scientific reports

OPEN

Helquat dyes targeting G-quadruplexes as a new class of anti-HIV-1 inhibitors

Marcela Pávová, Paul Eduardo Reyes-Gutiérrez, Jaroslav Kozák, Juraj Dobiaš, Yevgen Yurenko, Martin Lepšík, Filip Teplý & Jan Weber ✉

The secondary structure of nucleic acids containing quartets of guanines, termed G-quadruplexes, is known to regulate the transcription of many genes. Several G-quadruplexes can be formed in the HIV-1 long terminal repeat promoter region and their stabilization results in the inhibition of HIV-1 replication. Here, we identified helquat-based compounds as a new class of anti-HIV-1 inhibitors that inhibit HIV-1 replication at the stage of reverse transcription and provirus expression. Using Taq polymerase stop and FRET melting assays, we have demonstrated their ability to stabilize G-quadruplexes in the HIV-1 long-terminal repeat sequence. Moreover, these compounds were not binding to the general G-rich region, but rather to G-quadruplex-forming regions. Finally, docking and molecular dynamics calculations indicate that the structure of the helquat core greatly affects the binding mode to the individual G-quadruplexes. Our findings can provide useful information for the further rational design of inhibitors targeting G-quadruplexes in HIV-1.

It is well accepted that G-quadruplexes (G4s) play an important regulatory role in the promoters of many cellular genes[1], oncogenes in particular, and that their misregulation is associated with various human diseases[2–4]. Over the last few years, the presence of G4s in human pathogens[5–7], and especially in viruses[8], and their involvement in many key steps of viral life cycles have attracted increasing interest[9–15] (summarized in extensive reviews[16–19]).

G4s are promising pharmacological targets, because the distinct molecular features of the individual G4s enable selective recognition by small molecules with minimal side effects[20]. This approach is being investigated in anticancer research by targeting G4s located in telomeres or oncogenes[21–24] and can be adopted in antiviral research, as documented by Perrone et al., who originally demonstrated the anti-HIV activity of the G4 ligand BRACO-19[11]. This finding was followed by others, resulting in the publication of the antiviral properties of other G4-binding ligands[10,25–33].

Retroviruses are multifaceted viruses that infect almost all vertebrates and cause severe diseases, including immunodeficiency, neurological disorders, and various types of cancer, representing a major threat across the animal kingdom. The human immunodeficiency virus type 1 (HIV-1), the etiological agent of the acquired immunodeficiency syndrome (AIDS), establishes a persistent infection in cells of the human immune system. Recently, Bohálová et al. analysed all available viral genomes in the NCBI database (11,000 genomes in total) for putative G-quadruplex forming sequences (PQS) and showed that viruses causing persistent infection, such as HIV-1, are enriched for the presence of PQS whereas the viruses causing acute infections only are often reduced in PQS[34]. The presence of G4 structures has been described in both RNA and DNA forms of its genome, with implications throughout the HIV-1 life cycle[35,36]. Perhaps most importantly, the HIV-1 promoter region, specifically the U3 region of the 5′ LTR (long terminal repeat), contains a G-rich sequence of 50 nucleotides upstream from the transcription-starting site and close to the TATA box (Fig. 1). This region readily takes on different G4 configurations within both the RNA virus genome and the single-stranded DNA form during the transcription of the integrated provirus[36]. This region contains several G-patches (labelled as 1–6′ in Fig. 1) that are, based on computational, biophysical and biochemical analyses, able to form three mutually exclusive G4s. Several publications have identified G-patches 3″–6 as a region indispensable for the formation of all relevant G4s[11,36,37]. Moreover, Perrone et al. have observed the formation of an additional G4, exploiting G-patches 4–6′, after treatment with a G4-binding ligand[11], and Butovskaya et al. have confirmed this G4 as the major G4 structure formed in the HIV-1 LTR[38]. The core of the G4-forming region is composed of G-patches 4 and 5, as documented by two point mutations (G61A and G65A) that completely erase the ability of this region to form any G4[11].

Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, Prague 160 00, Czech Republic. ✉email: jan.weber@uochb.cas.cz
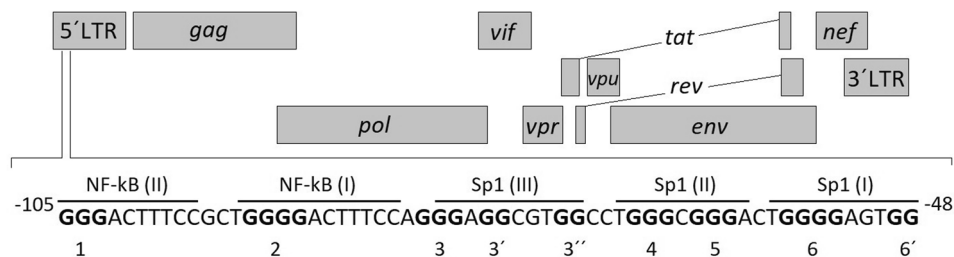
Figure 1. HIV-1 genomic organization and a sequence of the G-rich LTR region spanning nucleotides − 105/− 48. The lines indicate transcription factor binding sites. The G bases involved in the formation of G-quadruplexes are in bold and G-tracts are numbered.

Despite the considerable success of antiretroviral therapy in the treatment of HIV/AIDS, there is an urgent need for new anti-HIV drugs targeting highly conserved features of HIV-1, which will be difficult for the virus to bypass or compensate for. G4s in the HIV-1 promoter region may represent such a target as documented by their high evolutionary conservation across all HIV-1 subspecies[37,39]. Several studies have linked this region to the effectivity of provirus transcription as documented by the effect of point mutations that disrupt the G4 structures formed in the promoter, or by the binding of G4-stabilizing ligands[11,40,41]. Moreover, mutations in promoter regulatory elements are likely to impact viral fitness, limiting the emergence of resistant strains. Furthermore, the HIV-1 promoter sequence cannot be found in the human genome and the polymorphism of G4 structures indicates that it is possible to achieve a high degree of drug specificity. In addition, the importance of non-B DNA structures in the host genome, including G-quadruplexes, was recently demonstrated for productive and latent HIV-1 integration and the reactivation potential of latent proviruses[42]. Finally, the crucial role of RNA G-quadruplexes at the early steps of HIV-1 life cycle was unravelled when stabilization of HIV G4 structures by G4-binding ligands prevented the initiation of reverse transcription[43].

The general features of G4-binding compounds include an electron-rich aromatic core with potential for a flat conformation and an ability to participate in hydrogen bonding[44]. These requirements are perfectly fulfilled by helquats (Fig. 2). Helquat-based compounds can interact with high selectivity for ds-DNA and by changing certain structural features in the main helquat core, this class of compounds could be made selective for ss-DNA over ds-DNA[45], which makes them attractive for the development of a new class of G4 stabilizers. Here, we synthesized a group of helquat-based compounds capable of interacting with G4-forming sequences leading to their stabilization. These compounds were screened for an inhibitory effect on HIV-1 replication. Selected hits were then evaluated by several virological methods (adsorption and replication assays, a time-of-addition assay) to address the timing of their antiviral action and biophysical methods (FRET and *Taq* polymerase stop assays and light-up experiments) to examine their mode of action. Computational methods were subsequently used to estimate their binding modes to two G4 patches.

Results

An CPE-based antiviral assay in HIV-1-infected MT-4 cells was used to determine the EC_{50} of compounds.

Based on a facile $[2+2+2]$-cycloisomerization strategy, a robust three-step synthetic entry into a novel family of helical extended diquats (helquats) was developed[46]. After a straightforward Knoevenagel condensation of the respective helquat parent with different aldehydes (Fig. 2A), more than 500 helquat-based compounds were synthesized and screened for anti-HIV-1 activity in an assay based on the inhibition of the virus-induced cytopathic effect in MT-4 cells, evaluated after 5 days post-infection by an XTT assay. In parallel with the screening, the XTT assay was also used to determine the cytotoxicity of all compounds (Fig. 2B; for more information on the synthesis and characterization of the compounds, see the Supplementary Information).

The compounds were ranked by their EC_{50}, and the top 19 compounds that had their EC_{50} lower than 4 μM were selected for this study (Fig. 2). Only two compounds have their EC_{50} values below 1 μM (**1** and **2**). Out of the selected compounds, only two exhibit cytotoxicity to MT-4 cells. Compounds **2** and **3** have CC_{50} values of 24.4 μM and 15.2 μM, respectively. Because the remaining compounds do not exhibit cytotoxicity in the range tested, their CC_{50} values are described as greater than 50 μM.

Helquats inhibit different stages of the HIV-1 life cycle.

It has been reported that compounds with multiple positive charges can unspecifically block viral entry[47]. Because the compounds which we tested are protonated under physiological conditions[48], their antiviral effect could result from virion damage or reduced virion binding and entry into target cells. To investigate this possibility, we performed adsorption and replication assays.

