

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/260993851>

Topology of a DNA G-Quadruplex Structure Formed in the HIV-1 Promoter: A Potential Target for Anti-HIV Drug Development

ARTICLE *in* JOURNAL OF THE AMERICAN CHEMICAL SOCIETY · MARCH 2014

Impact Factor: 12.11 · DOI: 10.1021/ja501500c · Source: PubMed

CITATIONS

16

READS

29

6 AUTHORS, INCLUDING:



Marie-line L Andreola

French National Centre for Scientific Research

76 PUBLICATIONS 1,677 CITATIONS

SEE PROFILE

**G-quadruplexes in viruses: function, targeting and
therapeutic applications**

Journal:	<i>Nucleic Acids Research</i>
Manuscript ID:	Draft
Manuscript Type:	5 Surveys and Summaries Manuscript
Key Words:	G-quadruplex, viruses, function, targeting, therapeutic applications

SCHOLARONE™
Manuscripts

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

G-quadruplexes in viruses: function, targeting and therapeutic applications #

Mathieu Métifiot¹, Samir Amrane², Simon Litvak^{1*} and Marie-Line Andreola^{1*}

¹CNRS UMR-5234, Université de Bordeaux, 146 Rue Léo Saignat, 33076 Bordeaux, France;

²INSERM, U869, IECB, ARNA laboratory, Université de Bordeaux, 2 Rue Robert Escarpit 33600 Pessac, France.

*Corresponding authors: marie-aline.andreola@u-bordeaux.fr, simon.litvak@reger.u-bordeaux2.fr

This survey is dedicated to the memory of Dr. Wilhelm Guschlbauer, a dear friend, distinguished biophysicist and pioneer in the field of G-quadruplexes, who passed away on December 1st, 2012.

Abstract

G-rich DNA or RNA sequences can form non-canonical G-quadruplex structures in which four guanines fold in a planar arrangement through Hoogsteen hydrogen bonds. Although many biochemical and structural studies have focused on DNA sequences containing successive, adjacent guanines that spontaneously fold into G-quadruplex structures *in vitro*, evidence for their *in vivo* relevance has recently begun to accumulate. Complete sequencing of the human genome highlighted the presence of approximately 300,000 sequences that can potentially form G4s. Likewise, the presence of putative G4 sequences has been reported in the genomes of various viruses (*e.g.*, HIV-1, EBV, and papillomavirus). Many studies have focused on telomeric G-quadruplexes and how their dynamics are regulated to enable telomere synthesis. Moreover, a role for G4s has also been proposed in cellular and viral replication, recombination and gene expression control. In parallel, DNA aptamers that form G4s have been described as inhibitors and diagnostic tools to detect viruses (*e.g.*, hepatitis A, Epstein-Barr, cauliflower mosaic virus, SARS, and SV40). In our survey, special emphasis will be given to the possible role of these structures in a virus life cycle as well as the use of G4-forming oligonucleotides as potential antiviral agents and innovative tools.

From G-quartets to G-quadruplexes: discovery and topology

Almost a century ago, the ability of guanosine, but not guanine, to form viscous gels was described (1). Fifty years later, X-ray diffraction data clearly showed that the guanosine moieties in these gels were arranged in a tetrameric organization linked by eight Hoogsteen hydrogen bonds (Figure 1) (2,3). These hydrogen bonds differ from the bonds observed in canonical Watson-Crick pairing and are involved the interaction of the N7 group from one guanine with the exocyclic amino group from a neighboring base (Figure 1a). Therefore, a G-tetrad or a G-quartet results from planar association between four guanines that are held together by eight hydrogen bonds and coordinated with a central Na⁺ or K⁺ cation (4-8). In addition, nucleoside derivatives were also used to confirm the structural properties of G-quartets (9-14).

Conversely, a G-quadruplex is formed by nucleic acid sequences (DNA or RNA) containing G-tracts or G-blocks (adjacent runs of guanines) and composed of various numbers of guanines. Depending on the nucleotide sequence, the way G-quadruplexes can be formed presents a high degree of diversity. The core of a G4 is based on stacking between two or more G-tetrads, wherein the guanines can adopt either a *syn* or an *anti* glycosidic bond angle conformation. Consequently, each of the four G-tracts that form the core of the structure can run in the same or opposite direction with respect to its two neighbors, forming parallel, anti-parallel or hybrid core conformations. Depending on these orientations, the G-blocks delimit four negatively charged grooves of different sizes: narrow, medium or wide (Figure 1b-e). For intra-molecular structures (Figure 1b and 1c), the four G-tracts belong to the same oligonucleotide and are attached by linkers with variable nucleotide sequences and lengths. These loops can adopt three different conformations: lateral, diagonal or propeller (Figure 1b-d). The bi- or tetra-molecular G4 structures (Figure 1d-e) are assembled from G-tracts belonging to two or four different strands. The G-blocks can also be interrupted by one to seven non-G nucleotides, which result in bulges that protrude from the G4 core (Figure 1e). In contrast to the almost monomorphic canonical duplex, these variable structural parameters are directly related to the nucleotide

primary sequence. This unique family of globular-shaped nucleic acid structures (Figure 1f-j) presents a high level of plasticity that enables various applications (see section “*G4s as antiviral agents*”).

However, a potential physiological role for G4s has remained controversial for many years. Nonetheless, interest in G-quadruplexes has increased over time, which indicates a potential biological role for these structures (15). Since then, thousands of reports and reviews have been published on several aspects of G4s, including the biophysical and chemical characteristics as well as biological function in prokaryotes and eukaryotes (5,16-23).

Additionally, G4s present an intrinsic resistance to “regular” nucleases. However, specific nucleases have been isolated, first in yeast with KEM1/SEP1 (24) and later in humans with GQN1 (25). Later, a DNA binding protein (G4R1/RHAU) that can unfold RNA/DNA quadruplexes was isolated in human cells (26,27), which further highlights the biological relevance of G4s. Finally, development of G4 specific probes such as monoclonal antibodies (28-30) and small chemical ligands (31-36) facilitated *in vivo* studies that support quadruplex formation in cells. Taken together, G4s are likely present in several significant genomic regions and may be a key component in important cellular processes (37,38).

Biological roles of G4s in the genome

The sequencing of the human genome highlighted the presence of many sequences enriched in guanines that can potentially form quadruplexes [for a review, see (39)]. In theory, more than 350,000 putative G-tetrads may be formed with loops of 1–7 nucleotides, and over 700,000, with loops of up to 12 nucleotides. However, the genomic DNA is dynamic, and quadruplexes, duplexes or other structural forms at those sites will be influenced by chromatin and other DNA-binding proteins. Thus, G4 formation depends on the cell type, cell cycle and environmental conditions (stress).

The telomeres. Telomeres are found at the end of chromosomes, and their length has been linked to the lifespan of organisms. Initially, studies reported that the ciliates *Oxytricha*, *Stylonychia*, and *Euplotes* have G-strand overhangs of either 16 or 14 bases at the telomeric termini (40). Later, a similar arrangement was found in the telomeres from the ciliate *Tetrahymena thermophila* and slime mold *Didymium iridis*. Taken together, these data led to the proposal that chromosomal termini include an unusual folded back structure comprising the G-strand overhang (Figure 2a) (41). Accumulated experimental data indicate the presence of G4s in telomeric DNA (6,42-44). Finally, DNA2 is a helicase/nuclease that cleaves G4s and regulates telomere integrity (45,46). This latter observation clearly establishes that G-quadruplexes form in the telomere and are crucial structures for regulating telomere maintenance, which provides a mechanism for controlling cell proliferation (47,48).

Telomerase is a ribonucleoprotein enzyme that adds TTAGGG repeats to the 3' end of DNA strands in the telomere regions. Human telomerase is a heterodimeric complex composed of reverse transcriptase (TERT), RNA (TERC), and dyskerin (DKC1, a pseudouridine synthase). TERT elongates single-stranded DNA using TERC, which is a single-stranded RNA molecule, as a template. Because TERC contains a H/ACA motif (49), dyskerin binds TERC and stabilizes the telomerase complex (50). At the 3' ends of the telomeres, if this single-stranded DNA is longer than four TTAGGG repeats (>23 nucleotides), it can fold over itself and form secondary structures, including G-quadruplexes. Those structures prevent telomere elongation by the telomerase complex (51). Accordingly, G4-binding proteins and small G4-ligands (e.g., telomestatin) show anti-proliferative and potential anti-tumor activities through telomere interference (31,52-54). However, the exact mechanism is more complicated because telomestatin and derivatives are also active independent of telomerase, most likely through targeting the G4 involved in tumor growth at another site in the genome (55-57).

DNA replication. DNA replication is highly regulated, which ensures faithful reproduction of genetic information through each cell cycle. In eukaryotes, this process is initiated at thousands of DNA replication origins (ORIs) distributed along each chromosome (58). Initially, a pre-replicative complex (Pre-RC) is assembled during late mitosis and the early G1 phase; it evolves during the S phase to a pre-initiation complex through initiator protein activity (*e.g.*, cyclin-dependent kinases, cdc7, cdc45, and Mcm proteins). This step involves unwinding the origin and DNA polymerase recruitment. A tissue-specific temporal system exists, and only a few origins undergo “firing” (replication fork elongation). The remaining origins are dormant backups used under stress (*e.g.*, DNA damage and collisions with transcription machinery, Figure 2c-d). Studies on organisms ranging from drosophila to humans have highlighted that repeated elements (OGREs) are enriched with G residues (59). *In vitro*, OGRE sequences form G4s, and point mutations that affect the stability of these G4s also impair origin function. Even if a 200 bp *cis*-regulatory element is necessary for efficient initiation, G-quadruplex structure formation at ORIs might be the key to selecting the firing origins (60,61).

