence G4 binding affinity. Unlike the first-generation porphyrin TMPyP4 (H2T); LA, MA and PG compounds carry bulky substituents at the four *meso*-positions. In comparison with TMPyP4, H2LA and H2MA have lower affinities for the RNA and DNA G4. However, MA and LA porphyrins have much better binding selectivity for G4 with respect to the duplex DNA than does TMPyP4.

HIV-1 inhibition induced by the G4-binders correlates with their ability to stabilize and bind to the viral G4 structures

Next, we evaluated the ability of the six porphyrin derivatives to inhibit the replication of HIV-1 in human cells. Dose-response experiments were performed to establish IC_{50} values for all the compounds. We first used very short incubation times: porphyrins were incubated at the time

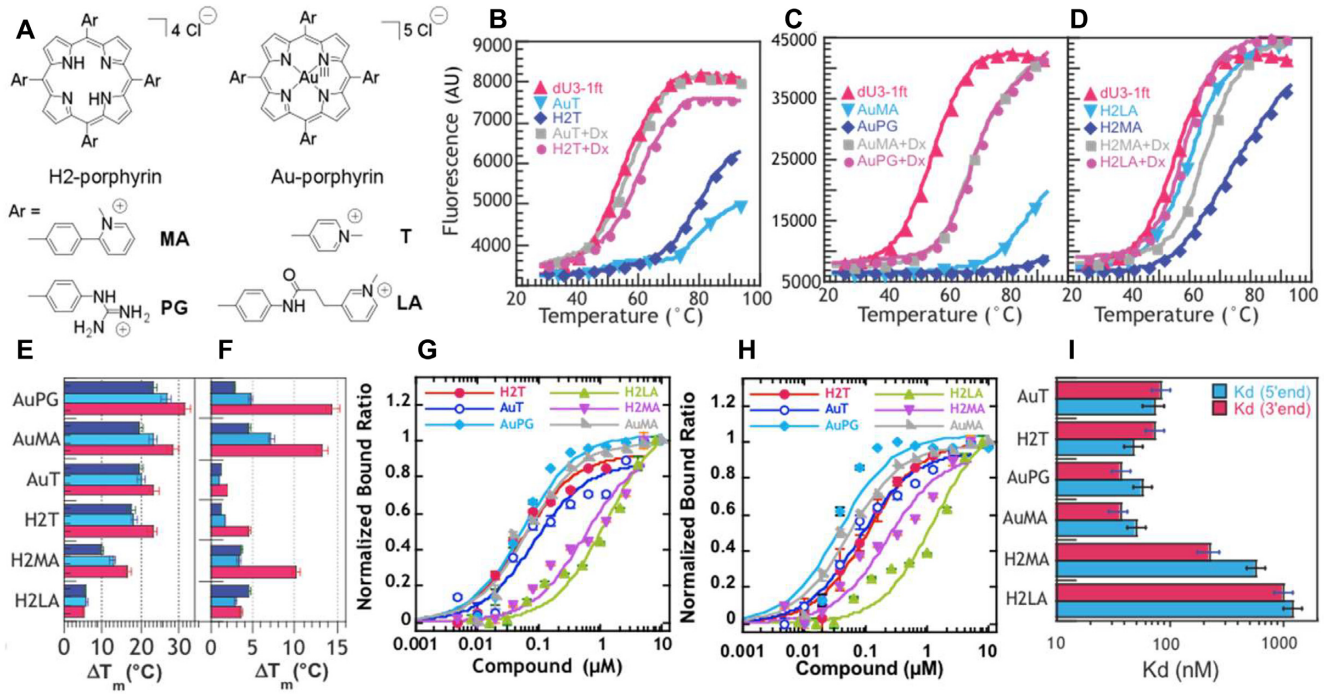


Figure 3. Interaction of porphyrin-based G4 binders with viral DNA and RNA G4 structures. (**A**) Chemical structures of the non-metalated and gold(III) porphyrin metallo-complexes studied here. The first generation porphyrin TMPyP4 is called here H2T (presence of two protons without gold(III)), which can be directly compared to the gold(III) metallo complex AuTMPyP4 called here AuT (two protons are replaced by gold(III)). (**B–F**) FRET melting assay. The DNA oligonucleotides were labeled with fluorophores (TAMRA and fluorescein) at either end and dissolved at a concentration of 0.2 μM in 10 mM KCl, 90 mM LiCl, and 10 mM lithium cacodylate at pH 7.2. Thermal denaturation experiments were performed by heating and recording the fluorescence emission of the fluorescein. Ligands were added at a concentration of 0.5 μM. As a first approximation, the higher the increase in $T_{1/2}$, the higher the affinity of the ligand for the target. (B–D) dU3-1-ft melting profiles were recorded in absence of ligand (pink triangles) or in the presence of 0.5 μM porphyrin. Competition experiments were performed in the presence of a competitor duplex (Table S1) at 10 μM (grey squares and violet circles). (E, F) Summary of the stabilization induced by the ligands on dU3-2 (dark blue), dU3-1 (red), and dU3-3 (cyan) quadruplexes in the (E) absence or (F) presence of 10 μM ds26 (Table S1b). (**G–I**) Ligand-induced fluorescence quenching assay evaluated the binding of the porphyrin ligands to rU3-1Cy5, which induces the quenching of the fluorescence of Cy5 dye. Conditions: 10 mM KCl, 90 mM LiCl, 10 mM Li-cacodylate pH 7.2 and 10 nM of rU3-1Cy5. (G) Saturation binding curves for the 5′ quartet of the 5′rU3-1Cy5 G4. (H) Saturation binding curves for the 3′ quartet of the 3′rU3-1Cy5 G4. The curves fitting allow the determination of K_d. (**I**) Histogram representation of the K_d values measured for porphyrin binding to the 5′ and 3′ end G quartets.

Table 2. Binding properties of the G4 ligands

Ligand	dU3-1-ft ΔT_m (°C)		dU3-2-ft ΔT_m (°C)		dU3-3-ft ΔT_m (°C)		rU3-1-cy5 K_d (μM)		HIV-1 Inhibition IC_{50} (μM)
	−Dx	+Dx	−Dx	+Dx	−Dx	+Dx	5′Cy5	3′Cy5	
H2LA	5.6	4.5	5.1	3.4	5.8	2.9	1.24 ± 0.24	1.04 ± 0.25	1.41 ± 0.13
H2MA	9.9	3.6	16.5	10.1	12.6	3.3	0.58 ± 0.17	0.22 ± 0.06	0.53 ± 0.04
AuMA	27.0	4.5	>30	13.2	25.0	7.2	0.06 ± 0.01	0.04 ± 0.01	0.39 ± 0.10
AuPG	>35	2.7	>35	14.4	>35	4.6	0.05 ± 0.01	0.04 ± 0.01	0.16 ± 0.01
AuT	>30	1.2	>30	1.8	>30	0.9	0.07 ± 0.02	0.08 ± 0.02	>20
TMPyP4 (H2T)	25.0	1.2	28.0	4.5	20.2	1.7	0.06 ± 0.01	0.07 ± 0.02	4.50 ± 0.30
Braco19	4.3	1.7	11.5	7.7	3.8	0.7	0.96 ± 0.27	0.20 ± 0.64	9.50 ± 0.10

ΔT_m: difference in T_m with and without G4 binder as determined by FRET melting experiments; the uncertainty of the value is ± 1°C. +/− Dx: FRET melting performed in the presence (+) or absence (−) of 10 μM (50-fold excess) of duplex competitor ds26 (Table S1).