To address virion binding and entry in a competitive setting, we performed an adsorption assay where test compounds and virions were added to MT-4 cells simultaneously and were washed out after 2 h of incubation. In parallel, we performed a replication assay where MT-4 cells were infected with HIV-1 and the tested compounds were added 1 h later. In both cases, the infection was evaluated by measuring reverse transcriptase activity in a cell-free supernatant 5 days after the infection (Fig. 3). As controls, PBS (phosphate-buffered saline) and inhibitors of different HIV-1 life-cycle steps were used: AMD (AMD3100), an inhibitor of gp120 attachment to CXCR4
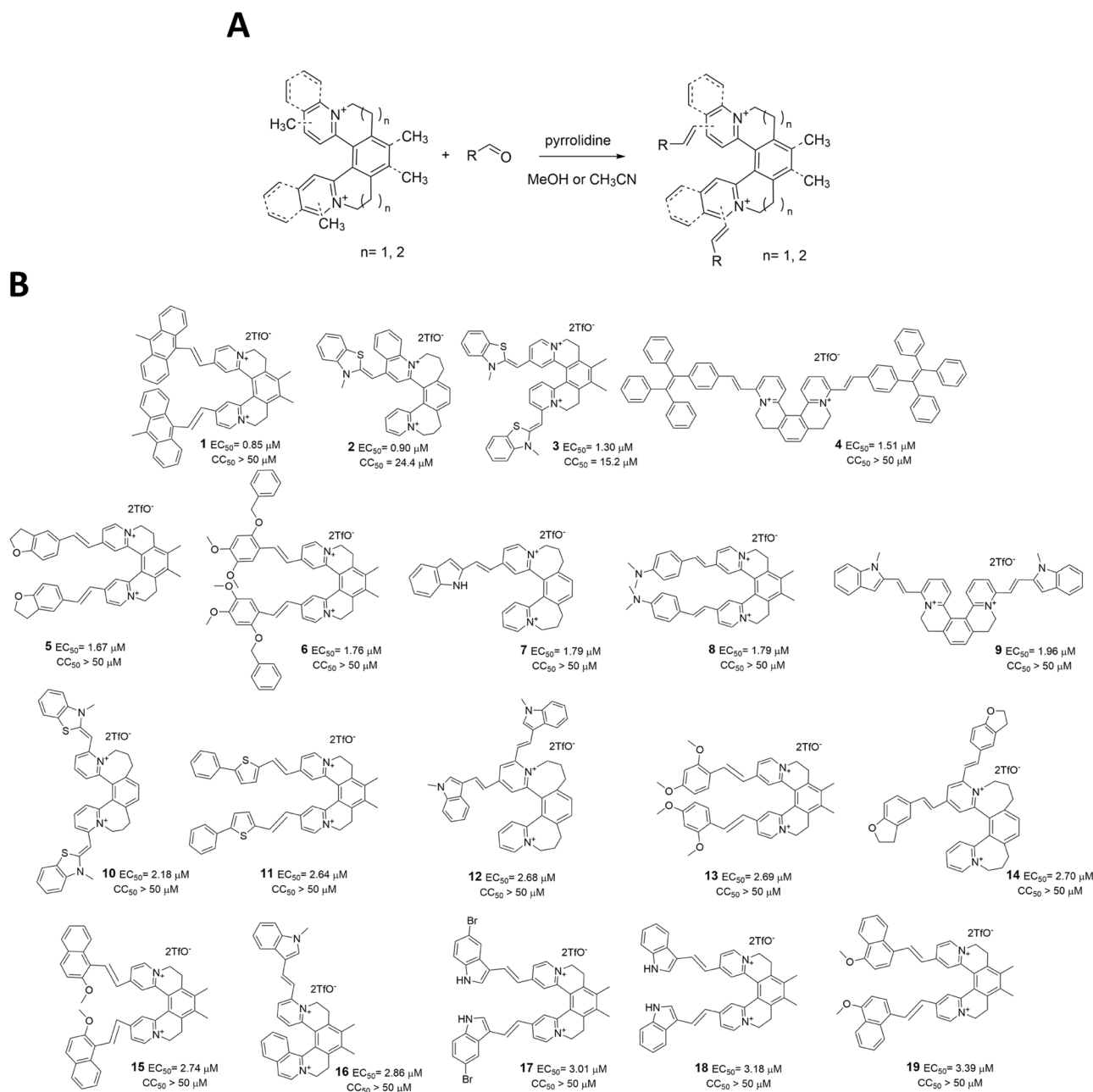
Figure 2. (**A**) Synthesis of substituted helquats using a Knoevenagel condensation. (**B**) Structures and EC_{50} and CC_{50} values of selected helquat compounds in cytopathic effect-based assay using MT-4 cells infected by HIV-1 (NL4-3 strain).

receptors of the target cell, thus preventing HIV-1 adsorption and entry; AZT (azidothymidine), an inhibitor of reverse transcriptase; and RAL (raltegravir), an integrase inhibitor.

Some compounds (namely **1**, **5**, **10**, **13**, **16**, **17**, **18** and **19**) inhibited neither adsorption nor replication, or their effect was weak, whereas other compounds (**2**, **4**, **8**, **9** and **11**) inhibited adsorption. Compounds compromising viral entry were excluded from further analyses because their effect is probably nonspecific, mainly because their cationic nature. The results of these two experiments led to the selection of compounds **3**, **6**, **7**, **12**, **14** and **15**, which had minimal or no effect on virus adsorption while significantly impairing viral replication.

Helquats regulate G-quadruplexes in HIV-1 at the level of both RNA and DNA. There are indications that G4s in the HIV-1 promoter might play roles in both RNA and DNA forms. G4s present in the LTR region of HIV-1 genomic RNA could influence reverse transcription. By contrast, G4s in the promoter region of the integrated HIV-1 provirus might regulate the transcription of viral genome and of viral mRNAs and thus the expression of viral proteins and the production of new virions - both the later stages of the HIV-1 life cycle. A modified time-of-addition assay was performed with an emphasis on the later stages of the HIV-1 life cycle, such as reverse transcription, integration and provirus expression. The tested compounds were added at several time
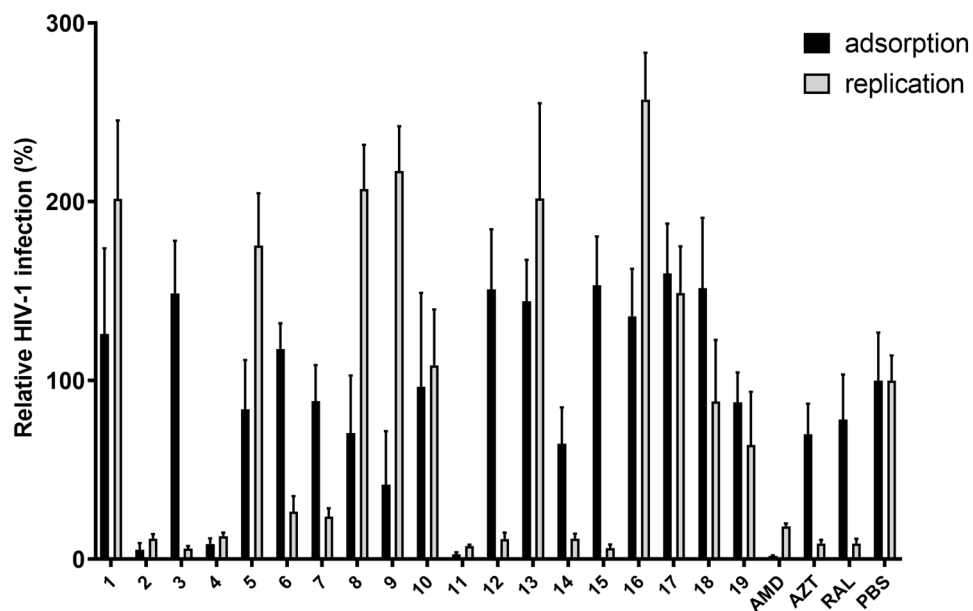
Figure 3. Adsorption and replication assays for individual compounds tested. The levels of HIV-1 infection were normalized against the phosphate-buffered saline (PBS) control in each individual experiment. AMD—AMD3100, AZT—azidothymidine and RAL—raltegravir were used as HIV-1 inhibition controls. Error bars represent standard errors from three independent experiments.

points, but the initial time point was 2 h post-infection to eliminate their virucidal/adsorption/entry effects. Several inhibitors of the individual steps of the HIV-1 life cycle were used as controls to resolve the post-entry steps of the HIV-1 life cycle. MT-4 cells were infected with NL4-3 HIV-1 and the infection was allowed to proceed for 1 h. Subsequently, the unbound virions were washed out, and the cells were resuspended in fresh media. The compounds were first added 2 h after the infection. The culture media were harvested 31 h later and the number of released virions was monitored through the activity of reverse transcriptase in the culture media (Fig. 4).