Gene expression. Altered gene expression levels are implicated in many human diseases. G4s have been found in the promoter region of many genes and implicated in transcription and protein translation regulation [for a review, see (62)]. Particularly interesting, the P1 region, which is upstream of the c-myc promoter, is highly sensitive to DNase I and S1 nucleases (nuclease-hypersensitive element or NHE). This purine-rich sequence can form an intramolecular G4 (63-66), the structure for which has been determined through NMR (67). This NHEIII₁ region controls up to 90% of gene transcription activation, and the G-quadruplex acts as a transcriptional repressor element (68,69). Because interactions between DNA-binding proteins and local DNA supercoiling can impact the equilibrium between duplex/single-stranded DNA (active transcription) and quadruplex DNA

(silent), c-myc expression and, ultimately, cell proliferation may be modulated. Nucleolin (a nucleolar phosphoprotein) and ADAR1 (a Z-DNA binding / RNA editing protein) both bind the c-myc promoter *in vivo* and stabilize the G4 structure of the promoter (70,71).

Due to the high therapeutic potential of genomic G4s, studies have also been performed on promoters of other genes implicated in cancer biology. For example, KRAS is one of the most frequently mutated genes in human cancer. Its promoter contains a G-rich nuclease hypersensitive element that is critical for transcription (72). The polypurine strand forms G-quadruplex structures that bind crucial DNA repair proteins (*e.g.*, PARP-1, Ku70) (73). Direct evidence for G-quadruplex formation was also reported for the proximal promoter region of the RET proto-oncogene, and targeting this region with a small molecule represses RET proto-oncogene transcription (74).

Although many G-rich sequences have been identified in the genome, they enable generation of corresponding G-rich RNAs when the sequences are located in transcription units. Ultimately, these RNAs can fold into quadruplex structures and impact gene expression (75). In 5'-untranslated regions (5'-UTRs), G-quadruplexes can act as translation repressors (76-78). G4 structures have also been proposed to regulate translation driven by IRES domains in the absence of an mRNA cap [for a recent review, see (79) and the references therein]. Finally, it has recently been shown that G4 RNAs can stimulate ribosomal frame-shifts *in vitro* and in cultured cells when they are in coding regions (80).

Genome stability. A major challenge for future studies is to better understand genome instability, cancer and severe neurological/neuromuscular defects origins. Repeated sequences, such as polynucleotide repeats (3 and above), mini- and mega-satellites, are source of genome instability through either expansion or contraction (81). However, DNA secondary structures associated with those tandem repeats may be crucial elements in the expansion phenomenon. For progressive myoclonus epilepsy type-1 (EPM1), expansion of the dodecamer sequence (CGCG₄CG₄) within the cystatin B

promoter is thought to depend on G-quadruplex formation (82).

In addition to this potential deleterious consequence, G quadruplexes have also been implicated in development of specific immune responses. In antigen-activated B cells, class switch recombination promotes deletion of a DNA fragment (several kilobases) that joins a new constant region to the variable immunoglobulin chain (83). These switch regions (S regions) are intronic, G-rich, repetitive sequences that form G-loops upon transcription. *In vivo*, class switch recombination depends on the activity of AID (a cytidine deaminase) and MutS α (MSH2/MSH6, and conserved DNA repair factors involved in mismatch repair). AID promotes creation of U:G mismatches, and MutS α binds with high affinity the G4 DNA formed upon transcription of the S regions.

In parallel, pathogens have also developed counter-measures to evade the immune system. The role of G-quartets in this evasion is related to antigenic variation. For *Plasmodium falciparum*, G-rich sequences have been found in the upstream region of group B *var* genes that can form *in vitro* stable G-tetrads (84). Similarly, a G-quadruplex-forming sequence (5'-G₃TG₃TTG₃TG₃) is located upstream of the *pilE* expression locus in *Neisseria gonorrhoeae* (85). Formation of parallel G4s is necessary for initiating this pilin antigenic variation through recombination, and interactions with RecA could facilitate this specialized reaction. Overall, immune evasion and immunoglobulin gene class switch recombination in vertebrates are clearly analogous mechanistically (86).

The role of G4s in the virus life cycle

Human immunodeficiency virus (HIV). Searches for G-quadruplexes and elucidation of their function in the viruses' genome have mainly focused on the human immunodeficiency virus (HIV), which causes acquired immunodeficiency syndrome (AIDS). Currently, approximately 35 million people are infected with HIV worldwide. With over 2 million new infections and approximately 1.6 million deaths from AIDS per year, the pandemic continues to spread. Even without a vaccine, development of highly active anti-retroviral therapies have allowed people to live with HIV as a chronic disease (87). However, viruses within T cells remain fully capable of replicating and infecting other cells if

the drug pressure is removed or when resistance emerges. Thus, new drugs must be developed to overcome the treatment's genetic barrier.

HIV-1 is an RNA virus in the *Lentivirus* genus and is part of the *Retroviridae* family. Lentiviruses are single-stranded, positive-sense, enveloped RNA viruses. HIV-1 particles contain two molecules of genomic RNA that are converted into double-stranded DNA by the viral reverse transcriptase (RT). The resulting viral DNA is then imported into the nucleus for insertion into the cellular DNA due to virally encoded integrase (IN). Once integrated, transcription from the viral promoter at the 5' long terminal repeat generates mRNAs that code various viral proteins and genomic RNA. Alternatively, the provirus may become latent, which allows the virus and its host cell to avoid detection by the immune system. The presence of G-quadruplex structures has been highlighted at both RNA and DNA levels with implications throughout the viral life cycle.

HIV-1 RNA dimerization and recombination. Retroviral RNAs dimerize in the cytoplasm of an infected cell allowing two copies of the genome to be encapsidated in the newly produced virion (88). While a single copy of the genome is sufficient for viral replication, the second copy is also used during reverse transcription, and the viral RT switches multiple times between the two RNA molecules (89,90). The strand transfers are partially responsible for the viral variability through production of recombinant molecules. Therefore, understanding the mechanisms that drive dimerization and recombination is essential.

Dimerization is a two-step process that involves sequences upstream of the splice donor site (91,92). The sequences involved in initial dimerization and encapsidation partially overlap at the 5' end of the viral genome. The first sequence is a highly conserved dimer initiation site (DIS) that forms a stem loop. A concentration-dependent kissing-loop interaction is initiated from contacts between consecutive guanines (93); the interaction then spreads to the stems. However, this interaction does not seem sufficiently strong to keep the two copies together during reverse transcription.

Several studies have identified a G-rich sequence that form bi-molecular G-quadruplex structures in the *gag* region of the HIV-1 genome, near the DIS (94-96). Localization of these G4-forming sequences correlates with recombination hot spots and exhibits an increased rate of template switching that highlights a potential role for these structures (97). Supporting this hypothesis, recombination in the U3 domain is cation-dependent and is lower in the presence of Li^+ , which is a metal ion that destabilizes G-quadruplexes.

Short RNA templates from the central region of the HIV-1 genome contain G-rich sequences near the central poly-purine tract (cPPT) at the 3' end of the *pol* gene (IN coding sequence); this is a region where one of the two primers used for synthesizing the (-) strand DNA is produced during reverse transcription. These sequences can form both intramolecular and dimeric G4 structures. Moreover, reconstituted systems have confirmed that G4 structures near the cPPT facilitate strand transfer and promote template switching by the RT (98). Interestingly, certain elements from the cPPT region are involved in forming the cPPT flap, which is a region that plays an important role in nuclear entry of the double-stranded DNA (99,100). Taken together, the G-rich regions located at the 5' end of the genome and in the central region are likely maintained in proximity through inter-RNA G4 formation with a crucial role in HIV-1 replication.

HIV-1 nucleocapsid: G4 chaperone properties. Retroviral nucleocapsid proteins (NCp) are multifunctional elements encoded in the *gag* gene. Notably, NCp participate in many retroviral cycle steps by remodeling nucleic acid structures to favor thermodynamically stable conformations. They are referred to as nucleic acid chaperones and interact with nucleic acid phosphodiester backbones through electrostatic interactions thanks to basic residues (especially residues in the N-terminus) [for reviews, see (101,102)]. Moreover, HIV-1 NCp (NCp7) exhibits sequence-specific binding to runs of Gs, UGs or TGs through interactions involving its two zinc fingers (CCHC motifs, which are separated by a proline-rich linker). NCp7 promotes and stabilizes G-quadruplex structures through an

unknown mechanism (103,104). A recent biophysical study using high-speed atomic force microscopy (HS-AFM) addressed direct and real-time investigations on the molecular chaperone activity at the single-molecule level (105). NCp7 can efficiently promote bimolecular G4 formation and is able to anneal the G4 structures. The G4 structure is induced by both unprocessed NCp15 and mature NCp7, which indicates that both proteins may participate in genome recognition, recombination, dimerization and packaging. NCp7 could act through a potential mechanism that involves synaptic G4 intermediates, as illustrated in Figure 3 (106).