of infection, i.e. infection is carried out immediately after ligands were added to the cell culture medium. No virus nor ligands washing step was performed. Of the porphyrins tested, AuPG and AuMA were most effective (IC_{50} values of 160 and 390 nM, respectively) (Table 2 and Figure 4A). TMPyP4 (H2T) and AuT, the two poorly selective G4 binders, had IC_{50} values higher than 4.5 μM, despite high binding affinities for G4 structures measured *in vitro*. We also used fluorescence microscopy to follow the cellular localization of the G4-ligands as a function of time (Figure S23A). Two hours after infection, ligands were observed in the cytoplasm of the cells and twenty-four hours after infection, ligands were also found in the nucleus. To confirm that the inhibition was not due to cellular toxicity of the ligands, viability assays were performed with AuMA and AuPG. No cytotoxic effect on human cells was observed at concentrations up to 10 μM for AuMa or AuPG, a concentration two orders of mag-

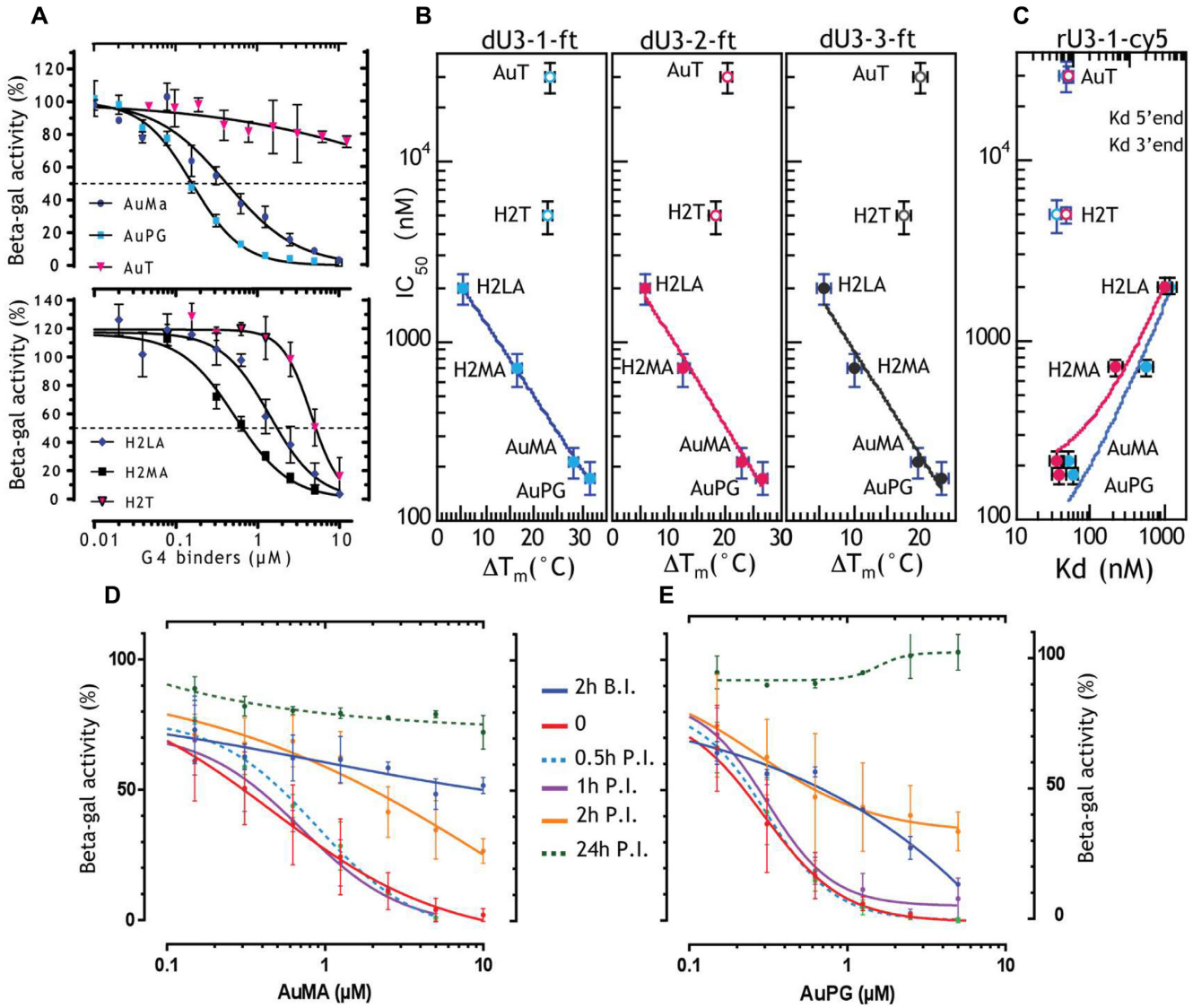


Figure 4. Analyses of effects of G4 binders on HIV-1 infectivity. (**A**) Dose effect of the G4 binders on HIV-1 infectivity in the presence of indicated porphyrins. Infection was performed with native HIVLai. The G4 ligands were added to the cells at the time of infection. β-Galactosidase activity was measured to evaluate the viral infectivity. Data are the results of at least three independent experiments each performed in duplicate. (**B**) Correlation between HIV-1 inhibition (IC_{50}) of each G4 binder and stabilization of DNA G4 conformations dU3-1ft (parallel), dU3-3ft (hybrid), and dU3-2ft (antiparallel) (ΔT_m, in°C). The best fit was obtaining using an exponential equation (AuT and H2T (TMPyP4)) were excluded from the fit). (**C**) Correlation between HIV-1 inhibition (IC_{50}) and K_d measured for the RNA rU3-1-cy5. The best fit was obtained using a linear equation (AuT and H2T were excluded from the fit). (**D**, **E**) Dose and incubation time effect of the G4 binders on HIV-1 infectivity in the presence of AuMA (D) or AuPG (E). Ligands were added two hours before infection (2 h B.I blue line) with a ligand washing step before infection; or at the time of infection (0 red line); or after infection (post-infection): half hour (dotted blue line), 1 h (violet line), 2 h (yellow line) or twenty four hours (green dotted line).

nitude above the IC_{50} values (Figure S24). In an additional experiment, no cytotoxic effects were observed when cells were incubated with 30 μM AuMa or AuPG for 96 hours (Figure S23B). We also verified that the ligands were not degraded in conditions that mimic the cellular medium (Figure S23C).

We then compared the effects on viral inhibition induced by the six ligands as a function of their ability to interact with the G4 structures *in vitro* (ΔT_m and K_d, Figure 4B, C). In the case of the three DNA G4s dU3-1, dU3-2 and dU3-3, we observed a good correlation (excluding TMPyP4 (H2T) and AuT) between the ΔT_m and HIV-1 IC_{50} inhibition for H2LA, H2MA, AuMA and AuPG (Figure 4B). AuT and TMPyP4 (H2T) were excluded due to their low binding specificity as shown by FRET competition experiments. AuPG and AuMA exhibited the highest ΔT_m and strongest antiviral effect (i.e. lowest IC_{50} values). We observed the same trend when comparing the K_d for U3 RNA and the IC_{50} (Figure 4C). For instance, when comparing H2LA and AuPG, a 25-fold decrease in K_d (i.e. stronger affinity) translated into a 9-fold decrease in IC_{50} (i.e. better inhibition). This correlation between *in vitro* interactions and *ex vivo* inhibition indirectly suggests that the observed

inhibition is due to the recognition of G4 structures located in the genomic RNA and/or DNA of the virus.

In order to further investigate the kinetics of the inhibition mechanism, we added AuMA or AuPG ligands at several times before or after infection is carried out (Figure 4D, E). When the ligands are added at the time of infection or thirty minutes or one hour after infection, a similar effect is observed in the three cases. When the compounds are added two hours after infection, a weaker inhibition is observed and no effects are detected when ligands are added twenty-four hours post-infection. We also added the ligands two hours before infection to allow the ligand to diffuse inside the cells. Then, we washed away the ligands that did not penetrate the cells before proceeding to infection. The ligand was still able to inhibit the virus but the effect was less efficient than when the molecules are added at the time of infection or shortly post-infection. Altogether these data suggest that the inhibition mechanism does not involve the viral entry step and occurs during the first hours post infection.