All the compounds presented in Fig. 4, except for **7**, which was excluded from further experiments due to large results variability, were active when added at early time points of the experiment (i.e. two and 3 h post-infection), but only some of them (**12** and **14**) preserved their activity at later time points. In comparison with the control compounds (AZT—inhibitor of reverse transcriptase and RAL—inhibitor of integrase), these results suggest that the effect of **3**, **6** and **15** is mainly manifested when these compounds are administered before or early during reverse transcription; therefore, these compounds are likely to act already at the RNA level. By contrast, the second group of compounds (**12** and **14**) can be added later (up to 8 h), when reverse transcription is mostly completed, indicating that these compounds might influence the G-quadruplexes also at the DNA level.

Helquats stabilize G-quadruplexes in the HIV-1 LTR promoter. Perrone et al.[11] have published that the conserved region of the HIV-1 promoter spanning the NF-κB site and Sp1 sites contains G-tracts involved in
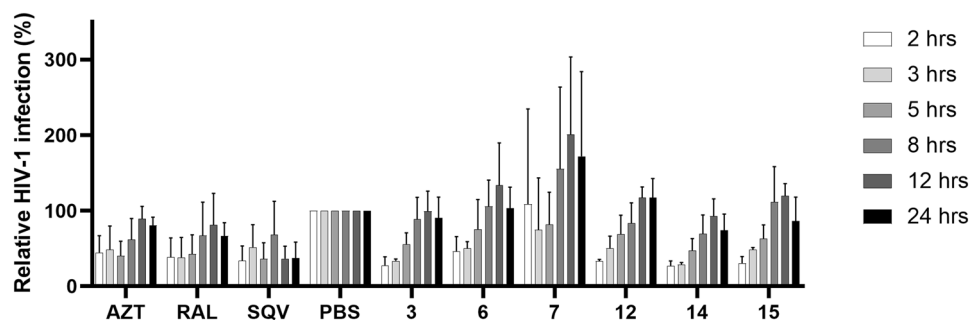

Figure 4. Modified time-of-addition assay. The panel of control compounds includes AZT (azidothymidine, an inhibitor of reverse transcriptase), RAL (raltegravir, an integrase inhibitor), SQV (saquinavir, a protease inhibitor) and PBS (phosphate-buffered saline, control). The levels of HIV-1 infection for each time point were normalized against the respective PBS control. Error bars represent standard errors from three independent experiments.

the formation of mutually exclusive G4 structures. Furthermore, they have shown the stabilization of these G4s by quadruplex-binding ligands, such as TMPyP4 or BRACO-19 and the formation of an additional G4, which is induced by these ligands. The aforementioned experiments have indicated that active compounds might modify the stability of HIV-1 LTR G4s. We have used a construct corresponding to the G-rich part of the core HIV-1 promoter (positions −105/−48 with respect to the transcription initiation site) with FAM and TAMRA moieties placed at their 5′- and 3′-ends, respectively. The stability of the G4s formed in the presence or absence of G4-binding ligands has been assessed by melting-curve experiments monitored based on the increase in FAM-fluorophore fluorescence (a representative examples of FRET melting curves is in Supplement Information Fig. S1).

The results show that all the compounds tested at the concentration of 2 μM stabilize G-quadruplexes in the HIV-1 LTR (0.2 μM concentration), as demonstrated by an increase in melting temperature (T_m), albeit to a different extent (ΔT_m of 10–28 °C, Table 1). This experiment has confirmed that the compounds tested are able to stabilize G-quadruplexes in vitro and indicates that the observed HIV-1 inhibitory effect of the tested compounds might be caused by their stabilization of G-quadruplexes in the HIV-1 LTR region.

Helquats bind mostly to the 6 and 6′ G-patches of the HIV-1 promoter.

To address specifically the positions in the HIV-1 promoter region that are affected by the binding of G4-stabilizing agents, we performed a *Taq* polymerase stop assay. The wild-type LTR oligonucleotide was extended at the 5′end to include a primer-annealing region and used as a template for a single-cycle *Taq* polymerase reaction. The elongation of the radioactively labelled primer was performed at 45 °C in the presence or absence of the tested compounds and controls. Most of the tested compounds (16 of the 19 compounds originally selected for this study) exhibit some degree of G-quadruplex stabilization at the concentration of 2.5 μM, as manifested by an increase in the intensity of the bands in the individual G-rich patches, even though the extent of the observed effect is different.

The majority of the active candidates selected from the previous experiments (Fig. 5A, the entire gels for all tested compounds can be found in Supplementary Fig. S2) mostly stabilize in the region corresponding to the G-patch 6. Compounds **3**, **6** and **15** show strong stop bands also in G-patch 6′. Some level of stabilization was observed for G-patch 5 as well, especially after incubation with **3** and **6** and partially also with **15** (Fig. 5B, quantification of G-patch regions 5, 6 and 6′). Interestingly, the tested helquats exhibit a different pattern, indicating the involvement of different bases/G-tracts in the stabilized G-quadruplexes.

To evaluate the potency of G-quadruplex-stabilizing compounds, we performed this assay at different compound concentrations (5, 2.5 and 1 μM; Supplementary Fig. S2). This experiment revealed the concentration dependency of the stabilization effect of helquats. The strongest pausing of *Taq* polymerase was observed for **3**, **6**, **15** and partially **12**, which exhibit their stabilization effect even at 1 μM concentration (Fig. 5C).

Helquats bind specifically to G-quadruplexes, not only to G-rich regions.

There are many small molecular probes for specific G4 visualization in cells[49]. Helquats have been shown to have certain optical properties that make them attractive for use as target-specific fluorescence light-up probes for the recognition of targets such as dsDNA or AT-rich DNA sequences[45]. Therefore, we speculated that the most potent helquats presented in our study might show fluorescence light-up in the presence of G-quadruplexes. To distinguish specifically the potential light-up triggered by helquat binding to G-quadruplexes in the HIV-1 LTR, we employed two control oligonucleotides, both unable to form G-quadruplexes (Fig. 6A): an LTR oligonucleotide sequence with two point mutations precluding G4 formation and an LTR oligonucleotide with a scrambled sequence (both adopted from[41]).

Among the compounds tested, only **12** exhibits strong and specific fluorescence after incubation with a quadruplex-forming oligonucleotide (Fig. 6B). The intensity of fluorescence measured after the incubation with both oligonucleotides incompetent of G-quadruplex formation was almost identical, with the maximum intensity 2.5 times lower than that of the G-quadruplex-forming oligonucleotide (wt). This indicates that **12** does not bind non-specifically to G-rich regions but requires and selectively recognizes G-quadruplex formation. To examine the specificity of ligand **12**, we have repeated the light-up experiment with single-stranded G4 LTR in parallel with double-stranded G4 LTR and additional G-quadruplexes (c-myc, c-kit1 and h-telo). We have observed that ligand **12** exhibit strong fluorescence after incubation with a quadruplex forming oligonucleotide ss G4 LTR and

Compound	T_m (°C)	ΔT_m (°C)
3	61 ± 0.1	10
6	65 ± 0.1	14
12	79 ± 0.2	28
14	64 ± 0.2	13
15	70 ± 1.0	19
PDS	93 ± 13	42
PBS	51 ± 0.1	0