Regulation of HIV-1 promoter activity. Viral DNA is transcribed by cellular RNA polymerase II from the viral promoter, which is located in the 5'-LTR of the proviral genome. The U3 region (Figure 4a) contains a G-rich sequence 50 nucleotides upstream from the transcription-starting site (TSS) and close to the TATA box (Figure 4b). This sequence overlaps with the so-called minimum promoter, which is composed of three SP1 as well as two NF-kB binding sites and is crucial for transcription initiation. The presence of eight blocks of guanines suggests that this region is a good candidate for G4 formation.

Two independent biophysical studies based on DMS-mediated foot-printing assays or nuclear magnetic resonance (NMR) recently evaluated the ability of this 45-nt G-rich sequence to form G4 structures (107,108). DNA fragments spanning this region (corresponding to two or three SP1 or NF-kB binding sites with oligonucleotides ranging from 19 to 32 bases) were all able to form stable parallel and anti-parallel G4 topologies with melting temperatures ranging from 43°C to 56°C in 100 mM KCl solutions. The three models are formed with rather long sequences (up to 32 nucleotides), allowing the formation of long loops of 5 to 12 nucleotides (figure 4 d, e, f). Interestingly, a hybrid structure was also observed in the NMR study with an anti-parallel G4 core composed of only two tetrads and an additional Watson Crick base pair that stacks on top of the upper tetrad (Figure 4g). Notably, these four topologies are mutually exclusive, and forming one of these G4s in the promoter

will prevent formation of the three alternative conformations. Thus, the equilibrium between these forms may play a role in regulating promoter activity.

Recently, the interaction between the SP1 protein and a fragment of the HIV-1 promoter sequence folded into a G-quadruplex was studied (98). Piekna-Przybylska *et al* used an affinity-based selection approach using biotinylated G4s immobilized on streptavidin-coated magnetic beads. The pull down experiments followed by Western blotting revealed that the SP1 protein can bind the HIV-1 promoter sequence when it adopts a G4 conformation. Perrone *et al* analyzed the effect of point mutations that disrupt the G4 structures formed in the promoter (107). The wild type (WT) and mutated LTR promoters were cloned upstream of a firefly gene in a promoter-free plasmid. In HEK 293T cells, the promoter activity of the mutated sequences (unable to form G4) was twice as high as the WT LTR. Furthermore, G-quadruplex ligands, such as BRACO-19, inhibited HIV-1 replication potentially by targeting the LTR promoter. Therefore, G4 structures might be critical for HIV-1 fitness and represent novel targets for antiviral drug development. A better understanding of the role for G-quadruplexes in regulating the HIV-1 promoter activity would also shed light on HIV-1 latency and reactivation mechanisms, from which a new landscape may emerge in clinical research to eradicate HIV from reservoirs (109).

G4s and HIV proteins expression. A G-rich sequence composed of three conserved clusters has been identified in the reading frame of the negative regulatory factor (Nef) (110). This coding sequence is located at the 3'-end of the genome and overlaps with the 3'-LTR (Figure 4a). Isolated nef G-stretches can form G-quadruplexes *in vitro*. Moreover, their stabilization represses Nef expression and decreases viral replication. Nef has multiple roles during HIV infection and has been implicated in immune evasion by promoting CD4 and CMH molecule down-regulation. Importantly, the absence of Nef seems to correlate with low viral load and inhibition of disease progression. Thus, targeting the G-quadruplexes located in the Nef coding sequence is a new attractive therapeutic

opportunity.

Simian virus 40 (SV40). SV40 is a polyomavirus with a 5 kb, closed circular and double-stranded DNA genome that codes for 6 proteins and includes a non-coding regulatory region (NCRR). This latter regulatory region contains the replication origin (ORI) and the encapsidation sequence (*ses*) but also controls the transcription direction (early *versus* late transcription). Notably, six GC boxes (GGGCGG) are present in this region, which can form an unusual quadruplex structure containing a C-tetrad stacked between G-tetrads *in vitro* (111). These repeated sequences are binding motifs for SP1 and, therefore, play an important role in early transcription. On the other hand, SV40 genome replication requires the large T-antigen (TAg), which is a multifunctional protein that binds the ORI, has ATP-dependent helicase activity and interacts with cellular proteins such as p53 and Rb (112). Interestingly, TAg can unwind G4 DNA structures (113,114); thus, it might play a crucial role in regulating replication as well as early and late transcription. Perylene di-imide derivatives (PDI) stabilize G-quadruplexe structures and inhibit both the G-quadruplex and TAg duplex DNA helicase activities. Hence, PDI provide tools for probing the role of the G-quadruplex helicase activity in SV40 replication (114,115) and introduces new insights into the link between helicases and tumorigenesis (116) or other human genetic diseases (117).

Human papillomaviruses (HPVs). The HPV family consists of more than 120 viruses; approximately half are sexually transmissible, and 15 are considered high risk due to their carcinogenic properties. These viruses are one of the most common sexually transmitted infections and induce cervical cancer. Notably, HPV-16 and HPV-18 contribute to 70% of cervical cancer induced by HPV infections. The HPV genome consists of 8 kb, circular, double-stranded DNA, and integration into the infected cell induces dramatic genome instability (118). The open reading frames encode six ‘early’ proteins (E1, E2, E4, E5, E6, and E7) and two ‘late’ structural proteins (L1 and L2). The late

1
2
3 proteins are major and minor capsids, respectively, and association with 72 L1/L2 heterodimers
4
5 forms star-shaped capsomeres. However, L1 can self-assemble into 55-60 nanometer, virus-like
6
7 particles, which can be used in prophylactic HPV vaccines to protect against an initial HPV infection.
8
9 G-rich regions have been found in HPV genomes, and their potential to fold into a G-quadruplex
10
11 structure has been described (119). G-rich loci that fulfill the criteria for G-quadruplex formation
12
13 have only been found in 8 types of HPV. However, a strong argument for the relevance of G4s in
14
15 HPV biology is that the viral protein E1 is a helicase that resembles SV40 TAg (112). Consequently,
16
17 E1 may also present a G4 unwinding activity. For HPV52 and HPV58, potential G-quadruplex-
18
19 forming sequences are located in the long control region (LCR), which is a regulatory sequence
20
21 composed of nearly 1 kb, suggesting a potential role in transcription and replication. The presence of
22
23 G4s in the sequence coding for the L2 protein (HPV57), E1 (HPV32, HPV42), and E4 (HPV3, HPV9,
24
25 HPV25) suggests that G-quadruplex formation may also alter alternative splicing necessary for
26
27 producing viral proteins from the overlapping ORFs. Targeting these G-quadruplexes could
28
29 potentially serve as a basis for novel antiviral therapies.
30
31
32
33
34
35

36 Epstein-Barr virus (EBV). Also referred to as human herpesvirus 4 (HHV-4), EBV is a virus of the
37
38 herpes family that can induce infectious mononucleosis, which includes particular cancers, such as
39
40 Hodgkin's lymphoma, Burkitt's lymphoma and nasopharyngeal carcinoma (120). In the case of HIV-
41
42 1 co-infection, EBV is also associated with hairy leukoplakia and central nervous system lymphoma.
43
44 The virus is approximately 120 nm to 180 nm in diameter and is composed of a 172 kb DNA double
45
46 helix that circularizes upon entry into the nucleus and becomes a viral episome. Human herpes
47
48 viruses are mostly asymptomatic due to latency.
49
50

51
52 The Epstein-Barr virus encodes a genome maintenance protein (nuclear antigen 1, EBNA1),
53
54 expressed in all EBV-associated malignancies. EBNA1 binds G-rich sequences at the viral
55
56 replication origin, recruits the replication complex and is involved in metaphase chromosome
57
58
59
60

attachment, which insures maintenance throughout mitosis (121). Thus, formation of secondary structures, such as G-quadruplexes, may play a role in regulating EBV replication.

The level of EBNA1 synthesized is tightly controlled; it is sufficiently high to maintain viral infection but sufficiently low to avoid immune recognition by the host's virus-specific T cells. Regulation occurs during translation due to secondary structures present in the mRNA (122). Destabilization of the G-quadruplex-forming sequence through antisense oligonucleotide annealing increases the translation rate and, consequently, promotes antigen presentation. On the other hand, stabilization of the G4 with a G4 ligand (pyridostatin) decreases EBNA1 synthesis and allows immune evasion. Interestingly, G-quadruplexes are similarly observed in mRNA for other maintenance genes in the gammaherpesvirus family. These findings suggest that this mode of translational regulation may be more general among proteins that self-regulate synthesis. In addition, one could imagine alternative therapeutic strategies focused on targeting RNA structures within viral ORFs to interfere with the virus cycle as well as to promote antigen presentation and to stimulate the host immune response.