The G4 binders prevent the initiation of reverse transcription

In order to determine the actual stage of the viral life cycle which is inhibited by AuMA and AuPG ligands, we quantified the viral RNA and DNA present at each step of the viral life cycle (Figure 5A) by RT-qPCR in cells infected with the replicative virus. The effects of G4 binders on cellular uptake were assessed by quantification of the viral RNAs inside cells 2 hours after infection. The cells were infected in the presence of 2 μM porphyrin, a concentration at which AuPG and AuMA strongly (100 and 90%, respectively) inhibit viral replication. Neither AuPG nor AuMA had any effect on viral entry as the amount of viral RNA detected was the same within experimental error with and without porphyrin (Figure 5B).

To assess the effects of G4 binders on the reverse transcription step, viral DNAs were extracted 3, 6 and 24 h after infection and quantified by qPCR (Figure 5C, D). At 3 h post-infection, viral DNA was detected in absence of drug but not in the presence of 10 μM AZT or 2 μM AuPG or 2 μM AuMA (not shown). At 6 h after infection in the presence of AuMA or AZT, viral DNA levels were 4% and 1% of those in the absence of drug; no DNA was detected in the presence of AuPG (Figure 5C). At 24 h post infection, AuPG and AuMA inhibited the synthesis of DNA by 97% and 90%, respectively, and AZT inhibited synthesis by 68% (Figure 5D). Therefore, the G4 binders unambiguously inhibited the viral replication between the entry and the reverse transcription step. Notably, AuMA and AuPG ligands also proved to be more potent inhibitors than AZT.

Since the HIV-1 LTR contains several DNA G4s in its U3 region, we wanted to assess if the ligands had an effect on the DNA proviral stage by targeting DNA G4s and inhibiting transcription. To do so, we used latent Jurkat cells carrying a GFP reporter gene under LTR promoter control (Figure S23D). Before transcription reactivation of the reporter gene, the expression of GFP in cells was silent, with or without ligands, as quantified by flow cytometry (0 and 0.1% positive cells). Transcription was then reactivated by addition of 8 or 4 nM of PMA and the proportion of positive cells reached around 40% and 57% respectively. The addition of AuMA or AuPG ligand did not significantly affect the expression of the reporter gene. This indicates that the expression from the HIV-1 LTR was not affected by the G4 ligands and confirms that their effect was strictly due to the RNA targeting of the incoming virus.

We then attempted to accurately identify the step of the reverse transcription process blocked by the G4-ligands. Reverse transcription process (Figure 6A) begins with the synthesis of a 175-nt cDNA composed of the U5 and R region that is also called strong-stop cDNA. The complete cDNA is obtained after completion of the first DNA strand (first jump) and the second strand following the second jump of reverse transcription. For this analysis, we used primers specific for the strong-stop DNA and for the first and second jump cDNAs. The strong-stop cDNA was detected in cells infected by the virus in absence of drug but not in the presence of AuMA or AuPG (Figure 6B). No strong-stop cDNA was detected in control experiments in uninfected cells or without cellular extracts. We did not observe any amplification with the primers we used to detect the first or second jump cDNAs in the presence of AuMA or AuPG, although these cDNAs were detected in cells infected with virus in the absence of porphyrin (Figure 6C). To ensure that the absence of detection of DNA was not due to a lack of sensitivity of the method, qPCR was used for the detection of strong-stop DNA with a different set of primers, none was detected in the presence of the G4 ligands (Figure 6E). These data show that the G4 binders did not inhibit viral entry but did block the initiation of reverse transcription (Figure 6F).

Direct *in-cell* detection of viral RNA G4s using biotinylated G4 ligands

In order to directly detect the formation of RNA G4s in HIV-1 infected cells, we synthesized biotinylated derivatives of AuMA (45) which we named AuMA-biotin (Figure 7A and supplementary methods). We first checked the interaction between AuMA-biotin and fluorescently labeled rU3-1-cy5 RNA using the ligand-induced fluorescence quenching assay (Figure S25). The addition of the biotin tag to the AuMA scaffold decreased its binding affinity from around 60 nM to 440 nM, which still proved to be a robust interaction. As a result, AuMA-biotin was also able to inhibit HIV-1 infectivity with a slightly decreased efficiency than AuMA (IC_{50} values of 1.5 and 0.4 μM, respectively) (Figure S25). Therefore, AuMA-biotin retained most of the activity of the parent AuMA compound.

To detect the formation of viral RNA G4s over the course of the first two hours post-infection, we implemented a G4 RNA pull-down approach (Figure 7B) using the AuMA-biotin G4 probe (see G4 pull-down section in the methods). Two hours after infection, cells were cross-linked to stabilize transient structural interactions and then lysed by sonication. The AuMA-biotin probe and AuMA (used as a negative control) were incubated in the cell extracts. Subsequently, RNAs were captured with streptavidin beads and quantified after amplification by RT-qPCR. In order to design the G4 specific qPCR primers, we first evaluated the

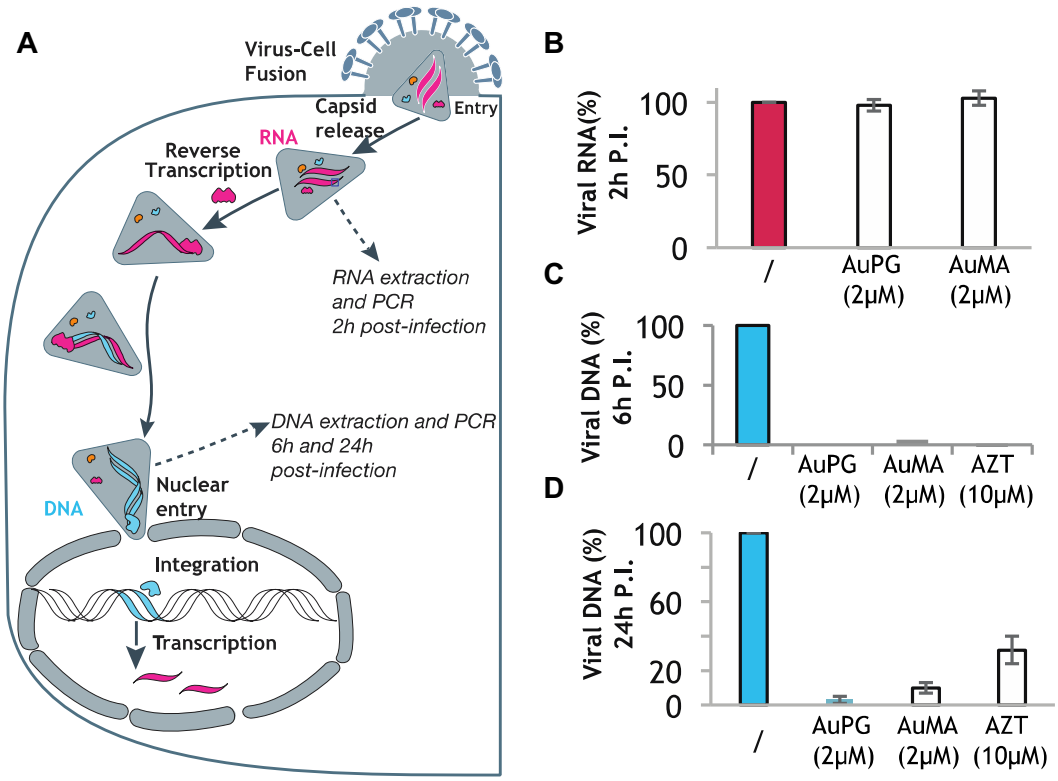


Figure 5. Determination of HIV-1 replication step inhibited by the G4 ligands. (**A**) Scheme of HIV-1 replication cycle highlighting the steps for viral nucleic acids extraction. (**B**) Quantification of viral RNA performed by RT-qPCR 2 h post infection in the absence of drug (red bar) and in the presence of 2 μM AuPG or 2 μM AuMA. (C and D) Quantification of viral DNA (**C**) 6 h and (**D**) 24 h post infection. Experiments were performed in the presence of 2 μM of G4 ligands or AZT at 10 μM concentration. Data are the results of at least three independent experiments each performed in duplicate.