Table 1. FRET analysis of G-quadruplex melting temperatures in the presence of tested compounds at a concentration of 2 μM. PDS (pyridostatin)—positive control, PBS (buffer)—negative control. The ΔT_m was calculated by subtracting the average T_m of PBS (control) from average T_m for each compound. The T_m results were averaged from three independent experiments, depicted as mean and standard error.
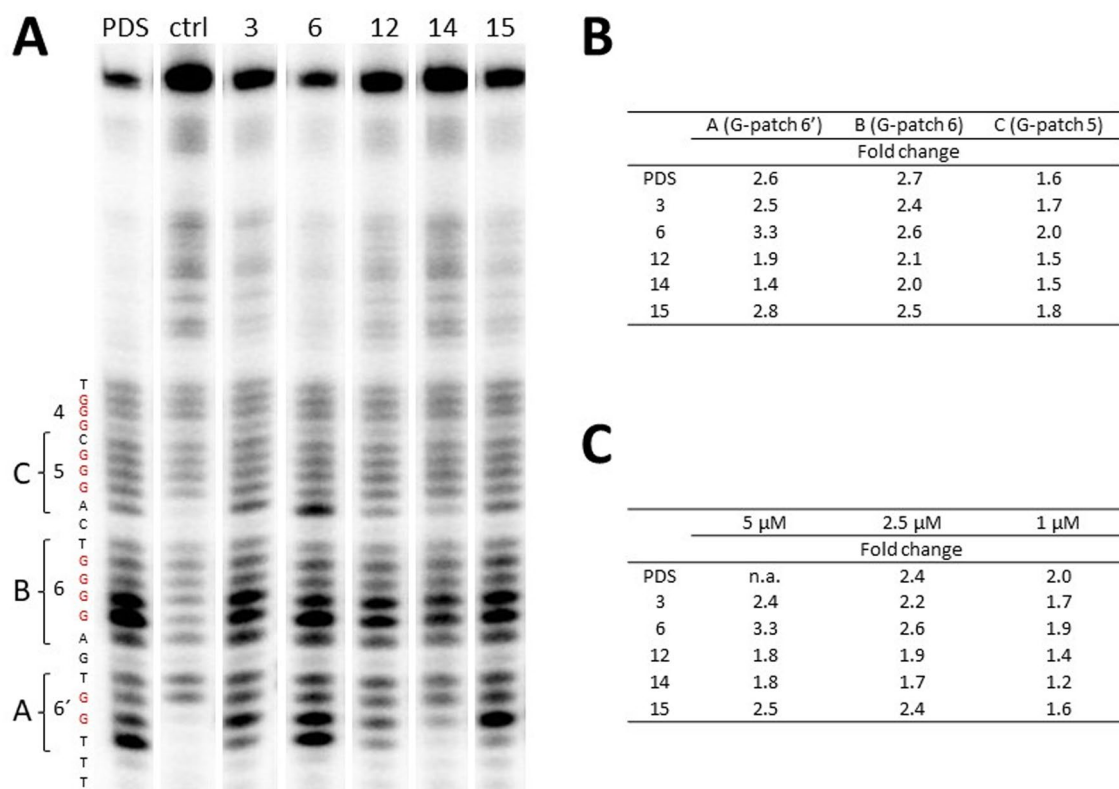
Figure 5. Identification of individual G-tracts stabilized by helquat binding. (**A**) A *Taq* polymerase stop assay in the presence of 2.5 μM of the compounds being tested, with stop regions visible as bands of higher intensity. The sequence and numbered G-tracts (highlighted in red) are indicated on the left. The brackets labelled with letters A, B and C mark the regions selected for quantification. This figure was made by the excision of selected lines from Supplementary Fig. 2C, which shows the entire original gel. (**B**) Quantification of selected regions, normalized against the negative control (ctrl). The values represent a fold change calculated as ratio between pixel intensity of individual region A (G-patch 6′), B (G-patch 6) or C (G-patch 5) and pixel intensity of negative control (ctrl) of the same regions. The automatic background subtraction was performed by Rolling Ball method in 1D gel analysis using ImageQuant TL 8.1. (**C**) Quantification of all regions (A, B and C) together at different concentrations (see Supplementary Fig. S2 for gels) normalized against the control. The values represent a fold change calculated as ratio between pixel intensity of all regions (A, B, C) at different concentrations of compounds and pixel intensity of negative control (ctrl) of all regions (A, B, C). The same background subtraction method was used as in (**B**). n.a.—not analysed, PDS (pyridostatin)—positive control.
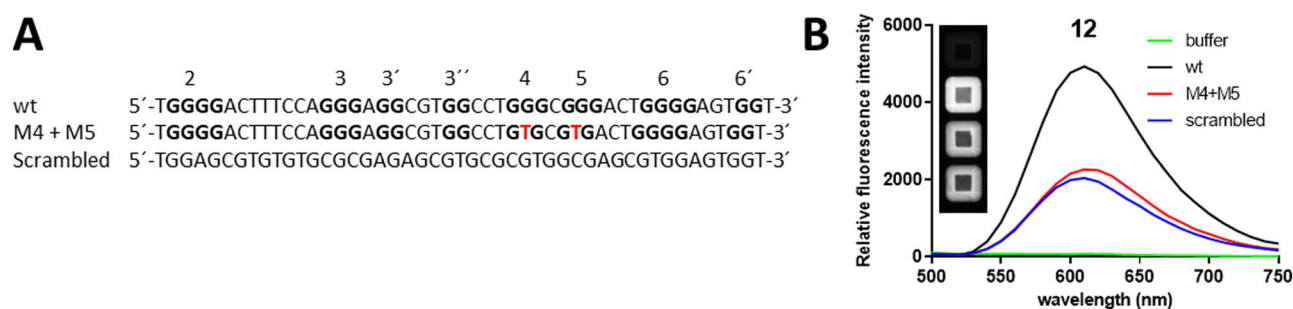

Figure 6. Light-up experiment. (**A**) Sequences of oligonucleotides used in the light-up experiment with the G bases involved in the formation of G-quadruplexes in bold and the G-tracts numbered. The abbreviation 'wt' indicates the LTR oligonucleotide with a wild-type sequence capable of folding into G-quadruplex structures. 'M4 + M5' indicates the LTR oligonucleotide with two point mutations (highlighted in red) that is unable to form G-quadruplexes. 'Scrambled' is the labelling of the LTR oligonucleotide with the identical base composition as the wild-type oligonucleotide, but with a scrambled sequence. (**B**) Emission spectra of **12** after incubation with different oligonucleotides or the reaction buffer alone after excitation at 280 nm. The inset represents a fluorescent image of the wells with compound **12** and a given oligonucleotide after UV-light excitation taken with the Bio-Rad ChemiDoc imaging system.

only very weak fluorescence after incubation with ds G4 LTR (Supplementary Fig. S3A). The incubation of ligand **12** with other G-quadruplexes showed strong fluorescence in the case of c-myc and c-kit1 oligonucleotides and low fluorescence in the case of h-telo oligonucleotide (Supplement Fig. S3A). This suggests that ligand **12** is not specific to HIV LTR quadruplexes but can bind also to other G-quadruplexes. To further evaluate the specificity of ligand **12**, we performed the light-up titration experiment incubating compound **12** with serial dilutions (1:1) of oligonucleotides (ss G4 LTR wild type, G4 c-myc and c-kit1) spanning final concentrations 2 to 0.002 μM (Supplement Fig. S3B). Here we observed that at oligonucleotides concentration of 0.5 μM and below, compound **12** exhibits stronger fluorescence after incubation with ss G4 LTR than after incubation with c-myc and c-kit1. This can indicate a slight preference of ligand **12** for ss G4 LTR than for c-myc at lower oligonucleotide concentrations.

Different classes of compounds have distinct modes of binding to G-quadruplexes.

It has already been shown that the promoter region of HIV-1 can form alternative-configuration G4s with different G-patches involved or even induced by G4-binding ligands[11]. The structures of two predominant HIV-1 LTR G4s (LTR-III, comprising G-patches 3–6, and LTR-IV, comprising G-patches 4–6′) have been characterized earlier by NMR spectroscopy[38,50] and are shown in Fig. 7A. Both G4s are mutually exclusive and show highly distinctive folding and features. Based on the structure of the helquat 'core', all the tested compounds can be categorized into two groups: HQ[6,6] dyes (**3**, **6** and **15**) and HQ[7,7] dyes (**12** and **14**). In accordance with previously described results, we selected one representative of each group, namely **6** and **12** (because they are not cytotoxic and also have a different binding patterns to HIV-1 LTR G-patches), and used computational methods to provide a rationale for their interaction with LTR-III and LTR-IV, respectively.

Both enantiomers (M- and P-) of **6** and **12** have been docked into ten conformations of the available NMR structures of HIV-1 LTR-III and LTR-IV[38,50]. For the assessment of their stability, we ran 1 μs molecular dynamics (MD) in explicit solvent. Most of the poses of **6** and **12** docked to LTR-III and LTR-IV were not stable, unbound and rebound on either side of the G-quartet (Supplementary Fig. S4), where they stayed for hundreds of ns. Those poses that were stable in MD bound at the junction of the G-quartet and the stem-loop of LTR-III and stacked on the bottom of the G-quartet of LTR-IV (Fig. 7B). These calculations indicate that the structures and the flexibilities of the individual G4s strongly affect the preferred binding mode of the helquats.

We have observed larger flexibility of the loops in the case of LTR-III (average RMSD of DNA of 6.8 Å) than in LTR-IV (3.6 Å). This behaviour is inherent in the structures of the LTRs (average RMSDs of 6.3 Å and 3.5 Å, respectively, from unliganded LTR MD). It should be noted that the flexibility is attributed to the loops whereas the G-quadruplexes are stable.