Targeting viral G4s: Recently, a G4 ligand called quarfloxin (CX-3543) entered phase II clinical trials. It is capable to suppress MYC transcription by inhibiting the interaction between nucleolin and the c-myc G4 motifs (123). Original anti-cancer applications based on such ligands seem underway (124). Taken together, all these studies suggest that targeting these G-quadruplexes could potentially serve as a basis for novel antiviral therapies. Hence, as described for cellular targets (125), small ligands that can stabilize the G-quadruplex structure may compose a potential new class of therapeutic agents to fight viral infections (55,126) [for reviews, see (127,128)].

Several laboratories are already working to identify G4 ligands that preferentially interact with the pathogen's G4 over cellular G-quadruplexes. This new antiviral strategy presents many advantages: (i) these compounds target viral DNAs and RNAs, which are the source of the disease; (ii) high

conservation of these targets across subspecies suggests that they are important for viruses, and mutations will likely impact viral fitness, limiting the emergence of resistant strains.

G4s as antiviral agents

Nucleic acids are currently used as therapeutics. Formivirsen (Vitravene[®]) is the first anti-sense oligonucleotide with a phosphorothioate backbone that was approved in 1998 to treat viral retinitis (CMV) (129). Pegaptanib (Macugen[®]) is another example of therapeutic nucleic acid that has been approved by the FDA in 2004. This 2'-fluoropyrimidine RNA-based aptamer targets VEGF to treat neovascular age-related macular degeneration (130). Regarding G-quadruplexes, certain G4 oligonucleotides and G4 ligands present therapeutic potential, including the thrombin-binding aptamer (anti-TBA, G₂T₂G₂TGTG₂T₂G₂) (131,132). More recently, quarfloxin (CX-3543) entered phase II clinical trials due to its capability to bind G4s; inhibit RNA PolI and nucleolin binding to G4 motifs, and suppress MYC transcription (123). Taken together, original anti-cancer applications based on G4 seem underway (124).

For most applications, G4-forming molecules have been selected through combinatorial methods (e.g., pegaptanib, anti-TBA). Two key protocols have been developed. First, polynucleotide arrays facilitated rapid detection of molecules with an affinity to the target. Then, the development of SELEX (systematic evolution of ligands by exponential enrichment) facilitates isolation of oligonucleotide sequences with the capacity to recognize virtually any class of target molecule with high affinity and specificity (133). These oligonucleotides, “aptamers”, are emerging as a class of molecules that rival antibodies in both therapeutic and diagnostic applications (134,135). In this section, we describe a few examples of aptamers specifically selected to interact with viral components and interfere with viral replication.

Hepatitis A virus (HAV). HAV belongs to the *picornaviridae* family. Even without a specific

treatment for HAV infection, a vaccine has been produced to protect against initial infection. However, only a limited number of countries recommend vaccination because HAV is rarely lethal. The HAV genome comprises a 7.5-kb single-stranded (+) RNA with a 5' untranslated region (5'-UTR) and 3' poly(A) tail. It encodes for a single polyprotein, which includes the 3C protease. This enzyme is crucial for the virus because it cleaves the polyprotein into several capsid proteins and non-structural proteins. Additionally, the 3C protease binds regulatory structural elements at the 5'-UTR, which control viral genome replication. Thus, the 3C protease is an attractive target for develop a specific antiviral treatment. Recently, hexadeoxyribonucleotides that specifically bind the HAV 3C protease were identified through a hexanucleotide array (136). *In vitro* experiments showed that the hexanucleotide GGGGGT (G₅T) forms a tetramolecular G-quadruplex, binds the C-terminal domain of HAV protease and is a potent protease inhibitor (137).

Influenza virus (the flu). Every year in winter, influenza viruses cause seasonal respiratory disease epidemics (138). Three subtypes have been defined (A, B and C), which depend on the surface proteins (hemagglutinin and neuraminidase). Sixteen hemagglutinin (H) and 9 neuraminidase (N) variants have been discovered, but only H 1, 2 and 3 as well as N 1 and 2 are commonly found in humans. These viruses belong to the *Orthomyxoviridae* family, which composes enveloped viruses with a single-stranded negative-sense RNA. The genome is 10 to 15 kb but is segmented into multiple (7 or 8) molecules. Each RNA is 0.9 to 2.3 kb and encodes for 1 or 2 proteins. The non-structural protein 1 (NS1) is a 26 kDa multifunctional protein and participates in protein–protein and protein–RNA interactions (139). During infection, NS1 interferes with cellular mRNA biology (splicing, maturation and translation). As a result, NS1 prevents interferon (IFN) production, which leads to inhibition of the host's innate immunity (140). After 15 cycles of SELEX, which employs a 45-base variable sequence, a G4-forming aptamer was selected that presents high affinity for the NS1

RNA binding domain ($K_d \approx 20$ nM) (141). In a cellular context, this G4 can block NS1 and restore INF production, which results in antiviral activity without cellular toxicity.

Severe acute respiratory syndrome coronavirus (SARS-CoV). SARS-CoV is an enveloped virus with a single-stranded RNA genome approximately 30 kb long. It encodes 2 poly-cistronic ORFs, which code for 16 non-structural proteins (Nsp). Since the 2003 outbreak, the three-dimensional structures of several key proteins have been determined (e.g., RDRP, 3C-like protease). The non-structural protein 3 (Nsp3) is one component of the viral replicase complex and contains a domain referred to as SUD (SARS unique domain), which interacts with G-quadruplexes. Thus, G-quadruplexes are also relevant for the coronavirus and might be involved either in viral replication or host immune evasion (142). On the other hand, using a SELEX approach with oligonucleotides that harbor a 30 nucleotide variable sequence and following 20 rounds of selection, DNA aptamers against the SARS-CoV helicase were isolated (143). These aptamers compose two distinct classes, G-quadruplex and non-G-quadruplex forming sequences, which were determined through circular dichroism and gel electrophoresis. Their inhibitory effect on viral replication is being studied with encouraging results.

Human immunodeficiency virus (HIV). Combinatorial approaches have been used to identify several aptamers that target HIV-1. The first RNA aptamer isolated using the SELEX approach was an RNA pseudoknot HIV-1 RT inhibitor (144). In 1994, a study showed that the molecule ($T_2G_4T_2$) inhibited cell-to-cell and virus-to-cell spread of the HIV-1 infection (145). Its G-quadruplex structure and phosphorothioate backbone were reported as essential to this inhibition. Later studies showed that phosphodiester oligonucleotides containing only G and T inhibit HIV-1 replication and the most potent molecule, $GTG_2TG_3TG_3TG_3T$ (T30177, EC_{50} at approximately 100 nM), formed a G4 *in vitro* (146-148). This oligonucleotide and related molecules, such as T30923 ($G_3TG_3TG_3TG_3T$, Figure 1j), are potent HIV-1 IN inhibitors *in vitro*. T30177 was the first IN inhibitor tested in clinical trials

(ZintevirTM developed by Aronex Pharmaceuticals in 1996) (149). However, the action mechanism in cells is more complex because it also targets viral entry.

SELEX approaches were developed to isolate DNA aptamers with high affinity for the RNase H domain of HIV-1 RT. Thus, the target protein used was either the isolated RNase H domain (p15) or the functional heterodimer p66/p51 with a counter-selection that used the p51/p51 form (truncated for the RNase H domain) (150,151). Interestingly, these selections mainly led to aptamers with G-rich sequences capable of forming G-quadruplexes. Some, but not all, of these G4 inhibited the RNase H activity of HIV-1 RT *in vitro* with an IC₅₀ in the nM range (150). Surprisingly, these G4 aptamers (93del and 112del) were also potent IN inhibitors *in vitro* with IC₅₀ values in the range 10-40 nM (152). This dual inhibition can be explained by the structural similarities between the IN active site and RT RNase H domain (153). The aptamer 93del can form an original dimeric interlocked G4 (Figure 1i), which is stable even at temperatures over 90°C (154). Through comparing the structures of the G4-forming aptamers that inhibit IN, IN inhibition likely requires a stack of 6 G-tetrads (148).

Similar to T30177, 93del is a potent antiviral agent with multimodal inhibition and, in cells, targets the viral entry step, reverse transcription and integration (155). Additional *in vitro* studies indicated that free 93del and T30923 can enter human cells, including epithelial (HeLa), hepatic (Huh7), and lymphocytes (H9) cells (156). However, striking differences were observed in the presence of viral particles; HIV-1 strongly stimulates cellular uptake of aptamer. This latter observation opens an opportunity for specific drug delivery to cells that are infected, which, thus, decreases the potential side effects from G4 off-targeting.

G4s as detection tools

Given the remarkable physical properties of G4s, scientists have engineered innovative tools using G4 folding to control ribozyme activity and constructed probes for state-of-the-art imagery as well as quantification techniques, among other approaches. As an example, the aforementioned G-quadruplex 93del was engineered into aptamer beacons to visualize endogenous protein HIV-1 reverse transcriptase in living cells (157). The following section describes two such innovations related to viruses.