sizes of the RNA fragments produced during the sonication process by capillary electrophoresis (Figure S26). The majority of the fragments were between 100 and 200 nucleotides in size and a minority of RNA fragments was up to 700 nucleotides-long. Therefore, the primers were designed to amplify products with sizes comprised between 118 and 219 nucleotides and displaying only one G4FS (Table S3). Of the 11 G4 sequences previously identified, we wanted to determine which of them were formed in infected cells. We did not consider the sequence 8-Env8, which does not form a G4, as well as the two sequences 3-POL3 and 7-ENV7 present only on the negative strand of the DNA genome, and thus absent on the viral RNA. We therefore selected primer pairs surrounding the remaining 8 G4FS as well as a viral RNA fragment (located in the middle of the gag gene) that do not present any putative G4FS (called Non-G4) and used here as negative control. The PCR products of 6 G4FS were separated on agarose gel (Figures 7C and S27) and showed a single band as expected. Amplification was unsuccessful for the sequences 6-VPR and 9-ENV9. Sonication might have resulted in RNA fragmentation patterns for which the primers were in two separated RNAs. The sequences 1-GAG, 2-POL2, 4-cPPT, 5-POL5, 10-PPT and 11-U3 were all successfully amplified in samples treated with the AuMA-biotin probe. Conversely, PCR products were barely detected in control samples incubated with AuMA, except for the sequences 2-POL2 and 11-U3 for which a less intense band was also detected. Furthermore, a precise comparison between the Ct values obtained with the AuMA-biotin probe or AuMA control showed that a six- to twenty-five-fold enrichment is obtained in the presence of G4 AuMA-biotin probe (Figure 7D). No enrichment was obtained in the case of AuMA control except for 11-U3 sequence that presented a two-fold enrichment which was still significantly smaller than the ten-fold enrichment measured for AuMA-biotin probe. The negative control unable to form a G-quadruplex (Non-G4) was hardly detected in both conditions highlighting the G4 specificity of the AuMA-biotin G4 probe. Hence, the 1-GAG, 2-POL2, 4-cPPT, 5-POL5, 10-PPT and 11-U3 RNA G4s were present in the cell two hours after infection and were detected by the biotinylated G4 probe.

DISCUSSION

HIV-1 is a rapidly evolving virus with high levels of genetic variability. This extreme diversity results from three important factors: the low fidelity of the reverse transcription process, RNA-level recombination events, and short generation times (66). This diversity prevents the immune system from eradicating the virus and renders it difficult to develop effective drugs and vaccines to fight infection. Conversely, some conserved secondary structures have important functions during the viral life cycle as illustrated by the TAR and RRE structural elements of HIV-1 (54). Our bioinformatics and biophysics analyses of HIV-1 genome identified evolu-

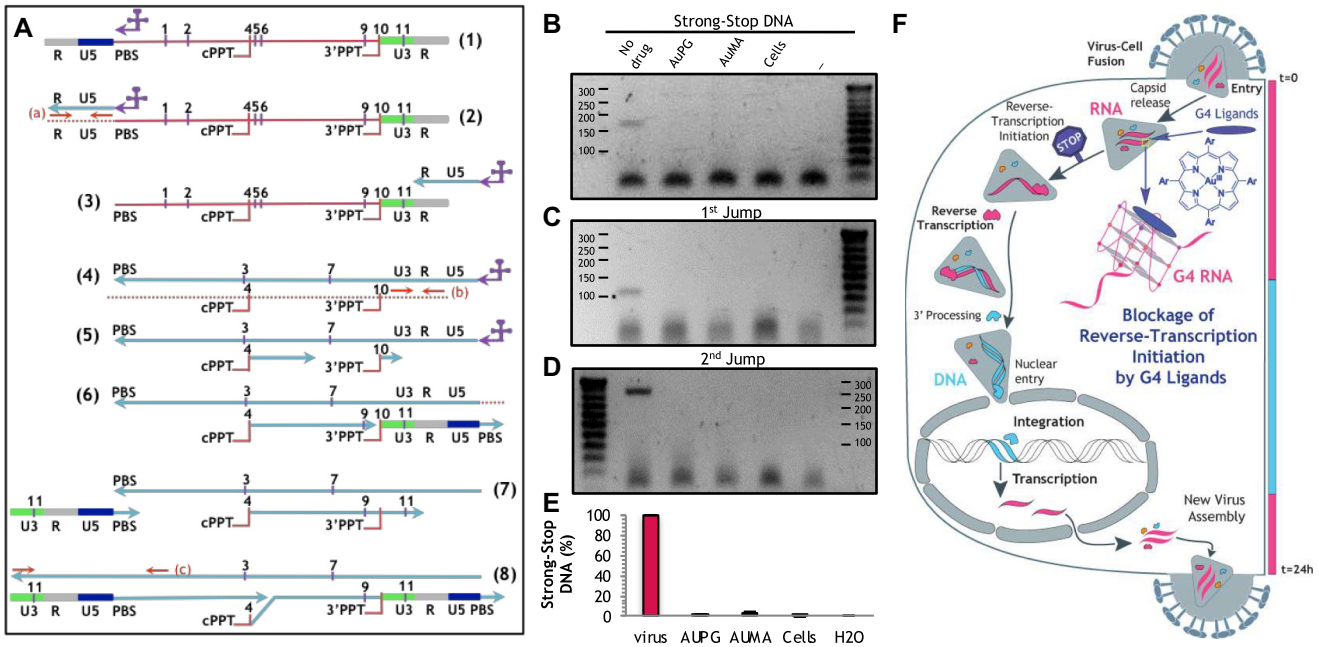


Figure 6. Effect of the G4 ligands on reverse transcription. (**A**) Scheme of the RT process. Viral DNA is in cyan, viral RNA is in red and $tRNA^{Lys}$ is in purple. The red arrows indicate positions of the pairs of primers used for PCR amplification of (a) strong-stop cDNA, (b) first jump cDNA, and (c) second jump cDNA. The G4FSs are indicated by numbers. The steps of the reverse transcription process are as follows: (1) The tRNA primer base pairs to the primer-binding site (PBS). (2) The reverse transcriptase initiates minus-strand DNA synthesis creating an RNA/DNA duplex on the U5-R region. RNase H degrades the RNA resulting in the strong-stop cDNA. (3) Via pairing with the R sequence of the strong-stop cDNA, the minus-strand DNA is translocated to the 3′ end of the viral RNA; this is the first jump cDNA. (4) The minus-strand DNA synthesis resumes. RNase H degrades the RNA that has been copied with exceptions of the PPT and cPPT sequences that are resistant to RNase H. (5) The PPTs serve as primers for plus-strand DNA synthesis. (6) The U3, R, U5 and PBS sequences of the plus DNA strand are synthesized. (7) After the removal of the tRNA primer, the PBS of the minus-strand DNA is translocated to the 5′ end of the minus-strand DNA. The plus-strand DNA synthesis resumes, and both the plus and minus-strand DNAs are then elongated. This leads to the second jump cDNA. (8) The plus strand synthesis initiated at the U3 junction displaces the segment of the plus-strand that was initiated from the cPPT, creating the central flap. (**B–E**) Detection and quantification of viral DNAs. Virus indicates DNA from cells infected with virus in the absence of drug; AuPG and AuMA indicate DNA from infected cells treated with 2 μM porphyrin; cells indicates DNA from uninfected cells; and the line indicates no DNA. (B) Gel separation of products of amplification with primers to strong-stop cDNA. (C) Gel separation of products of amplification with primers to first jump DNA. (D) Gel separation of products of amplification with primers to second jump cDNA. (E) Taqman amplification with primers to strong-stop cDNA. Percent strong-stop cDNA relative to virus-infected cells. (F) Mechanism of action of the porphyrin G4-binders during HIV-1 viral cycle. The main steps from the virus-cell fusion (top) to the assembly of new viruses (bottom) are represented. Viral RNA is colored in red and viral DNA is colored in cyan. The G4-binder is represented in blue. The G4 binders act during the two first hours after cell entry and prevent reverse transcription of viral RNA.