The calculations of interaction energy using MM-GBSA have predicted the order of affinity (from the strongest to the weakest) as LTR-III/6 > LTR-III/12 > LTR-IV/12 > LTR-IV/6. This indicates that: (i) both dyes prefer the LTR-III conformation of G4 over the LTR-IV conformation, (ii) compound **6** binds to LTR-IV much more weakly than compound **12**, and (iii) the affinity differences between the M- and P-stereoisomers are larger for LTR-III (7.6 and 8.2 kcal/mol for **6** and **12**, respectively, in favour of the M isomer) than for LTR-IV (0.6 and 5.3 kcal/mol in favour of 6/M and 12/P, respectively). These calculations indicate that the structure of the helquat core strongly affects the binding mode to the individual G4s.
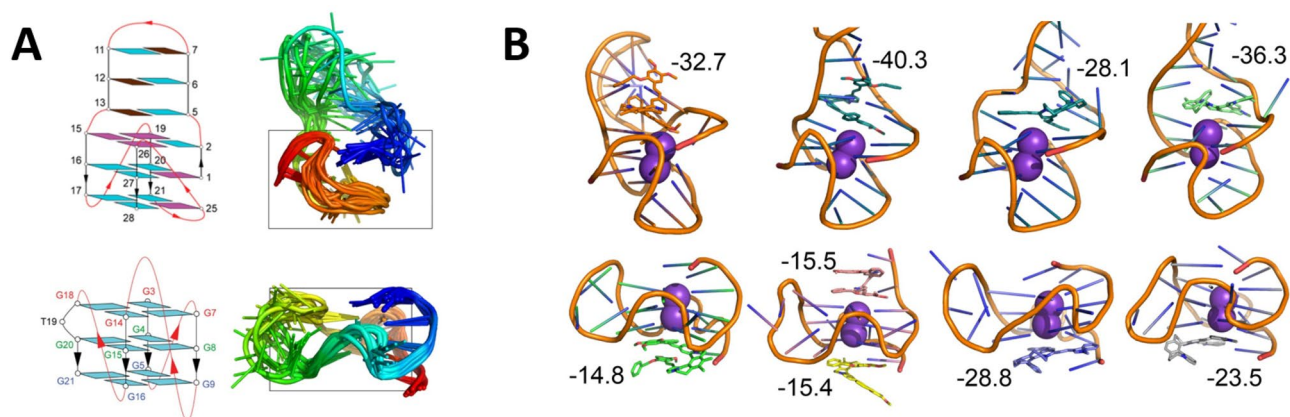

Figure 7.. (**A**) Structures of HIV-1 LTR-III (top; PDB: 6H1K)[38] and LTR-IV (bottom; PDB: 2N4Y)[50] G4s as determined by NMR. The structures are coloured in a rainbow pattern from the 5′ terminus (blue) to the 3′ terminus (red). (**B**) Computationally predicted most stable poses of **6** and **12** (in sticks) in LTR-III (top row) and LTR-IV (bottom row). G4s are shown as a cartoon, two K+ ions as magenta spheres. The compounds/enantiomers are arranged from left to right: **6**/P, **6**/M, **12**/P, **12**/M. MM-GBSA binding free energies are shown in kcal/mol. For the enantiomer **6**/M, binding to LTR-IV, the iso-energetic poses are shown to be stacked on the top and bottom of the G-quadruplex.

Discussion

More than 500 helquat-based compounds were synthesized and subsequently screened for their anti-HIV-1 activity. This activity was a factor for the selection of 19 compounds with EC_{50} < 4 μM, whose mechanism of action was then evaluated. After the exclusion of compounds that targeted virus attachment or entry, the timing of the action was determined for the remaining candidates. It was revealed that in order for some compounds (**3**, **6** and **15**) to be active, they had to be administered within 5 h after infection, which corresponded to the time frame of the initial phase of reverse transcription. This implies that they might act via binding to G4s at the RNA level and interfere with the course of reverse transcription. A similar activity time frame was previously published for the G4 ligand BRACO-19, which inhibited the reverse transcription process at the template level[40]. G4s have multiple roles during HIV-1 reverse transcription. It is known that the HIV-1 genome consists of two single-stranded RNA molecules and that both RNA strands are required for the reverse transcription to be successfully completed. HIV-1 RNA dimerises directly without the need for a protein co-factor, and it has been described that this interaction is driven by the formation of an interstrand G4[51]. Furthermore, it is known that the G4 regions of the HIV-1 RNA genome form bi-molecular G4 structures corresponding to HIV-1 recombination hotspots[36].

Other tested compounds (**12** and **14**) could be administered later (up to 8 h post-infection) and still retain some level of antiviral activity, which provides a sufficient time frame for the completion of reverse transcription. Consequently, they might express their activity also at the DNA level, either before or after integration, but this does not exclude the possibility for them to act at the RNA level as well. It should be noted that the reference compounds in the TOA assay are added at 100-fold their EC_{50} values, but our compounds did not exhibit sufficient selectivity (with the selectivity index ranging between 11.7 and 29.9). Therefore, it is remarkable that we observed such a clear effect even at concentrations only five times greater than their EC_{50} values. Based on the antiviral activity of the G4-binding ligands and its timing, it can be assumed that during viral infection, the G4 structure is formed both in the RNA genome and in the promoter region of the HIV-1 provirus and that it, presumably, has important biological roles in both environments.

The G-cluster in the HIV-1 promotor has previously been shown to be a fine modulator of HIV-1 transcription. Several studies have shown that LTR G4s act as repressor elements of viral transcription, the effect being augmented by G4 ligands[36,40,41]. G-patches 3″–6′ are conserved across HIV-1 sequences, with slight differences in the number of G residues (two to four), in length of linkers (two to five nucleotides) and in their base composition. Interestingly, the common feature shared by all of these sequences is their ability to form stable G4 structures[37]. This observation further supports the assumption that the G4 structure is an important element of the HIV-1 promoter region that is maintained by a strong evolutionary pressure[39].

It has already been shown that the promoter region of HIV-1 can form alternative G4s with different G-patches involved and that additional conformations of G4 can even be induced by G4-binding ligands[11] or proteins[52]. The structures of two predominant HIV-1 LTR G4s (LTR-III, comprising G-patches 3–6, and LTR-IV, consisting of G-patches 4–6′) were characterized by NMR spectroscopy[38,50]. The two G4s are mutually exclusive and show very distinctive folding (Fig. 7). Furthermore, their effect on HIV-1 transcription is quite the opposite: The stabilization of LTR-III inhibits viral transcription and thus might promote latency[11] whereas the formation of LTR-IV increases transcriptional activity and might result in provirus reactivation[50]. These observations have led to the assumption that the balance of LTR-III and LTR-IV may act as a fine regulator of HIV-1 promoter activity.

Indeed, we have observed different stop patterns after the treatment of the LTR template with the compounds studied. All compounds belonging to the HQ[6,6] group of dyes (**3**, **6** and **15**) have shown a similar pattern with strong stop bands at G-patches 6 and, 6′ and weaker *Taq* polymerase pausing also before G-patch 5. By contrast, the representatives of HQ[7,7] group of dyes (**12** and **14**) have forced stop bands only in G-patch 6.

We tried to link these results to LTR structures, dynamics and compound binding. Using molecular dynamics, we confirmed the structural stability of both LTR-III and LTR-IV G4 and the greater flexibility of the loops of the former. We obtained stable binding modes of two selected representatives (**6** from the HQ[6,6] group of dyes and **12** from the HQ[7,7] group of dyes) in both stereoisomeric forms. They were characterized by stacking on one side of the G4s coupled with interactions with the loops. These are typical modes of binding to G4s and have been observed for many ligands by X-ray and NMR methods[53]. The predicted affinities, such as binding free energies, were calculated by the MM-GBSA method as the sum of interaction and deformation free energies (see the "Materials and methods"). These numbers should be interpreted with great caution due to inaccuracies of nucleic-acid and ligand force fields, incomplete sampling and approximations in MD (e.g. the lack of polarization) as well as MM-GBSA (e.g. implicit-solvent) models. Therefore, not even the related, more advanced MM-PBSA method can quantitatively describe G4 conformations from 2.5-μs-long MD simulations[54]. Moreover, entropy effects beyond the scope of the current study, such as the effect of two isoenergetic binding sites on G4 or M/P stereoisomer interconversion in solution, can further contribute to the binding of **6**. These challenges in G4-specific ligand design warrant a multidisciplinary approach involving biology, chemistry and computations.