Hepatitis delta virus (HDV). HDV is a small, enveloped virus that depends on hepatitis B virus (HBV) for propagation (158). Its genome is a circular, single-stranded (-) RNA molecule 1679 nucleotides long; however, due to self-complementarity over approximately 70% of its length, it folds into a partially double-stranded molecule (rod-like RNA structure). Because HDV does not encode for a polymerase, replication of its genome relies on cellular enzymes. Upon polymerization following the 'double rolling circle' mechanism, genomic and anti-genomic circular RNA is produced with linear polyadenylated mRNA that only codes for the HDV protein, delta antigen (HDAg). However, HDV also encompasses an 85 nucleotide ribozyme with self-cleavage activity. Molecular engineering led to development of a "G-quartzyme", which is a ribozyme controlled by a G-quadruplex structure (159,160). Stabilization of the G-quadruplex at the 3' end by monovalent ions (K^+) activates the ribozyme.

Cauliflower mosaic virus (CaMV). CaMV is a pararetrovirus that infects plants (161). Its genome is an 8 kb, circular, double-stranded DNA molecule produced through reverse transcription of a pre-genomic mRNA. Therefore, transcription of the genome is a crucial step for two reasons: i) it is necessary to code for several viral proteins, and ii) it generates 35S RNA that is slightly longer than the genome (terminally redundant), which serves as a matrix to replicate the DNA genome. Because the unique CaMV promoter, the 35S promoter, is a strong and constitutive promoter, it has been

widely used to develop genetically modified organisms (GMOs). Consequently, there is a growing need to detect GMOs in the food industry, and many methods have been developed based on ELISA, PCR, and other techniques. Notably, a fluorescent assay has been engineered based on G-quadruplex formation. It involves two specific primers that recognize the 35S promoter. Each primer is linked to a double repeat of the human telomeric sequence TTAGGG. As a consequence, binding the 2 primers facilitates inter-strand quadruplex formation. Upon adding berberine (a selective G4-binder), a strong fluorescent signal is produced (162), which facilitates promoter detection.

Conclusion

The X-ray structure of G-tetrads was determined in the early 60s; in the 90s, G-quadruplexes remained an intriguing deviation from the Watson-Crick canonical structure. However, a remarkable quantity of information has been obtained over the past two decades, which has allowed this field to grow from basic science to clinical application (163,164). Thus, the presence of G-quadruplexes in telomeres and oncogenic promoters has opened broad opportunities for understanding and generating new treatments against cancer.

The viral world is extensive and includes viruses with replication schemes based on RNA (*e.g.*, flaviviruses), DNA (*e.g.*, papillomavirus) or intermediates thereof (*e.g.*, retroviruses), and G4s are part of their life cycle. First, they must address the presence of G4s in their hosts, and second, they must also contain G4-forming sequences that regulate important replication steps. Several challenges remain to better define the features that provide specific G4 motifs with the ability to function as structural elements. Nevertheless, G-quadruplex sequences and G4-binders have been identified in some of the most pathogenic viruses; thus, they are attractive targets for controlling viral infections.

References

1. Bang, I. (1910) Untersuchungen über die Guanylsäure. *Biochemische Zeitschrift*, **26**, 293-311.

2. Gellert, M., Lipsett, M.N. and Davies, D.R. (1962) Helix formation by guanylic acid. *Proceedings of the National Academy of Sciences of the United States of America*, **48**, 2013-2018.
3. Sasisekharan, V., Zimmerman, S. and Davies, D.R. (1975) The structure of helical 5'-guanosine monophosphate. *Journal of molecular biology*, **92**, 171-179.
4. Howard, F.B. and Miles, H.T. (1982) Poly(inosinic acid) helices: essential chelation of alkali metal ions in the axial channel. *Biochemistry*, **21**, 6736-6745.
5. Williamson, J.R. (1993) Guanine quartets. *Current Opinion in Structural Biology*, **3**, 357-362.
6. Sundquist, W.I. and Klug, A. (1989) Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature*, **342**, 825-829.
7. Miura, T., Benevides, J.M. and Thomas, G.J., Jr. (1995) A phase diagram for sodium and potassium ion control of polymorphism in telomeric DNA. *Journal of molecular biology*, **248**, 233-238.
8. Bouaziz, S., Kettani, A. and Patel, D.J. (1998) A K cation-induced conformational switch within a loop spanning segment of a DNA quadruplex containing G-G-G-C repeats. *Journal of molecular biology*, **282**, 637-652.
9. Chantot, J. and Guschlbauer, W. (1969) Physicochemical properties of nucleosides 3. Gel formation by 8-bromoguanosine. *FEBS letters*, **4**, 173-176.
10. Chantot, J.F., Sarocchi, M.T. and Guschlbauer, W. (1971) Physico-chemical properties of nucleosides. 4. Gel formation by guanosine and its analogues. *Biochimie*, **53**, 347-354.
11. Tougard, P., Chantot, J.F. and Guschlbauer, W. (1973) Nucleoside conformations. X. An X-ray fiber diffraction study of the gels of guanine nucleosides. *Biochimica et biophysica acta*, **308**, 9-16.
12. Delabar, J.M. and Guschlbauer, W. (1973) Nucleoside conformations. XI. Solvent effects on optical properties of guanosine and its derivatives in dilute solutions. *Journal of the American Chemical Society*, **95**, 5729-5735.
13. Leicknam, J.P., Chauvelier, C., Chantot, J.F. and Guschlbauer, W. (1973) Nucleoside conformations. 12. An infrared study of the polymorphism of guanine nucleosides in the solid state. *Biophysical chemistry*, **1**, 134-139.
14. Chantot, J.F., Haertle, T. and Guschlbauer, W. (1974) Nucleoside conformations. XIII. Circular dichroism of guanosine gels and the conformation of GpG and poly (G). *Biochimie*, **56**, 501-507.
15. Guschlbauer, W., Chantot, J.F. and Thiele, D. (1990) Four-stranded nucleic acid structures 25 years later: from guanosine gels to telomer DNA. *Journal of biomolecular structure & dynamics*, **8**, 491-511.
16. Williamson, J.R. (1994) G-quartet structures in telomeric DNA. *Annual review of biophysics and biomolecular structure*, **23**, 703-730.
17. Gottarelli, G., Spada, G.P. and Garbesi, A. (1996) In J.-M. Lehn, J. L. A., D. D. MacNicol, J. A. D. Davies, F. Vögtle, J.-P. Sauvage, M. W. Hosseini (ed.), *Comprehensive Supramolecular Chemistry*. Pergamon, Vol. 9, pp. 483-506.
18. Gilbert, D.E. and Feigon, J. (1999) Multistranded DNA structures. *Curr Opin Struct Biol*, **9**, 305-314.
19. Keniry, M.A. (2000) Quadruplex structures in nucleic acids. *Biopolymers*, **56**, 123-146.
20. Hardin, C.C., Perry, A.G. and White, K. (2000) Thermodynamic and kinetic characterization of the dissociation and assembly of quadruplex nucleic acids. *Biopolymers*, **56**, 147-194.
21. Shafer, R.H. and Smirnov, I. (2000) Biological aspects of DNA/RNA quadruplexes. *Biopolymers*, **56**, 209-227.
22. Suhnel, J. (2001) Beyond nucleic acid base pairs: from triads to heptads. *Biopolymers*, **61**, 32-51.

23. Simonsson, T. (2001) G-quadruplex DNA structures--variations on a theme. *Biological chemistry*, **382**, 621-628.
24. Liu, Z. and Gilbert, W. (1994) The yeast KEM1 gene encodes a nuclease specific for G4 tetraplex DNA: implication of in vivo functions for this novel DNA structure. *Cell*, **77**, 1083-1092.
25. Sun, H., Yabuki, A. and Maizels, N. (2001) A human nuclease specific for G4 DNA. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 12444-12449.
26. Creacy, S.D., Routh, E.D., Iwamoto, F., Nagamine, Y., Akman, S.A. and Vaughn, J.P. (2008) G4 resolvase 1 binds both DNA and RNA tetramolecular quadruplex with high affinity and is the major source of tetramolecular quadruplex G4-DNA and G4-RNA resolving activity in HeLa cell lysates. *The Journal of biological chemistry*, **283**, 34626-34634.
27. Vaughn, J.P., Creacy, S.D., Routh, E.D., Joyner-Butt, C., Jenkins, G.S., Pauli, S., Nagamine, Y. and Akman, S.A. (2005) The DEXH protein product of the DHX36 gene is the major source of tetramolecular quadruplex G4-DNA resolving activity in HeLa cell lysates. *The Journal of biological chemistry*, **280**, 38117-38120.
28. Fernando, H., Rodriguez, R. and Balasubramanian, S. (2008) Selective recognition of a DNA G-quadruplex by an engineered antibody. *Biochemistry*, **47**, 9365-9371.
29. Henderson, A., Wu, Y., Huang, Y.C., Chavez, E.A., Platt, J., Johnson, F.B., Brosh, R.M., Jr., Sen, D. and Lansdorp, P.M. (2014) Detection of G-quadruplex DNA in mammalian cells. *Nucleic acids research*, **42**, 860-869.
30. Schaffitzel, C., Berger, I., Postberg, J., Hanes, J., Lipps, H.J. and Pluckthun, A. (2001) In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with *Styloynchia lemnae* macronuclei. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 8572-8577.
31. 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. *Nature chemistry*, **6**, 75-80.
32. Biffi, G., Tannahill, D., McCafferty, J. and Balasubramanian, S. (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nature chemistry*, **5**, 182-186.
33. Rizzo, A., Salvati, E., Porru, M., D'Angelo, C., Stevens, M.F., D'Incalci, M., Leonetti, C., Gilson, E., Zupi, G. and Biroccio, A. (2009) Stabilization of quadruplex DNA perturbs telomere replication leading to the activation of an ATR-dependent ATM signaling pathway. *Nucleic acids research*, **37**, 5353-5364.
34. Rodriguez, R., Miller, K.M., Forment, J.V., Bradshaw, C.R., Nikan, M., Britton, S., Oelschlaegel, T., Xhemalce, B., Balasubramanian, S. and Jackson, S.P. (2012) Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. *Nature chemical biology*, **8**, 301-310.
35. Shin-ya, K., Wierzba, K., Matsuo, K., Ohtani, T., Yamada, Y., Furihata, K., Hayakawa, Y. and Seto, H. (2001) Telomestatin, a novel telomerase inhibitor from *Streptomyces anulatus*. *Journal of the American Chemical Society*, **123**, 1262-1263.
36. Tauchi, T., Shin-ya, K., Sashida, G., Sumi, M., Okabe, S., Ohyashiki, J.H. and Ohyashiki, K. (2006) Telomerase inhibition with a novel G-quadruplex-interactive agent, telomestatin: in vitro and in vivo studies in acute leukemia. *Oncogene*, **25**, 5719-5725.
37. Huppert, J.L. and Balasubramanian, S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic acids research*, **33**, 2908-2916.
38. Yang, D. and Okamoto, K. (2010) Structural insights into G-quadruplexes: towards new anticancer drugs. *Future medicinal chemistry*, **2**, 619-646.
39. Murat, P. and Balasubramanian, S. (2014) Existence and consequences of G-quadruplex structures in DNA. *Curr Opin Genet Dev*, **25C**, 22-29.