tionary conserved G4s with mutation rates even lower than those of RRE and TAR structural elements. In this context, the conservation and formation of the viral RNA G4s (1-GAG, 2-POL2, 4-cPPT, 5-POL5, 10-PPT and 11-U3) during the two first hours post-infection strongly suggest that these structures play a role in the virus life cycle, after cell entry and before reverse transcription starts. Some of the G4s studied here were identified in earlier studies: Evidence suggests that RNA G4s within the 1-GAG, 4-cPPT, and 11-U3 function in viral RNA recombination and/or dimerization processes (67–69). At the DNA level, the 11-U3 appears to be involved in transcription regulation of the provirus (42, 70). A role of 4-cPPT in the migration of the pre-integration complex into the nucleus was suggested (45), and an interaction of 10-PPT with topoisomerase I has been described (71).

In previous studies, G4 ligands including naphthalene diimides (72) and acridiniums (73) have been shown to inhibit HIV-1 replication. These antiviral effects pointed to a likely role of G4 structures in HIV-1, but the G4-based mechanisms of action remained elusive in most cases. Data from previous investigations suggest that G4 binders act mainly during the reverse transcription step with a minor contribution at post-integration levels. In our study, we used porphyrin-based G4 binders to probe the formation of viral G4s during HIV-1 infection. We showed that the porphyrin-based compounds acted after viral entry and impeded the initiation of the reverse transcription process. This mechanism of action indicates that viral RNA G4s are formed and are targetable during the early stages of the viral cycle within the first hours post-infection (Figure 6F).

There are several possible molecular mechanisms by which the G4 binders might block initiation of reverse transcription. Piekna-Przybylska *et al.* and earlier Sundquist *et al* proposed that genetic recombination and/or dimerization events between the two encapsulated viral RNAs are stabilized by formation of a bimolecular G4 involving the

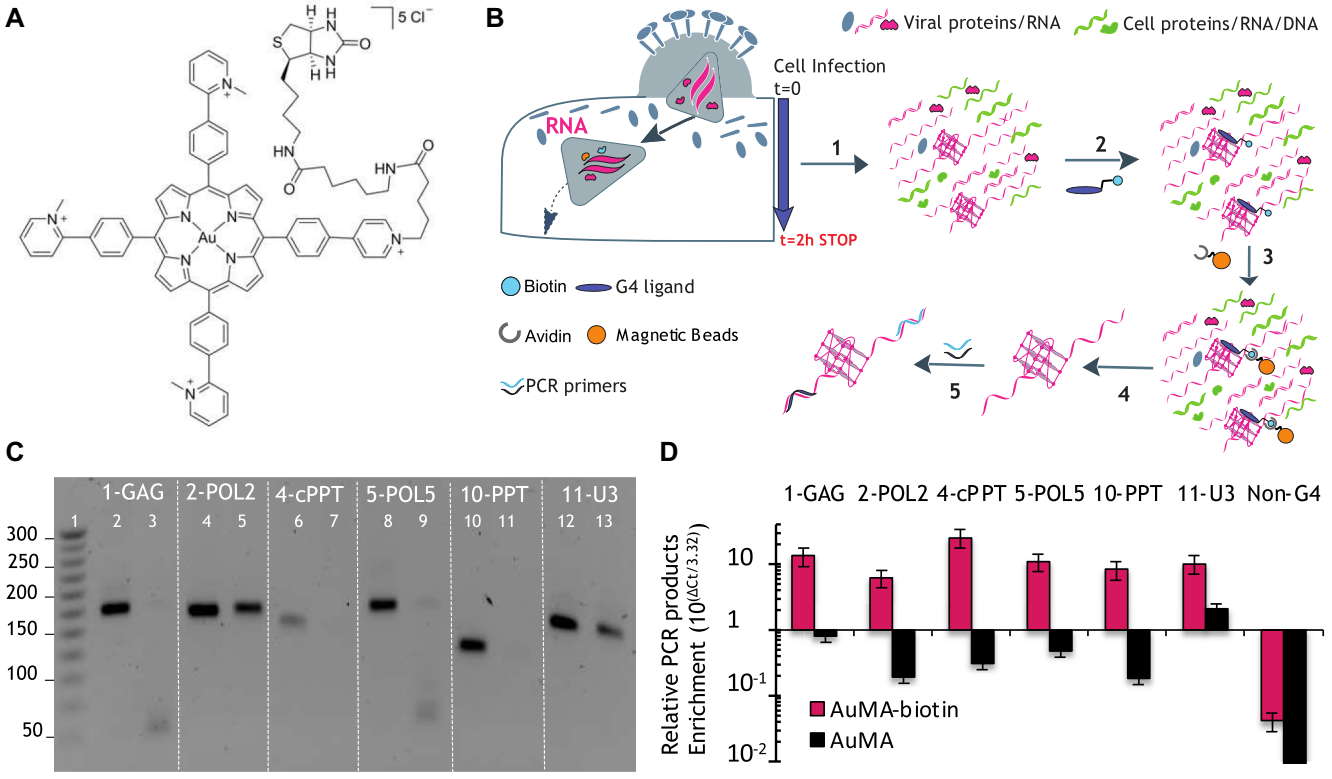


Figure 7. G4 RNA pull-down. (**A**) AuMA-biotin G4 probe. (**B**) G4 RNA pull-down strategy: 1) after 2 h infection, cells were fixed, cross-linked and lysed by sonication; 2) overnight incubation with AuMA or AuMA-biotin; 3) 2 h incubation with Streptavidin beads; 4) washing, reverse cross-linking and RNA extraction; 5) RNAs amplification and quantification by RT-qPCR. (**C**) Analysis of PCR products. Pull down and amplification of viral RNA was performed using primers described in Table S3. Lane 1: 25 bp DNA step ladder. Lanes 2–13: 10 µl of RT-qPCR products were injected on a 2% agarose gel stained with SYBR safe. For each G4FS are shown the incubation products for AuMA-biotin (left) and AuMA (right); a cell negative control is provided in figure S27. (**D**) Relative PCR products enrichment for 6 G4FS and a non-G4 control. The ΔCt values for each condition were obtained by substracting the Ct value of the input by the Ct value of AuMA−biotin or AuMA (Table S3). We considered that a ΔCt value of 3.32 results from a 10-time enrichment of the sequence. (=) The actual value for non-G4/AuMA is 0.001.

rCPPT-G4, the rGAG-G4 or the rU3-G4 (67–69). These G4s could form transient G4 bridges that facilitate the exchange of material between the donor RNA and the recipient RNA and may also help the reverse transcriptase switch templates during DNA synthesis. The NCp7 chaperone protein, known to facilitate the packaging of the viral RNA, could also play a role in dimerization or recombination events by promoting the formation of a bimolecular G4s (74–76). Recently, Butovskaya *et al.* showed that NCp7 is able to bind and unfold rU3-G4s *in vitro* but this effect can be partially counteracted by the presence of G4 binders (77). In theory, the other viral G4FS identified here (rENV9, rVPR, rPOL2 and rPOL5) could also play a role in dimerization or recombination.