The regulatory mechanisms of G4s involve not only steric hindrance to transcription and/or translation, but also the binding of the protein factors that modulate G4 conformation and serve as a scaffold for the recruitment of additional protein regulators. Cellular proteins regulate viral latency and effective transcription by inducing or unfolding HIV-1 LTR G4s[41,52,55]. The G4-forming sequence in HIV-1 LTR overlaps with three Sp1 as well as two NFκB binding sites and is crucial for transcription initiation[11]. Sp1 has been shown to bind not only the putative 5′-GGGCGG-3′ sequence in double-stranded DNA, but also the G4 secondary structure[56], the latter of which it even binds to with a stronger affinity[57]. Sp1 has an ambiguous role in the HIV-1 life cycle: It is a strong activator of provirus expression, but it is also associated with maintaining virus latency[58]. The intriguing fact that HIV-1 LTR possesses two modes of Sp1 binding (G4 structure and consensus sequence) opens new questions about G4 folding and subsequently the role of Sp1 in regulating provirus transcription and HIV-1 latency. Such diversity indicates that HIV-1 uses the G4-forming ability of its promoter region as a complex switching

mechanism involving subtle Sp1 interaction differences to fine-tune the transcriptional output. This carefully balanced checkpoint of the HIV-1 life cycle can be exploited as a new target in antiretroviral therapy. In this study, we present a group of chemically similar compounds where subtle modifications of the helquat core lead to a different mode of binding to the G4 region of the HIV-1 promoter. Having strong positive charge these compounds represent poor candidates for antiviral agents at this stage, but they might still serve as lead compounds for further rationale-based drug design. Our experiments show that the presented helquat dyes selectively bind G4s but not only to HIV-1 LTR G4s but can bind also to other G4s e.g. c-myc (as documented by the light-up experiment with **12**). There is a great structural variability of G quadruplexes, differing in the composition and length of the loops and stems[59,60]. It is therefore difficult, if not impossible, to find structural reasons why any ligand, such as helquat would not bind to other types of G quadruplexes. On the other hand, absolute specificity to HIV-1 G4 to achieve antiviral activity might not be necessary[19]. Each infected cells produces many viral genomes, for example the estimated burst size (number of virions released from infected cell) for HIV is between 1000 and 3000 virions per infected cells. Thus we can assume that HIV G4s exceeds the host cells G4s.

Finally, these helquat dyes are also able to distinguish even the distinct G4 conformations (as documented by the different affinities of **6**, and to a lesser extent even of **12**, to LTR-III and LTR-IV quadruplexes). Such diversity could be exploited to design specific ligands able to bind selectively different G4s formed in the HIV-1 LTR and thus specifically regulate the HIV-1 life cycle. Furthermore, these compounds might be a useful tool for determining how and when G4s regulate HIV-1 transcription, which would enhance our understanding of HIV-1 latency and reactivation.

Materials and methods

Oligonucleotides, drugs, plasmids and cells. All oligonucleotides were purchased from Sigma-Aldrich (Supplementary Table S1). The following reagents were obtained through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH: saquinavir, AMD3100 and raltegravir, MT-4 cells (CD4+ T-cell line) from Dr. Douglas Richman; pNL4-3 plasmid from Dr. Malcolm Martin. HEK293T cells (human embryonic kidney cell line transformed with large T antigen) were obtained from the ATCC (Manassas, VA). MT-4 cells were maintained in RPMI1640 medium with L-glutamine supplemented with 10% fetal bovine serum (FBS), 100 U/ml penicillin, 100 µg/ml streptomycin and 10 mM HEPES (all Sigma-Aldrich). HEK293T cells were maintained in Dulbecco's Modified Eagle's Medium (DMEM) with L-glutamine, 10% FBS, 100 U/ml penicillin and 100 µg/ml streptomycin (all Sigma-Aldrich). Cells were grown in a humidified incubator maintained at 37 °C with 5% CO_2. HEK293T cells were transfected with pNL4-3 to produce NL4-3 HIV-1, as described previously[61].

Synthesis and characterization of the compounds tested. The methodology of organic synthesis, together with the structures, MS analysis, NMR characterization and evaluation of the anti-HIV-1 activity of the tested compounds, is described in the Supplementary Information.

Adsorption and replication assays. MT-4 cells were seeded in 80 µl of the Roswell Park Memorial Institute (RPMI) medium in the amount of 2×10^5 cells per well of a 96-well plate, and 10 µl of the virus stock were immediately added with the final multiplicity of infection (MOI) of 0.05. For a replication assay, the infected cells were incubated at 37 °C for 1 h before the addition of compounds being tested (10 µM final concentration) and subsequently cultivated at 37 °C for another 5 days. For an adsorption test, by contrast, the compounds (10 µM final concentration) were added simultaneously with the virus, and after 2 h of incubation, the cells were washed three times with PBS to remove the compounds and unbound virions; after the addition of 100 µl of the RPMI medium, the cells were cultivated at 37 °C for 5 days. Five days post-infection, aliquots of virus-containing cell-free supernatants were harvested and analysed by reverse transcriptase (RT) assay (described in the Supplementary Information). The experiments were performed as biological replicates where the sample size for each condition was n = 3.

Time-of-addition assay. MT-4 cells were infected with NL4-3 HIV-1 at a MOI of 1. After 1 h, the cells were washed, seeded (100,000 cells/well) and incubated at 37 °C. The tested compounds were added at 2, 3, 5, 8, 12 and 24 h post-infection at 10 µM final concentration, while the reference compounds were added at a concentration corresponding to 100-fold their EC_{50}. HIV-1 production was determined 31 h post-infection by measuring RT activity in the supernatant.

FRET melting assay. The G4 LTR FRET oligonucleotide was diluted to 0.4 µM in LiCaco buffer (10 mM lithium cacodylate at pH 7.2 and 100 mM KCl), heat-denatured at 95 °C for 5 min, and folded into a G4 structure at room temperature for 4 h. After incubation, the LiCaco buffer was added alone or with the compounds being tested (4 µM in LiCaco buffer and 1% DMSO). The final concentrations for compound screening were: 0.2 µM oligonucleotide, 2 µM tested compound, 100 mM KCl and 0.5% DMSO. Fluorescence melting curves were determined using a Realplex⁴ Mastercycler (Eppendorf). After the first equilibration step at 25 °C for 5 min, a stepwise increase of temperature with a resolution of 0.1 °C was performed to reach 95 °C, with FAM emission monitored in each step. T_m was either calculated directly by the Realplex⁴ Mastercycler, or raw data were analysed in GraphPad Prism 7, where the asymmetric (five-parameter) model was used to obtain melting temperatures T_m (°C) for the further calculation of ΔT_m. The ΔT_m was calculated by subtracting the average T_m of PBS (control) from the average T_m for each compound. Three independent experiments were performed.

***Taq* polymerase stop assay.**   A *Taq* polymerase stop assay was carried out as described in a previous study[11]. Briefly, the primer (G4 LTR primer, Supplementary Table S1) was 5′-end labelled with [γ-^{32}P]ATP using T4 polynucleotide kinase at 37 °C for 30 min. The T4 polynucleotide kinase was then inactivated by incubation at 65 °C for 5 min. The radioactively labelled primer was column-purified and stored at −20 °C for further use. The labelled primer (1.7 pmol) was annealed to the template (2.1 pmol) in the LiCaco buffer and extended with the Go *Taq* DNA polymerase (2U/reaction, Promega) at 45 °C for 45 min. Samples were incubated with G-quadruplex-ligands at 45 °C for 45 min before DNA polymerase addition. The reactions were stopped by a denaturing solution (80% formamide, 10% glycerol, 0.025% xylen cyanol, 0.025% bromphenol blue and 10 mM EDTA pH 8); extension products were separated on 15% denaturing gel and visualized by phosphorimaging. The pixel intensity of bands was calculated using 1D gel analysis in ImageQuant TL software version 8.1 (GE HealthCare) after automatic background subtraction by Rolling Ball method.

Light-up experiments.   The selected compounds were incubated in the presence of the G4 LTR wild-type (wt) oligonucleotide, the G4 LTR M4 + M5 oligonucleotide containing mutations that disabled G-quadruplex formation (described previously[41]), and the G4 LTR scrambled-sequence oligonucleotide, or the buffer-only control. The compounds were tested at a compound:oligonucleotide ratio of 2:1 (4 μM compounds and 2 μM oligonucleotide) in PBS in a 384-well plate. Emission spectra were measured after excitation at 280 nm in a Tecan Infinite M1000 plate reader. The fluorescent image of the plate was taken with the Bio-Rad ChemiDoc imaging system after UV-light excitation. The Light-up titration experiment was performed similarly as described above with the exception that 4 μM compound 12 was incubated with serial dilutions (1:1) of oligonucleotides (ss G4 LTR wild type, G4 c-myc and G4 c-kit1) spanning final concentrations 2 to 0.002 μM. The excitation at 280 nm and emission at 600 nm were used to monitor the Light-up.