40. Klobutcher, L.A., Swanton, M.T., Donini, P. and Prescott, D.M. (1981) All gene-sized DNA molecules in four species of hypotrichs have the same terminal sequence and an unusual 3' terminus. *Proceedings of the National Academy of Sciences of the United States of America*, **78**, 3015-3019.
41. Blackburn, E.H. and Karrer, K.M. (1986) Genomic reorganization in ciliated protozoans. *Annual review of genetics*, **20**, 501-521.
42. Henderson, E., Hardin, C.C., Walk, S.K., Tinoco, I., Jr. and Blackburn, E.H. (1987) Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine-guanine base pairs. *Cell*, **51**, 899-908.
43. Henderson, E.R. and Blackburn, E.H. (1989) An overhanging 3' terminus is a conserved feature of telomeres. *Molecular and cellular biology*, **9**, 345-348.
44. Williamson, J.R., Raghuraman, M.K. and Cech, T.R. (1989) Monovalent cation-induced structure of telomeric DNA: the G-quartet model. *Cell*, **59**, 871-880.
45. Masuda-Sasa, T., Polaczek, P., Peng, X.P., Chen, L. and Campbell, J.L. (2008) Processing of G4 DNA by Dna2 helicase/nuclease and replication protein A (RPA) provides insights into the mechanism of Dna2/RPA substrate recognition. *The Journal of biological chemistry*, **283**, 24359-24373.
46. Lin, W., Sampathi, S., Dai, H., Liu, C., Zhou, M., Hu, J., Huang, Q., Campbell, J., Shin-Ya, K., Zheng, L. *et al.* (2013) Mammalian DNA2 helicase/nuclease cleaves G-quadruplex DNA and is required for telomere integrity. *The EMBO journal*, **32**, 1425-1439.
47. Prescott, J.C. and Blackburn, E.H. (1999) Telomerase: Dr Jekyll or Mr Hyde? *Curr Opin Genet Dev*, **9**, 368-373.
48. Zhu, J., Wang, H., Bishop, J.M. and Blackburn, E.H. (1999) Telomerase extends the lifespan of virus-transformed human cells without net telomere lengthening. *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 3723-3728.
49. Mitchell, J.R., Wood, E. and Collins, K. (1999) A telomerase component is defective in the human disease dyskeratosis congenita. *Nature*, **402**, 551-555.
50. Angrisani, A., Vicidomini, R., Turano, M. and Furia, M. (2014) Human dyskerin: beyond telomeres. *Biological chemistry*, **395**, 593-610.
51. Wang, Q., Liu, J.Q., Chen, Z., Zheng, K.W., Chen, C.Y., Hao, Y.H. and Tan, Z. (2011) G-quadruplex formation at the 3' end of telomere DNA inhibits its extension by telomerase, polymerase and unwinding by helicase. *Nucleic acids research*, **39**, 6229-6237.
52. Redman, J.E., Ladame, S., Reszka, A.P., Neidle, S. and Balasubramanian, S. (2006) Discovery of G-quadruplex stabilizing ligands through direct ELISA of a one-bead-one-compound library. *Organic & biomolecular chemistry*, **4**, 4364-4369.
53. Jain, A.K. and Bhattacharya, S. (2012) Recent developments in the chemistry and biology of G-quadruplexes with reference to the DNA groove binders. *Current pharmaceutical design*, **18**, 1917-1933.
54. Kar, R.K., Suryadevara, P., Jana, J., Bhunia, A. and Chatterjee, S. (2013) Novel G-quadruplex stabilizing agents: in-silico approach and dynamics. *Journal of biomolecular structure & dynamics*, **31**, 1497-1518.
55. Tsai, Y.C., Qi, H., Lin, C.P., Lin, R.K., Kerrigan, J.E., Rzuczek, S.G., LaVoie, E.J., Rice, J.E., Pilch, D.S., Lyu, Y.L. *et al.* (2009) A G-quadruplex stabilizer induces M-phase cell cycle arrest. *The Journal of biological chemistry*, **284**, 22535-22543.
56. De Cian, A., Cristofari, G., Reichenbach, P., De Lemos, E., Monchard, D., Teulade-Fichou, M.P., Shin-Ya, K., Lacroix, L., Lingner, J. and Mergny, J.L. (2007) Reevaluation of telomerase inhibition by quadruplex ligands and their mechanisms of action. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 17347-17352.
57. Miyazaki, T., Pan, Y., Joshi, K., Purohit, D., Hu, B., Demir, H., Mazumder, S., Okabe, S., Yamori, T., Viapiano, M. *et al.* (2012) Telomestatin impairs glioma stem cell survival and

growth through the disruption of telomeric G-quadruplex and inhibition of the proto-oncogene, c-Myb. *Clinical cancer research : an official journal of the American Association for Cancer Research*, **18**, 1268-1280.

58. Aladjem, M.I. (2007) Replication in context: dynamic regulation of DNA replication patterns in metazoans. *Nature reviews. Genetics*, **8**, 588-600.

59. Cayrou, C., Coulombe, P., Puy, A., Rialle, S., Kaplan, N., Segal, E. and Mechali, M. (2012) New insights into replication origin characteristics in metazoans. *Cell cycle*, **11**, 658-667.

60. Besnard, E., Babled, A., Lapasset, L., Milhavet, O., Parrinello, H., Dantec, C., Marin, J.M. and Lemaître, J.M. (2012) Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nature structural & molecular biology*, **19**, 837-844.

61. Valton, A.L., Hassan-Zadeh, V., Lema, I., Boggetto, N., Alberti, P., Saintome, C., Riou, J.F. and Prioleau, M.N. (2014) G4 motifs affect origin positioning and efficiency in two vertebrate replicators. *The EMBO journal*, **33**, 732-746.

62. Brooks, T.A., Kendrick, S. and Hurley, L. (2010) Making sense of G-quadruplex and i-motif functions in oncogene promoters. *The FEBS journal*, **277**, 3459-3469.

63. Simonsson, T., Pecinka, P. and Kubista, M. (1998) DNA tetraplex formation in the control region of c-myc. *Nucleic acids research*, **26**, 1167-1172.

64. Grand, C.L., Han, H., Munoz, R.M., Weitman, S., Von Hoff, D.D., Hurley, L.H. and Bearss, D.J. (2002) The cationic porphyrin TMPyP4 down-regulates c-MYC and human telomerase reverse transcriptase expression and inhibits tumor growth in vivo. *Molecular cancer therapeutics*, **1**, 565-573.

65. Siddiqui-Jain, A., Grand, C.L., Bearss, D.J. and Hurley, L.H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 11593-11598.

66. Seenisamy, J., Rezler, E.M., Powell, T.J., Tye, D., Gokhale, V., Joshi, C.S., Siddiqui-Jain, A. and Hurley, L.H. (2004) The dynamic character of the G-quadruplex element in the c-MYC promoter and modification by TMPyP4. *Journal of the American Chemical Society*, **126**, 8702-8709.

67. Phan, A.T., Modi, Y.S. and Patel, D.J. (2004) Propeller-type parallel-stranded G-quadruplexes in the human c-myc promoter. *Journal of the American Chemical Society*, **126**, 8710-8716.

68. Davis, T.L., Firulli, A.B. and Kinniburgh, A.J. (1989) Ribonucleoprotein and protein factors bind to an H-DNA-forming c-myc DNA element: possible regulators of the c-myc gene. *Proceedings of the National Academy of Sciences of the United States of America*, **86**, 9682-9686.