In the initial stages of the reverse transcription process, a complex mechanism enables transfer of the growing minus-strand DNA from the 5′ end of the viral RNA to a remote region at around 9000 nucleotides away located at the 3′ end of the genome. This process is called minus-strand DNA transfer (Figure 6A, steps 2 and 3). It has been proposed that the primer $tRNA^{Lys3}$ simultaneously binds the PBS (primer binding sequence) at the 5′ end and the S1 sequence (5′-GCCUGGGCGGGACU-3′) located at the 3′ end. This dual interaction induces a circularization of the RNA genome that brings the two R regions into proximity and triggers the minus-strand transfer (78–80). Interestingly, the S1 sequence is embedded in the U3-G4 studied in this work (Table 1). Therefore, it is possible that the stabilization of U3-G4 by the ligands may impede the formation of this ternary complex, thus blocking the initiation of the reverse transcription process.

It is also possible that G4s participate to plus-strand and/or minus-strand DNA synthesis. The rPPT-G4FS, located at the boundary between the NEF gene and the U3 region, contains the 3′ polypurine tract (5′-AAAGAAAAGGGGGGACU-3′). The CPPT-G4FS, located in the center of the genome, contains the central polypurine tract (5′-AAAGAAAAGGGGGGAUU-3′). Both sequences are important in the reverse transcription process (Figure 6A, steps 5 to 6). After the synthesis of the minus strand, all the viral RNA that initially served as template is degraded by RNase H except for these two RNA tracts that are used as primers for the initiation of plus-strand synthesis. Notably, several studies have shown that RNase H is able to bind to G4 structures such as the 93del aptamer (81). The resistance of these polypurine tracts may be related to their abilities to form a G4 structure and to bind to RNase H.

CONCLUSIONS

During the early stages of HIV-1 life cycle, the RNA genetic material must be prepared for initiation of the reverse transcription. This process involves perfectly synchronized 3D rearrangements that will influence the overall progression of viral replication. Nucleocapsid proteins and the reverse transcriptase are also involved in this process. Our investigations demonstrated the formation of highly conserved viral RNA G4s during this early stage of the viral cycle. Some of these RNA G4s contain crucial regulatory elements like the PPT and cPPT sequences as well as the U3 region suggesting an important function in HIV-1 life cycle. Indeed, we have shown that, during the first 2 hours of the viral cycle, the stabilization of a viral G4 intermediate by the G4 binders locks the viral replication system and impede the initiation of reverse transcription. This unprecedented inhibition mechanism suggests that stabilizing G4s with small molecules during the early stages of the viral cycle would block infection. It is interesting to note that these G4 targets are highly conserved in the HIV-1 genome as shown by our analysis of about 2200 HIV-1 sequences. Therefore, targeting these G4 RNA structures represents a promising new antiviral avenue with limited risks of developing resistance. Indeed, such drugs designed to bind conserved viral elements should counteract the ability of HIV-1 to evolve drug resistance. Gold(III) porphyrins could also inhibit infections by other viruses that employ RNA G4s during their replication cycles.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank Dr. David Monchaud for his assistance in the design of the G4RP experiment. This work was conducted with the support of the *ANRS Maladies Infectieuses Émergentes – Agence autonome de l'Inserm*, the *Centre National de la Recherche Scientifique* (CNRS), the *Institut National de la Santé et de la Recherche Médicale* (Inserm), and the *Université de Bordeaux*.
Author contributions: S.A., M.L.A., GP designed the study. S.A., M.L.A., J.L.M., G.P. co-wrote the paper. A.B., O.G., F.L. performed the bioinformatics analysis. S.A., A.G., A.B., L.B performed the biophysical analysis of the G4s. S.A., A.G., A.D.R., I.D performed the biophysical analysis of the G4-binders. G.P., T.R. synthesized the porphyrin based G4-binders. C.J, C.A., P.P., M.L.A performed the viral infectivity assays, the qRT-PCR analysis, and the pull down experiments.

FUNDING

Agence Nationale de Recherches sur le Sida et les Hepatites Virales (ANRS) [ECTZ35927, ECTZ103899, ANRS AAP1-2015]. Funding for open access charge: INSERM. O.G. is supported by PRAIRIE (ANR-19-P3IA-0001). C.J. and A.D.R. benefited from post-doctoral fellowships from the ANRS (ECTZ35927 and ANRS-AAP1-2015). A.D.R. thanks the Belgian Fonds national de la Recherche Scientifique (FNRS) for her Postdoctoral Researcher grant. L.B. benefited from PhD fellowship from the region Nouvelle Aquitaine and INSERM.
Conflict of interest statement. None declared.

REFERENCES

1. Patel,D.J., Phan,A.T. and Kuryavyi,V. (2007) Human telomere, oncogenic promoter and 5′-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics. *Nucleic Acids Res.*, **35**, 7429–7455.
2. Guédin,A., Alberti,P. and Mergny,J.-L. (2009) Stability of intramolecular quadruplexes: sequence effects in the central loop. *Nucleic Acids Res.*, **37**, 5559–5567.
3. Varshney,D., Spiegel,J., Zyner,K., Tannahill,D. and Balasubramanian,S. (2020) The regulation and functions of DNA and RNA G-quadruplexes. *Nat. Rev. Mol. Cell Biol.*, **21**, 459–474.
4. Marquevielle,J., De Rache,A., Vialet,B., Morvan,E., Mergny,J.-L. and Amrane,S. (2022) G-quadruplex structure of the *C. elegans* telomeric repeat: a two tetrads basket type conformation stabilized by a non-canonical C-T base-pair. *Nucleic Acids Res.*, **50**, 7134–7146.
5. Paeschke,K., Simonsson,T., Postberg,J., Rhodes,D. and Lipps,H.J. (2005) Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. *Nat. Struct. Mol. Biol.*, **12**, 847–854.
6. Schaffitzel,C., Berger,I., Postberg,J., Hanes,J., Lipps,H.J. and Plückthun,A. (2001) In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 8572–8577.
7. Paeschke,K. and Burkovics,P. (2021) Mgs1 function at G-quadruplex structures during DNA replication. *Curr. Genet.*, **67**, 225–230.
8. Kim,N. (2019) The Interplay between G-quadruplex and Transcription. *Curr. Med. Chem.*, **26**, 2898–2917.
9. Schult,P. and Paeschke,K. (2020) The DEAH helicase DHX36 and its role in G-quadruplex-dependent processes. *Biol. Chem.*, **402**, 581–591.
10. Varshney,D., Spiegel,J., Zyner,K., Tannahill,D. and Balasubramanian,S. (2020) The regulation and functions of DNA and RNA G-quadruplexes. *Nat. Rev. Mol. Cell Biol.*, **21**, 459–474.
11. Bedrat,A., Lacroix,L. and Mergny,J.-L. (2016) Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.*, **44**, 1746–1759.
12. Brázda,V., Kolomazník,J., Lýsek,J., Bartas,M., Fojta,M., Šťastný,J. and Mergny,J.-L. (2019) G4Hunter web application: a web server for G-quadruplex prediction. *Bioinforma. Oxf. Engl.*, **35**, 3493–3495.
13. Lee,D.S.M., Ghanem,L.R. and Barash,Y. (2020) Integrative analysis reveals RNA G-quadruplexes in UTRs are selectively constrained and enriched for functional associations. *Nat. Commun.*, **11**, 527.
14. Lam,E.Y.N., Beraldi,D., Tannahill,D. and Balasubramanian,S. (2013) G-quadruplex structures are stable and detectable in human genomic DNA. *Nat. Commun.*, **4**, 1796.
15. Biffi,G., Tannahill,D., McCafferty,J. and Balasubramanian,S. (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.*, **5**, 182–186.
16. Biffi,G., Di Antonio,M., Tannahill,D. and Balasubramanian,S. (2014) Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. *Nat. Chem.*, **6**, 75–80.
17. Saranathan,N. and Vivekanandan,P. (2019) G-Quadruplexes: more than just a kink in microbial genomes. *Trends Microbiol.*, **27**, 148–163.
18. Métifiot,M., Amrane,S., Litvak,S. and Andreola,M.-L. (2014) G-quadruplexes in viruses: function and potential therapeutic applications. *Nucleic Acids Res.*, **42**, 12352–12366.
19. Harris,L.M. and Merrick,C.J. (2015) G-quadruplexes in pathogens: a common route to virulence control?*PLoS Pathog.*, **11**, e1004562.
20. Meier-Stephenson,V. (2022) G4-quadruplex-binding proteins: review and insights into selectivity. *Biophys. Rev.*, **14**, 635–654.
21. Ray,S, Bandaria,JN, Qureshi,MH, Yildiz,A and Balci,H (2014) G-quadruplex formation in telomeres enhances POT1/TPP1 protection against RPA binding. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 2990–2995.