Computational modelling.   Both P- and M-enantiomers of compounds 6 and 12 were docked to all 10 + 10 conformations of the available NMR structures of G4s (LTR-III: 6H1K, LTR-IV: 2N4Y) using Glide SP[62]. Based on visual inspection and the docking score, 49 of the resulting complexes were selected for further evaluation. To assess the stability of the binding modes, 1 μs molecular dynamics (MD) simulations were run in an explicit solvent for the complexes under study, free ligands and G4s. The ligands were parametrized by the antechamber module of AMBER (ambertools20, forcefield: GAFF2 [Amber 2020, University of California, San Francisco] with RESP charges (HF/6-31G*) calculated using the Gaussian 16 package (Gaussian 16, revision A.03, Gaussian, Inc., Wallingford CT, 2016). The systems were prepared in Leap (DNA forcefield: OL15, water model: TIP3P), neutralized with Na$^+$, and the NaCl concentration was adjusted to 0.15 M. Hydrogen mass repartitioning was applied to make it possible to use a time step of 4 fs. The systems were simulated using the GPU-accelerated pmemd code in AMBER20 in the following steps: 1000 cycles of minimization, 50 ps of heating to 300 K with the solute restrained (force constant: 2.0 kcal/mol/Å^2), 200 ps of NpT equilibration with the solute restrained (force constant: 2.0 kcal/mol/Å^2), 500 ps of NpT equilibration with the solute restrained (force constant: 0.1 kcal/mol/Å^2) and 1 μs of production without restraints, using a time step of 4 fs. The stability of the 500 frames obtained was analysed using pytraj, and the stable poses were subjected to binding free energy calculations using the MM-GBSA approach[63] by MMPBSA.py in ambertools20 with default parameters except for ionic strength, which was set to 0.1 M (istrng = 0.1). The individual terms were calculated according to Eqs. (1–4).

$$\text{binding energy} \ = \ E\big(\text{complex}\big) \ - \ E_{min}(\text{DNA, free}) \ - \ E_{min}\big(\text{ligand, free}\big) \tag{1}$$

$$\text{interaction energy} \ = \ E\big(\text{complex}\big) \ - \ E\big(\text{DNA, cplx}\big) \ - \ E\big(\text{ligand, cplx}\big) \tag{2}$$

$$\text{deformation energy (DNA)} = \ E\big(\text{DNA, cplx}\big) \ - \ E_{min}(\text{DNA, free}) \tag{3}$$

$$\text{deformation energy} \big(\text{ligand}\big) \ = \ E\big(\text{ligand, cplx}\big) \ - \ E_{min}\big(\text{ligand, free}\big) \tag{4}$$

Data availability
All data generated and analysed during this study are included in this published article (and its Supplementary Information files).

References

1. Lago, S. *et al.* Promoter G-quadruplexes and transcription factors cooperate to shape the cell type-specific transcriptome. *Nat. Commun.* **12**, 3885. https://doi.org/10.1038/s41467-021-24198-2 (2021).
2. Haeusler, A. R. *et al.* C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature* **507**, 195–200. https://doi.org/10.1038/nature13124 (2014).
3. Sket, P. *et al.* Characterization of DNA G-quadruplex species forming from C9ORF72 G4C2-expanded repeats associated with amyotrophic lateral sclerosis and frontotemporal lobar degeneration. *Neurobiol. Aging* **36**, 1091–1096. https://doi.org/10.1016/j.neurobiolaging.2014.09.012 (2015).
4. Varshney, D., Spiegel, J., Zyner, K., Tannahill, D. & Balasubramanian, S. The regulation and functions of DNA and RNA G-quadruplexes. *Nat. Rev. Mol. Cell Biol.* **21**, 459–474. https://doi.org/10.1038/s41580-020-0236-x (2020).

5. Perrone, R. *et al.* Mapping and characterization of G-quadruplexes in *Mycobacterium tuberculosis* gene promoter regions. *Sci. Rep.* **7**, 5743. https://doi.org/10.1038/s41598-017-05867-z (2017).

6. Zuffo, M. *et al.* Carbohydrate-naphthalene diimide conjugates as potential antiparasitic drugs: Synthesis, evaluation and structure-activity studies. *Eur. J. Med. Chem.* **163**, 54–66. https://doi.org/10.1016/j.ejmech.2018.11.043 (2019).

7. Yadav, P. *et al.* G-quadruplex structures in bacteria: Biological relevance and potential as an antimicrobial target. *J. Bacteriol.* **203**, e0057720. https://doi.org/10.1128/JB.00577-20 (2021).

8. Lavezzo, E. *et al.* G-quadruplex forming sequences in the genome of all known human viruses: A comprehensive guide. *PLoS Comput. Biol.* **14**, e1006675. https://doi.org/10.1371/journal.pcbi.1006675 (2018).

9. Metifiot, M., Amrane, S., Litvak, S. & Andreola, M. L. G-quadruplexes in viruses: Function and potential therapeutic applications. *Nucleic Acids Res.* **42**, 12352–12366. https://doi.org/10.1093/nar/gku999 (2014).

10. Murat, P. *et al.* G-quadruplexes regulate Epstein–Barr virus-encoded nuclear antigen 1 mRNA translation. *Nat. Chem. Biol.* **10**, 358–364. https://doi.org/10.1038/nchembio.1479 (2014).

11. Perrone, R. *et al.* A dynamic G-quadruplex region regulates the HIV-1 long terminal repeat promoter. *J. Med. Chem.* **56**, 6521–6530. https://doi.org/10.1021/jm400914r (2013).

12. Tan, J. *et al.* The SARS-unique domain (SUD) of SARS coronavirus contains two macrodomains that bind G-quadruplexes. *PLoS Pathog.* **5**, e1000428. https://doi.org/10.1371/journal.ppat.1000428 (2009).

13. Kong, J. N. *et al.* Identification and characterization of G-quadruplex formation within the EP0 promoter of pseudorabies virus. *Sci. Rep.* **8**, 14029. https://doi.org/10.1038/s41598-018-32222-7 (2018).

14. Zahin, M. *et al.* Identification of G-quadruplex forming sequences in three manatee papillomaviruses. *PLoS ONE* **13**, e0195625. https://doi.org/10.1371/journal.pone.0195625 (2018).

15. Frasson, I., Solda, P., Nadai, M., Lago, S. & Richter, S. N. Parallel G-quadruplexes recruit the HSV-1 transcription factor ICP4 to promote viral transcription in herpes virus-infected human cells. *Commun. Biol.* **4**, 510. https://doi.org/10.1038/s42003-021-02035-y (2021).

16. Ruggiero, E. & Richter, S. N. Viral G-quadruplexes: New frontiers in virus pathogenesis and antiviral therapy. *Annu. Rep. Med. Chem.* **54**, 101–131. https://doi.org/10.1016/bs.armc.2020.04.001 (2020).

17. Abiri, A. *et al.* Unlocking G-quadruplexes as antiviral targets. *Pharmacol. Rev.* **73**, 897–923. https://doi.org/10.1124/pharmrev.120.000230 (2021).

18. Ruggiero, E., Zanin, I., Terreri, M. & Richter, S. N. G-quadruplex targeting in the fight against viruses: An update. *Int. J. Mol. Sci.* https://doi.org/10.3390/ijms222010984 (2021).

19. Ruggiero, E. & Richter, S. N. Targeting G-quadruplexes to achieve antiviral activity. *Bioorg. Med. Chem. Lett.* **79**, 129085. https://doi.org/10.1016/j.bmcl.2022.129085 (2023).

20. Balasubramanian, S. & Neidle, S. G-quadruplex nucleic acids as therapeutic targets. *Curr. Opin. Chem. Biol.* **13**, 345–353. https://doi.org/10.1016/j.cbpa.2009.04.637 (2009).

21. Balasubramanian, S., Hurley, L. H. & Neidle, S. Targeting G-quadruplexes in gene promoters: A novel anticancer strategy?. *Nat. Rev. Drug Discov.* **10**, 261–275. https://doi.org/10.1038/nrd3428 (2011).

22. Lopergolo, A. *et al.* Targeting of RET oncogene by naphthalene diimide-mediated gene promoter G-quadruplex stabilization exerts anti-tumor activity in oncogene-addicted human medullary thyroid cancer. *Oncotarget* **7**, 49649–49663. https://doi.org/10.18632/oncotarget.10105 (2016).

23. Savva, L. & Georgiades, S. N. Recent developments in small-molecule ligands of medicinal relevance for harnessing the anticancer potential of G-quadruplexes. *Molecules* https://doi.org/10.3390/molecules26040841 (2021).

24. Kosiol, N., Juranek, S., Brossart, P., Heine, A. & Paeschke, K. G-quadruplexes: A promising target for cancer therapy. *Mol. Cancer* **20**, 40. https://doi.org/10.1186/s12943-021-01328-4 (2021).

25. Artusi, S. *et al.* The Herpes Simplex Virus-1 genome contains multiple clusters of repeated G-quadruplex: Implications for the antiviral activity of a G-quadruplex ligand. *Antiviral Res.* **118**, 123–131. https://doi.org/10.1016/j.antiviral.2015.03.016 (2015).

26. Callegaro, S. *et al.* A core extended naphtalene diimide G-quadruplex ligand potently inhibits herpes simplex virus 1 replication. *Sci. Rep.* **7**, 2341. https://doi.org/10.1038/s41598-017-02667-3 (2017).

27. Perrone, R. *et al.* 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. https://doi.org/10.1021/acs.jmedchem.5b01283 (2015).