69. Berberich, S.J. and Postel, E.H. (1995) PuF/NM23-H2/NDPK-B transactivates a human c-myc promoter-CAT gene via a functional nuclease hypersensitive element. *Oncogene*, **10**, 2343-2347.

70. Gonzalez, V., Guo, K., Hurley, L. and Sun, D. (2009) Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein. *The Journal of biological chemistry*, **284**, 23622-23635.

71. Kang, H.J., Le, T.V., Kim, K., Hur, J., Kim, K.K. and Park, H.J. (2014) Novel Interaction of the Z-DNA Binding Domain of Human ADAR1 with the Oncogenic c-Myc Promoter G-Quadruplex. *Journal of molecular biology*.

72. Jordano, J. and Perucho, M. (1986) Chromatin structure of the promoter region of the human c-K-ras gene. *Nucleic acids research*, **14**, 7361-7378.

73. Cogoi, S., Quadrifoglio, F. and Xodo, L.E. (2004) G-rich oligonucleotide inhibits the binding of a nuclear protein to the Ki-ras promoter and strongly reduces cell growth in human carcinoma pancreatic cells. *Biochemistry*, **43**, 2512-2523.
74. Guo, K., Pourpak, A., Beetz-Rogers, K., Gokhale, V., Sun, D. and Hurley, L.H. (2007) Formation of pseudosymmetrical G-quadruplex and i-motif structures in the proximal promoter region of the RET oncogene. *Journal of the American Chemical Society*, **129**, 10220-10228.
75. Kumari, S., Bugaut, A., Huppert, J.L. and Balasubramanian, S. (2007) An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. *Nature chemical biology*, **3**, 218-221.
76. Morris, M.J. and Basu, S. (2009) An unusually stable G-quadruplex within the 5'-UTR of the MT3 matrix metalloproteinase mRNA represses translation in eukaryotic cells. *Biochemistry*, **48**, 5313-5319.
77. Gomez, D., Guedin, A., Mergny, J.L., Salles, B., Riou, J.F., Teulade-Fichou, M.P. and Calsou, P. (2010) A G-quadruplex structure within the 5'-UTR of TRF2 mRNA represses translation in human cells. *Nucleic acids research*, **38**, 7187-7198.
78. Beaudoin, J.D. and Perreault, J.P. (2010) 5'-UTR G-quadruplex structures acting as translational repressors. *Nucleic acids research*, **38**, 7022-7036.
79. Bugaut, A. and Balasubramanian, S. (2012) 5'-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic acids research*, **40**, 4727-4741.
80. Yu, C.H., Teulade-Fichou, M.P. and Olsthoorn, R.C. (2014) Stimulation of ribosomal frameshifting by RNA G-quadruplex structures. *Nucleic acids research*, **42**, 1887-1892.
81. Lopez Castel, A., Cleary, J.D. and Pearson, C.E. (2010) Repeat instability as the basis for human diseases and as a potential target for therapy. *Nature reviews. Molecular cell biology*, **11**, 165-170.
82. Saha, T. and Usdin, K. (2001) Tetraplex formation by the progressive myoclonus epilepsy type-1 repeat: implications for instability in the repeat expansion diseases. *FEBS letters*, **491**, 184-187.
83. Larson, E.D., Duquette, M.L., Cummings, W.J., Streiff, R.J. and Maizels, N. (2005) MutSalpα binds to and promotes synapsis of transcriptionally activated immunoglobulin switch regions. *Current biology : CB*, **15**, 470-474.
84. Smargiasso, N., Gabelica, V., Damblon, C., Rosu, F., De Pauw, E., Teulade-Fichou, M.P., Rowe, J.A. and Claessens, A. (2009) Putative DNA G-quadruplex formation within the promoters of *Plasmodium falciparum* var genes. *BMC genomics*, **10**, 362.
85. Kuryavyi, V., Cahoon, L.A., Seifert, H.S. and Patel, D.J. (2012) RecA-binding pilE G4 sequence essential for pilin antigenic variation forms monomeric and 5' end-stacked dimeric parallel G-quadruplexes. *Structure*, **20**, 2090-2102.
86. Maizels, N. and Gray, L.T. (2013) The G4 genome. *PLoS genetics*, **9**, e1003468.
87. Metifiot, M., Marchand, C. and Pommier, Y. (2013) HIV integrase inhibitors: 20-year landmark and challenges. *Advances in pharmacology*, **67**, 75-105.
88. Jouvenet, N., Laine, S., Pessel-Vivares, L. and Mougél, M. (2011) Cell biology of retroviral RNA packaging. *RNA biology*, **8**, 572-580.
89. Hu, W.S. and Temin, H.M. (1990) Retroviral recombination and reverse transcription. *Science*, **250**, 1227-1233.
90. Charpentier, C., Nora, T., Tenaillon, O., Clavel, F. and Hance, A.J. (2006) Extensive recombination among human immunodeficiency virus type 1 quasispecies makes an important contribution to viral diversity in individual patients. *Journal of virology*, **80**, 2472-2482.
91. Ly, H., Nierlich, D.P., Olsen, J.C. and Kaplan, A.H. (1999) Moloney murine sarcoma virus genomic RNAs dimerize via a two-step process: a concentration-dependent kissing-loop

- interaction is driven by initial contact between consecutive guanines. *Journal of virology*, **73**, 7255-7261.
92. Ly, H., Nierlich, D.P., Olsen, J.C. and Kaplan, A.H. (2000) Functional characterization of the dimer linkage structure RNA of Moloney murine sarcoma virus. *Journal of virology*, **74**, 9937-9945.
93. Paillart, J.C., Skripkin, E., Ehresmann, B., Ehresmann, C. and Marquet, R. (1996) A loop-loop "kissing" complex is the essential part of the dimer linkage of genomic HIV-1 RNA. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 5572-5577.
94. Sundquist, W.I. and Heaphy, S. (1993) Evidence for interstrand quadruplex formation in the dimerization of human immunodeficiency virus 1 genomic RNA. *Proceedings of the National Academy of Sciences of the United States of America*, **90**, 3393-3397.
95. Awang, G. and Sen, D. (1993) Mode of dimerization of HIV-1 genomic RNA. *Biochemistry*, **32**, 11453-11457.
96. Marquet, R., Paillart, J.C., Skripkin, E., Ehresmann, C. and Ehresmann, B. (1994) Dimerization of human immunodeficiency virus type 1 RNA involves sequences located upstream of the splice donor site. *Nucleic acids research*, **22**, 145-151.
97. Shen, W., Gorelick, R.J. and Bambara, R.A. (2011) HIV-1 nucleocapsid protein increases strand transfer recombination by promoting dimeric G-quartet formation. *The Journal of biological chemistry*, **286**, 29838-29847.
98. 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. *The Journal of biological chemistry*, **288**, 24140-24150.
99. Zennou, V., Petit, C., Guetard, D., Nerhbass, U., Montagnier, L. and Charneau, P. (2000) HIV-1 genome nuclear import is mediated by a central DNA flap. *Cell*, **101**, 173-185.
100. Riviere, L., Darlix, J.L. and Cimarrelli, A. (2010) Analysis of the viral elements required in the nuclear import of HIV-1 DNA. *Journal of virology*, **84**, 729-739.
101. Darlix, J.L., Lapadat-Tapolsky, M., de Rocquigny, H. and Roques, B.P. (1995) First glimpses at structure-function relationships of the nucleocapsid protein of retroviruses. *Journal of molecular biology*, **254**, 523-537.
102. Godet, J., Boudier, C., Humbert, N., Ivanyi-Nagy, R., Darlix, J.L. and Mely, Y. (2012) Comparative nucleic acid chaperone properties of the nucleocapsid protein NCp7 and Tat protein of HIV-1. *Virus research*, **169**, 349-360.
103. Lonnais, 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 research*, **31**, 5754-5763.
104. Kankia, B.I., Barany, G. and Musier-Forsyth, K. (2005) Unfolding of DNA quadruplexes induced by HIV-1 nucleocapsid protein. *Nucleic acids research*, **33**, 4395-4403.
105. Rajendran, A., Endo, M., Hidaka, K., Tran, P.L., Mergny, J.L., Gorelick, R.J. and Sugiyama, H. (2013) HIV-1 nucleocapsid proteins as molecular chaperones for tetramolecular antiparallel G-quadruplex formation. *Journal of the American Chemical Society*, **135**, 18575-18585.
106. Kuryavyi, V., Phan, A.T. and Patel, D.J. (2010) Solution structures of all parallel-stranded monomeric and dimeric G-quadruplex scaffolds of the human c-kit2 promoter. *Nucleic acids research*, **38**, 6757-6773.
107. Perrone, R., Nadai, M., Frasson, I., Poe, J.A., Butovskaya, E., Smithgall, T.E., Palumbo, M., Palu, G. and Richter, S.N. (2013) A dynamic G-quadruplex region regulates the HIV-1 long terminal repeat promoter. *Journal of medicinal chemistry*, **56**, 6521-6530.