22. Hwang,H., Buncher,N., Opresko,P.L. and Myong,S. (2012) POT1–TPP1 regulates telomeric overhang structural dynamics. *Structure*, **20**, 1872–1880.
23. Budhathoki,JB, Ray,S, Urban,V, Janscak,P, Yodh,JG and Balci,H (2014) RecQ-core of BLM unfolds telomeric G-quadruplex in the absence of ATP. *Nucleic Acids Res.*, **42**, 11528–11545.
24. Paeschke,K, Bochman,ML, Garcia,PD, Cejka,P, Friedman,KL, Kowal- czykowski,SC and Zakian,VA (2013) Pif1 family helicases suppress genome instability at G-quadruplex motifs. *Nature*, **497**, 458–462.
25. Da Ros,S., Nicoletto,G., Rigo,R., Ceschi,S., Zorzan,E., Dacasto,M., Giantin,M. and Sissi,C. (2021) G-Quadruplex modulation of SP1 functional binding sites at the KIT proximal promoter. *Int. J. Mol. Sci.*, **22**, 329.
26. Edwards,A.D., Marecki,J.C., Byrd,A.K., Gao,J. and Raney,K.D. (2021) G-Quadruplex loops regulate PARP-1 enzymatic activation. *Nucleic Acids Res.*, **49**, 416–431.
27. González,V., Guo,K., Hurley,L. and Sun,d,D. (2009) Identification and Characterization of Nucleolin as a c-myc G-quadruplex-binding Protein. *J. Biol. Chem.*, **284**, 23622–23635.
28. Murat,P., Marsico,G., Herdy,B., Ghanbarian,A., Portella,G. and Balasubramanian,S. (2018) RNA G-quadruplexes at upstream open reading frames cause DHX36- and DHX9-dependent translation of human mRNAs. *Genome Biol.*, **19**, 229.
29. Phan,A.T., Kuryavyi,V., Darnell,J.C., Serganov,A., Majumdar,A., Ilin,S., Raslin,T., Polonskaia,A., Chen,C., Clain,D. *et al.* (2011) Structure-function studies of FMRP RGG peptide recognition of an RNA duplex-quadruplex junction. *Nat. Struct. Mol. Biol.*, **18**, 796–804.
30. Chen,M.C., Tippana,R., Demeshkina,N.A., Murat,P., Balasubramanian,S., Myong,S. and Ferré-D'Amaré,A.R. (2018) Structural basis of G-quadruplex unfolding by the DEAH/RHA helicase DHX36. *Nature*, **558**, 465–469.
31. Ladame,S., Schouten,J.A., Roldan,J., Redman,J.E., Neidle,S. and Balasubramanian,S. (2006) Exploring the recognition of quadruplex DNA by an engineered Cys2-His2 zinc finger protein. *Biochemistry*, **45**, 1393–1399.
32. Horvath,M.P. and Schultz,S.C. (2001) DNA G-quartets in a 1.86 Å resolution structure of an *Oxytricha nova* telomeric protein-DNA complex11Edited by I. Tinoco. *J. Mol. Biol.*, **310**, 367–377.
33. Yao,H., He,S., Henry,C., Baocheng,L., Yaqiang,W., Lukas,S., Zhou,Z.H. and Feigon,J. (2022) Structure of tetrahymena telomerase-bound CST with polymerase α-primase. *Nature*, **608**, 813–818.
34. Neidle,S. (2016) Quadruplex nucleic acids as novel therapeutic targets. *J. Med. Chem.*, **59**, 5987–6011.
35. Abiri,A., Lavigne,M., Rezaei,M., Nikzad,S., Zare,P., Mergny,J.-L. and Rahimi,H.-R. (2021) Unlocking G-quadruplexes as antiviral targets. *Pharmacol. Rev.*, **73**, 897–923.
36. Luo,J., Wei,W., Waldispühl,J. and Moitessier,N. (2019) Challenges and current status of computational methods for docking small molecules to nucleic acids. *Eur. J. Med. Chem.*, **168**, 414–425.
37. Kar,R.K., Suryadevara,P., Jana,J., Bhunia,A. and Chatterjee,S. (2013) Novel G-quadruplex stabilizing agents: in-silico approach and dynamics. *J. Biomol. Struct. Dyn.*, **31**, 1497–1518.
38. Wang,Y.-H., Yang,Q.-F., Lin,X., Chen,D., Wang,Z.-Y., Chen,B., Han,H.-Y., Chen,H.-D., Cai,K.-C., Li,Q. *et al.* (2021) G4LDB 2.2: a database for discovering and studying G-quadruplex and i-motif ligands. *Nucleic Acids Res.*, **50**, D150–D160.
39. Ruggiero,E. and Richter,S.N. (2020) Viral G-quadruplexes: New frontiers in virus pathogenesis and antiviral therapy. *Annu. Rep. Med. Chem.*, **54**, 101–131.
40. Jaubert,C., Bedrat,A., Bartolucci,L., Di Primo,C., Ventura,M., Mergny,J.-L., Amrane,S. and Andreola,M.-L. (2018) RNA synthesis is modulated by G-quadruplex formation in Hepatitis C virus negative RNA strand. *Sci. Rep.*, **8**, 8120.
41. Ruggiero,E. and Richter,S.N. (2018) G-quadruplexes and G-quadruplex ligands: targets and tools in antiviral therapy. *Nucleic Acids Res.*, **46**, 3270–3283.
42. Amrane,S., Kerkour,A., Bedrat,A., Vialet,B., Andreola,M.-L. and Mergny,J.-L. (2014) Topology of a DNA G-quadruplex structure formed in the HIV-1 promoter: a potential target for anti-HIV drug development. *J. Am. Chem. Soc.*, **136**, 5249–5252.
43. Perrone,R., Nadai,M., Frasson,I., Poe,J.A., Butovskaya,E., Smithgall,T.E., Palumbo,M., Palù,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.
44. De Nicola,B., Lech,C.J., Heddi,B., Regmi,S., Frasson,I., Perrone,R., Richter,S.N. and Phan,A.T. (2016) Structure and possible function of a G-quadruplex in the long terminal repeat of the proviral HIV-1 genome. *Nucleic Acids Res.*, **44**, 6442–6451.
45. Lyonnais,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.
46. 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.
47. Lemoine,F. and Gascuel,O. (2021) Gotree/Goalign: toolkit and Go API to facilitate the development of phylogenetic workflows. *NAR Genomics Bioinforma.*, **3**, lqab075.
48. Mergny,J.-L., Li,J., Lacroix,L., Amrane,S. and Chaires,J.B. (2005) Thermal difference spectra: a specific signature for nucleic acid structures. *Nucleic Acids Res.*, **33**, e138–e138.
49. De Rache,A. and Mergny,J.-L. (2015) Assessment of selectivity of G-quadruplex ligands via an optimised FRET melting assay. *Biochimie*, **115**, 194–202.
50. D. Le,D., Di Antonio,M., M. Chan,L.K. and Balasubramanian,S. (2015) G-quadruplex ligands exhibit differential G-tetrad selectivity. *Chem. Commun.*, **51**, 8048–8050.
51. Ventura,M., Tarrago-Litvak,L., Dollé,V., Nguyen,C.H., Legraverend,M., Fleury,H.J. and Litvak,S. (1999) Effect of nucleoside analogs and non-nucleoside inhibitors of HIV-1 reverse transcriptase on cell-free virions. *Arch. Virol.*, **144**, 513–523.
52. Charneau,P., Mirambeau,G., Roux,P., Paulous,S., Buc,H. and Clavel,F. (1994) HIV-1 reverse transcription. A termination step at the center of the genome. *J. Mol. Biol.*, **241**, 651–662.
53. Yang,S.Y., Lejault,P., Chevrier,S., Boidot,R., Robertson,A.G., Wong,J.M.Y. and Monchaud,D. (2018) Transcriptome-wide identification of transient RNA G-quadruplexes in human cells. *Nat. Commun.*, **9**, 4730.
54. Watts,J.M., Dang,K.K., Gorelick,R.J., Leonard,C.W., Bess,J.W. Jr, Swanstrom,R., Burch,C.L. and Weeks,K.M. (2009) Architecture and Secondary Structure of an Entire HIV-1 RNA Genome. *Nature*, **460**, 711–716.
55. Firth,A.E. (2014) Mapping overlapping functional elements embedded within the protein-coding regions of RNA viruses. *Nucleic Acids Res.*, **42**, 12425–12439.
56. Lefort,V., Desper,R. and Gascuel,O. (2015) FastME 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program. *Mol. Biol. Evol.*, **32**, 2798–2800.
57. Guindon,S., Dufayard,J.-F., Lefort,V., Anisimova,M., Hordijk,W. and Gascuel,O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.*, **59**, 307–321.
58. Butovskaya,E., Heddi,B., Bakalar,B., Richter,S.N. and Phan,A.T. (2018) Major G-quadruplex form of HIV-1 LTR reveals a (3 + 1) folding topology containing a stem-loop. *J. Am. Chem. Soc.*, **140**, 13654–13662.
59. Dixon,I.M., Lopez,F., Tejera,A.M., Estève,J.-P., Blasco,M.A., Pratviel,G. and Meunier,B. (2007) A G-quadruplex ligand with 10000-fold selectivity over duplex DNA. *J. Am. Chem. Soc.*, **129**, 1502–1503.
60. Pipier,A., De Rache,A., Modeste,C., Amrane,S., Mothes-Martin,E., Stigliani,J.-L., Calsou,P., Mergny,J.-L., Pratviel,G. and Gomez,D. (2019) G-Quadruplex binding optimization by gold(iii) insertion into the center of a porphyrin. *Dalton Trans. Camb. Engl.*, **48**, 6091–6099.
61. Amrane,S., Andreola,M., Pratviel,G. and Mergny,J.L. (2020) Derivatives of porphyrins, their process of preparation and their use for treating viral infections. *US Patent*, **10**, 577–375.
62. Sabater,L., Nicolau-Travers,M.-L., De Rache,A., Prado,E., Dejeu,J., Bombarde,O., Lacroix,J., Calsou,P., Defrancq,E., Mergny,J.-L. *et al.* (2015) The nickel(II) complex of guanidinium phenyl porphyrin, a specific G-quadruplex ligand, targets telomeres and leads to POT1 mislocalization in culture cells. *J. Biol. Inorg. Chem.*, **20**, 729–738.