28. Majee, P. *et al.* Inhibition of Zika virus replication by G-quadruplex-binding ligands. *Mol. Ther. Nucleic Acids* **23**, 691–701. https://doi.org/10.1016/j.omtn.2020.12.030 (2021).

29. Zou, M. *et al.* G-quadruplex binder pyridostatin as an effective multi-target ZIKV inhibitor. *Int. J. Biol. Macromol.* **190**, 178–188. https://doi.org/10.1016/j.ijbiomac.2021.08.121 (2021).

30. Lavigne, M. *et al.* SARS-CoV-2 Nsp3 unique domain SUD interacts with guanine quadruplexes and G4-ligands inhibit this interaction. *Nucleic Acids Res.* **49**, 7695–7712. https://doi.org/10.1093/nar/gkab571 (2021).

31. Carvalho, J. *et al.* Human papillomavirus G-rich regions as potential antiviral drug targets. *Nucleic Acid Ther.* **31**, 68–81. https://doi.org/10.1089/nat.2020.0869 (2021).

32. Artusi, S. *et al.* Antiviral activity of the G-quadruplex ligand TMPyP4 against herpes simplex virus-1. *Viruses* https://doi.org/10.3390/v13020196 (2021).

33. Lv, L., Cui, H., Chen, Z., Zhou, Y. & Zhang, L. G-quadruplex ligands inhibit chikungunya virus replication. *J. Med. Virol.* **94**, 2519–2527. https://doi.org/10.1002/jmv.27622 (2022).

34. Bohalova, N. *et al.* Analyses of viral genomes for G-quadruplex forming sequences reveal their correlation with the type of infection. *Biochimie* **186**, 13–27. https://doi.org/10.1016/j.biochi.2021.03.017 (2021).

35. Harpster, C., Boyle, E., Musier-Forsyth, K. & Kankia, B. HIV-1 genomic RNA U3 region forms a stable quadruplex-hairpin structure. *Biophys. Chem.* **272**, 106567. https://doi.org/10.1016/j.bpc.2021.106567 (2021).

36. Piekna-Przybylska, D., Sullivan, M. A., Sharma, G. & Bambara, R. A. U3 region in the HIV-1 genome adopts a G-quadruplex structure in its RNA and DNA sequence. *Biochemistry* **53**, 2581–2593. https://doi.org/10.1021/bi4016692 (2014).

37. Amrane, S. *et al.* 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. https://doi.org/10.1021/ja501500c (2014).

38. Butovskaya, E., Heddi, B., Bakalar, B., Richter, S. N. & Phan, A. T. 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. https://doi.org/10.1021/jacs.8b05332 (2018).

39. Ruggiero, E., Tassinari, M., Perrone, R., Nadai, M. & Richter, S. N. Stable and conserved G-quadruplexes in the long terminal repeat promoter of retroviruses. *ACS Infect. Dis.* **5**, 1150–1159. https://doi.org/10.1021/acsinfecdis.9b00011 (2019).

40. Perrone, R. *et al.* Anti-HIV-1 activity of the G-quadruplex ligand BRACO-19. *J. Antimicrob. Chemother.* **69**, 3248–3258. https://doi.org/10.1093/jac/dku280 (2014).

41. Tosoni, E. *et al.* Nucleolin stabilizes G-quadruplex structures folded by the LTR promoter and silences HIV-1 viral transcription. *Nucleic Acids Res.* **43**, 8884–8897. https://doi.org/10.1093/nar/gkv897 (2015).

42. Ajoge, H. O. *et al.* G-quadruplex DNA and other non-canonical B-form DNA motifs influence productive and latent HIV-1 integration and reactivation potential. *Viruses* https://doi.org/10.3390/v14112494 (2022).

43. Amrane, S. *et al.* Deciphering RNA G-quadruplex function during the early steps of HIV-1 infection. *Nucleic Acids Res.* **50**, 12328–12343. https://doi.org/10.1093/nar/gkac1030 (2022).

44. Spiegel, J., Adhikari, S. & Balasubramanian, S. The structure and function of DNA G-quadruplexes. *Trends Chem.* **2**, 123–136. https://doi.org/10.1016/j.trechm.2019.07.002 (2020).
45. Kužmová, E. *et al.* *HeliDye1: Helquat Fluorogenic Probe Specific for AT-rich DNA Duplexes* Preprint at https://doi.org/10.21203/rs.3.rs-737372/v2 (2022).
46. Reyes-Gutierrez, P. E. *et al.* Functional helquats: Helical cationic dyes with marked, switchable chiroptical properties in the visible region. *Chem. Commun.* **51**, 1583–1586. https://doi.org/10.1039/c4cc08967g (2015).
47. Wilkinson, R. A. *et al.* Novel compounds containing multiple guanide groups that bind the HIV coreceptor CXCR4. *Antimicrob. Agents Chemother.* **55**, 255–263. https://doi.org/10.1128/AAC.00709-10 (2011).
48. Adriaenssens, L. *et al.* Helquats: A facile, modular, scalable route to novel helical dications. *Chem.-Eur. J.* **15**, 1072–1076. https://doi.org/10.1002/chem.200801904 (2009).
49. Chilka, P., Desai, N. & Datta, B. Small molecule fluorescent probes for G-quadruplex visualization as potential cancer theranostic agents. *Molecules* https://doi.org/10.3390/molecules24040752 (2019).
50. De Nicola, B. *et al.* 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. https://doi.org/10.1093/nar/gkw432 (2016).
51. Sundquist, W. I. & Heaphy, S. Evidence for interstrand quadruplex formation in the dimerization of human immunodeficiency virus 1 genomic RNA. *Proc. Natl. Acad. Sci. USA* **90**, 3393–3397 (1993).
52. Ruggiero, E. *et al.* Fused in liposarcoma protein, a new player in the regulation of HIV-1 transcription, binds to known and newly identified LTR G-quadruplexes. *ACS Infect. Dis.* **8**, 958–968. https://doi.org/10.1021/acsinfecdis.1c00508 (2022).
53. Haider, S. M., Neidle, S. & Parkinson, G. N. A structural analysis of G-quadruplex/ligand interactions. *Biochimie* **93**, 1239–1251. https://doi.org/10.1016/j.biochi.2011.05.012 (2011).
54. Islam, B., Stadlbauer, P., Neidle, S., Haider, S. & Sponer, J. Can we execute reliable MM-PBSA free energy computations of relative stabilities of different guanine quadruplex folds?. *J. Phys. Chem. B* **120**, 2899–2912. https://doi.org/10.1021/acs.jpcb.6b01059 (2016).
55. Scalabrin, M. *et al.* The cellular protein hnRNP A2/B1 enhances HIV-1 transcription by unfolding LTR promoter G-quadruplexes. *Sci. Rep.* **7**, 45244. https://doi.org/10.1038/srep45244 (2017).
56. Raiber, E.-A., Kranaster, R., Lam, E., Nikan, M. & Balasubramanian, S. A non-canonical DNA structure is a binding motif for the transcription factor SP1 in vitro. *Nucleic Acids Res.* **40**, 1499–1508. https://doi.org/10.1093/nar/gkr882 (2012).
57. Kim, N. The interplay between G-quadruplex and transcription. *Curr. Med. Chem.* **26**, 2898–2917. https://doi.org/10.2174/0929867325666171229132619 (2019).
58. Jiang, G., Espeseth, A., Hazuda, D. J. & Margolis, D. M. 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. https://doi.org/10.1128/JVI.01208-07 (2007).
59. Bedrat, A., Lacroix, L. & Mergny, J. L. Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.* **44**, 1746–1759. https://doi.org/10.1093/nar/gkw006 (2016).
60. Lightfoot, H. L., Hagen, T., Tatum, N. J. & Hall, J. The diverse structural landscape of quadruplexes. *FEBS Lett.* **593**, 2083–2102. https://doi.org/10.1002/1873-3468.13547 (2019).
61. Hodek, J. *et al.* Protective hybrid coating containing silver, copper and zinc cations effective against human immunodeficiency virus and other enveloped viruses. *BMC Microbiol.* **16**(Suppl 1), 56. https://doi.org/10.1186/s12866-016-0675-x (2016).
62. Friesner, R. A. *et al.* Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **47**, 1739–1749. https://doi.org/10.1021/jm0306430 (2004).
63. Miller, B. R. III. *et al.* MMPBSA.py: An efficient program for end-state free energy calculations. *J. Chem. Theory Comput.* **8**, 3314–3321. https://doi.org/10.1021/ct300418h (2012).

Acknowledgements

Author contributions

M.P., P.E.R.-G., J.K., J.D., Y.Y., M.L., and F.T. designed and performed research; M.P. and J.W. designed research and analyzed data; J.D., Y.Y., and M.L. performed computational modelling; M.P. wrote the original draft; J.W. supervised the project, reviewed, and edited the manuscript. All authors reviewed and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-33263-3.

Correspondence and requests for materials should be addressed to J.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.