63. 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.
64. Sabater,L., Fang,P.-J., Chang,C.-F., De Rache,A., Prado,E., Dejeu,J., Garofalo,A., Lin,J.-H., Mergny,J.-L., Defrancq,E. *et al.* (2015) Cobalt(III)porphyrin to target G-quadruplex DNA. *Dalton Trans.*, **44**, 3701–3707.
65. Rundstadler,T., Mothes,E., Amrane,S., Stigliani,J.-L., Verhaeghe,P. and Pratviel,G. (2021) Gold(III) porphyrins: synthesis and interaction with G-quadruplex DNA. *J. Inorg. Biochem.*, **223**, 111551.
66. Andrews,S.M. and Rowland-Jones,S. (2017) Recent advances in understanding HIV evolution. *F1000Research*, **6**, 597.
67. Piekna-Przybylska,D., Sullivan,M.A., Sharma,G. and Bambara,R.A. (2014) U3 region in the HIV-1 genome adopts a G-quadruplex structure in its RNA and DNA sequence. *Biochemistry*, **53**, 2581–2593.
68. 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.
69. Piekna-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.
70. Perrone,R., Lavezzo,E., Palù,G. and Richter,S.N. (2017) Conserved presence of G-quadruplex forming sequences in the long terminal repeat promoter of lentiviruses. *Sci. Rep.*, **7**, 2018.
71. Arimondo,P.B., Riou,J.-F., Mergny,J.-L., Tazi,J., Sun,J.-S., Garestier,T. and Hélène,C. (2000) Interaction of human DNA topoisomerase I with G-quartet structures. *Nucleic Acids Res.*, **28**, 4832–4838.
72. Perrone,R., Doria,F., Butovskaya,E., Frasson,I., Botti,S., Scalabrin,M., Lago,S., Grande,V., Nadai,M., Freccero,M. *et al.* (2015) Synthesis, binding and antiviral properties of potent core-extended naphthalene diimides targeting the HIV-1 long terminal repeat promoter G-quadruplexes. *J. Med. Chem.*, **58**, 9639–9652.
73. Perrone,R., Butovskaya,E., Daelemans,D., Palù,G., Pannecouque,C. and Richter,S.N. (2014) Anti-HIV-1 activity of the G-quadruplex ligand BRACO-19. *J. Antimicrob. Chemother.*, **69**, 3248–3258.
74. Lyonnais,S., Gorelick,R.J., Mergny,J.-L., Le Cam,E. and Mirambeau,G. (2003) G-quartets direct assembly of HIV-1 nucleocapsid protein along single-stranded DNA. *Nucleic Acids Res.*, **31**, 5754–5763.
75. Rajendran,A., Endo,M., Hidaka,K., Tran,P.L.T., Mergny,J.-L., Gorelick,R.J. and Sugiyama,H. (2013) HIV-1 nucleocapsid proteins as molecular chaperones for tetramolecular antiparallel G-quadruplex formation. *J. Am. Chem. Soc.*, **135**, 18575–18585.
76. 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.
77. Butovskaya,E., Soldà,P., Scalabrin,M., Nadai,M. and Richter,S.N. (2019) HIV-1 nucleocapsid protein unfolds stable RNA G-quadruplexes in the viral genome and is inhibited by G-quadruplex ligands. *ACS Infect. Dis.*, **5**, 2127–2135.
78. Piekna-Przybylska,D., Dykes,C., Demeter,L.M. and Bambara,R.A. (2011) Sequences in the U3 region of human immunodeficiency virus 1 improve efficiency of minus strand transfer in infected cells. *Virology*, **410**, 368–374.
79. Piekna-Przybylska,D., DiChiacchio,L., Mathews,D.H. and Bambara,R.A. (2010) A sequence similar to tRNA 3 Lys gene is embedded in HIV-1 U3-R and promotes minus-strand transfer. *Nat. Struct. Mol. Biol.*, **17**, 83–89.
80. Piekna-Przybylska,D. and Bambara,R.A. (2011) Requirements for efficient minus strand strong-stop DNA transfer in human immunodeficiency virus 1. *RNA Biol.*, **8**, 230–236.
81. Faure-Perraud,A., Métifiot,M., Reigadas,S., Recordon-Pinson,P., Parissi,V., Ventura,M. and Andréola,M.-L. (2011) The guanine-quadruplex aptamer 93del inhibits HIV-1 replication ex vivo by interfering with viral entry, reverse transcription and integration. *Antivir. Ther.*, **16**, 383–394.