108. 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. *Journal of the American Chemical Society*, **136**, 5249-5252.
109. 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.
110. Perrone, R., Nadai, M., Poe, J.A., Frasson, I., Palumbo, M., Palu, G., Smithgall, T.E. and Richter, S.N. (2013) Formation of a unique cluster of G-quadruplex structures in the HIV-1 Nef coding region: implications for antiviral activity. *PloS one*, **8**, e73121.
111. Patel, P.K., Bhavesh, N.S. and Hosur, R.V. (2000) NMR observation of a novel C-tetrad in the structure of the SV40 repeat sequence GGGCGG. *Biochemical and biophysical research communications*, **270**, 967-971.
112. Topalis, D., Andrei, G. and Snoeck, R. (2013) The large tumor antigen: a "Swiss Army knife" protein possessing the functions required for the polyomavirus life cycle. *Antiviral research*, **97**, 122-136.
113. Baran, N., Pucshansky, L., Marco, Y., Benjamin, S. and Manor, H. (1997) The SV40 large T-antigen helicase can unwind four stranded DNA structures linked by G-quartets. *Nucleic acids research*, **25**, 297-303.
114. Plyler, J., Jasheway, K., Tuesuwan, B., Karr, J., Brennan, J.S., Kerwin, S.M. and David, W.M. (2009) Real-time investigation of SV40 large T-antigen helicase activity using surface plasmon resonance. *Cell biochemistry and biophysics*, **53**, 43-52.
115. Tuesuwan, B., Kern, J.T., Thomas, P.W., Rodriguez, M., Li, J., David, W.M. and Kerwin, S.M. (2008) Simian virus 40 large T-antigen G-quadruplex DNA helicase inhibition by G-quadruplex DNA-interactive agents. *Biochemistry*, **47**, 1896-1909.
116. Sharma, S., Doherty, K.M. and Brosh, R.M., Jr. (2005) DNA helicases as targets for anti-cancer drugs. *Current medicinal chemistry. Anti-cancer agents*, **5**, 183-199.
117. Sun, H., Karow, J.K., Hickson, I.D. and Maizels, N. (1998) The Bloom's syndrome helicase unwinds G4 DNA. *The Journal of biological chemistry*, **273**, 27587-27592.
118. Akagi, K., Li, J., Broutian, T.R., Padilla-Nash, H., Xiao, W., Jiang, B., Rocco, J.W., Teknos, T.N., Kumar, B., Wangsa, D. *et al.* (2014) Genome-wide analysis of HPV integration in human cancers reveals recurrent, focal genomic instability. *Genome research*, **24**, 185-199.
119. Tluckova, K., Marusic, M., Tothova, P., Bauer, L., Sket, P., Plavec, J. and Viglasky, V. (2013) Human papillomavirus G-quadruplexes. *Biochemistry*, **52**, 7207-7216.
120. Kutok, J.L. and Wang, F. (2006) Spectrum of Epstein-Barr virus-associated diseases. *Annual review of pathology*, **1**, 375-404.
121. Norseen, J., Johnson, F.B. and Lieberman, P.M. (2009) Role for G-quadruplex RNA binding by Epstein-Barr virus nuclear antigen 1 in DNA replication and metaphase chromosome attachment. *Journal of virology*, **83**, 10336-10346.
122. Murat, P., Zhong, J., Lekieffre, L., Cowieson, N.P., Clancy, J.L., Preiss, T., Balasubramanian, S., Khanna, R. and Tellam, J. (2014) G-quadruplexes regulate Epstein-Barr virus-encoded nuclear antigen 1 mRNA translation. *Nature chemical biology*, **10**, 358-364.
123. Balasubramanian, S., Hurley, L.H. and Neidle, S. (2011) Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nature reviews. Drug discovery*, **10**, 261-275.
124. Shalaby, T., Fiaschetti, G., Nagasawa, K., Shin-ya, K., Baumgartner, M. and Grotzer, M. (2013) G-quadruplexes as potential therapeutic targets for embryonal tumors. *Molecules*, **18**, 12500-12537.
125. Renaud de la Faverie, A., Hamon, F., Di Primo, C., Largy, E., Dausse, E., Delauriere, L., Landras-Guetta, C., Toulme, J.J., Teulade-Fichou, M.P. and Mergny, J.L. (2011) Nucleic acids targeted to drugs: SELEX against a quadruplex ligand. *Biochimie*, **93**, 1357-1367.

126. 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. *Nature chemical biology*, **1**, 167-173.
127. Hurley, L.H., Wheelhouse, R.T., Sun, D., Kerwin, S.M., Salazar, M., Fedoroff, O.Y., Han, F.X., Han, H., Izbicka, E. and Von Hoff, D.D. (2000) G-quadruplexes as targets for drug design. *Pharmacology & therapeutics*, **85**, 141-158.
128. Ou, T.M., Lu, Y.J., Tan, J.H., Huang, Z.S., Wong, K.Y. and Gu, L.Q. (2008) G-quadruplexes: targets in anticancer drug design. *ChemMedChem*, **3**, 690-713.
129. Perry, C.M. and Balfour, J.A. (1999) Fomivirsen. *Drugs*, **57**, 375-380; discussion 381.
130. Ng, E.W., Shima, D.T., Calias, P., Cunningham, E.T., Jr., Guyer, D.R. and Adamis, A.P. (2006) Pegaptanib, a targeted anti-VEGF aptamer for ocular vascular disease. *Nature reviews. Drug discovery*, **5**, 123-132.
131. Bock, L.C., Griffin, L.C., Latham, J.A., Vermaas, E.H. and Toole, J.J. (1992) Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature*, **355**, 564-566.
132. Avino, A., Fabrega, C., Tintore, M. and Eritja, R. (2012) Thrombin binding aptamer, more than a simple aptamer: chemically modified derivatives and biomedical applications. *Current pharmaceutical design*, **18**, 2036-2047.
133. Tuerk, C. and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, **249**, 505-510.
134. Bunka, D.H., Platonova, O. and Stockley, P.G. (2010) Development of aptamer therapeutics. *Current opinion in pharmacology*, **10**, 557-562.
135. Famulok, M. and Mayer, G. (1999) Aptamers as tools in molecular biology and immunology. *Current topics in microbiology and immunology*, **243**, 123-136.
136. Mescalchin, A., Wunsche, W. and Sczakiel, G. (2011) Specific recognition of proteins by array-bound hexanucleotides. *Angewandte Chemie*, **50**, 1052-1054.
137. Blaum, B.S., Wunsche, W., Benie, A.J., Kusov, Y., Peters, H., Gauss-Muller, V., Peters, T. and Sczakiel, G. (2012) Functional binding of hexanucleotides to 3C protease of hepatitis A virus. *Nucleic acids research*, **40**, 3042-3055.
138. Lynch, J.P., 3rd and Walsh, E.E. (2007) Influenza: evolving strategies in treatment and prevention. *Seminars in respiratory and critical care medicine*, **28**, 144-158.
139. Suarez, D.L. and Perdue, M.L. (1998) Multiple alignment comparison of the non-structural genes of influenza A viruses. *Virus research*, **54**, 59-69.
140. Donelan, N.R., Basler, C.F. and Garcia-Sastre, A. (2003) A recombinant influenza A virus expressing an RNA-binding-defective NS1 protein induces high levels of beta interferon and is attenuated in mice. *Journal of virology*, **77**, 13257-13266.
141. Woo, H.M., Kim, K.S., Lee, J.M., Shim, H.S., Cho, S.J., Lee, W.K., Ko, H.W., Keum, Y.S., Kim, S.Y., Pathinayake, P. et al. (2013) Single-stranded DNA aptamer that specifically binds to the influenza virus NS1 protein suppresses interferon antagonism. *Antiviral research*, **100**, 337-345.
142. Tan, J., Vonnrhein, C., Smart, O.S., Bricogne, G., Bollati, M., Kusov, Y., Hansen, G., Mesters, J.R., Schmidt, C.L. and Hilgenfeld, R. (2009) The SARS-unique domain (SUD) of SARS coronavirus contains two macrodomains that bind G-quadruplexes. *PLoS pathogens*, **5**, e1000428.
143. Shum, K.T. and Tanner, J.A. (2008) Differential inhibitory activities and stabilisation of DNA aptamers against the SARS coronavirus helicase. *ChemBiochem : a European journal of chemical biology*, **9**, 3037-3045.
144. Tuerk, C., MacDougall, S. and Gold, L. (1992) RNA pseudoknots that inhibit human immunodeficiency virus type 1 reverse transcriptase. *Proceedings of the National Academy of Sciences of the United States of America*, **89**, 6988-6992.

145. Wyatt, J.R., Vickers, T.A., Roberson, J.L., Buckheit, R.W., Jr., Klimkait, T., DeBaets, E., Davis, P.W., Rayner, B., Imbach, J.L. and Ecker, D.J. (1994) Combinatorially selected guanosine-quartet structure is a potent inhibitor of human immunodeficiency virus envelope-mediated cell fusion. *Proceedings of the National Academy of Sciences of the United States of America*, **91**, 1356-1360.
146. Rando, R.F., Ojwang, J., Elbaggari, A., Reyes, G.R., Tinder, R., McGrath, M.S. and Hogan, M.E. (1995) Suppression of human immunodeficiency virus type 1 activity in vitro by oligonucleotides which form intramolecular tetrads. *The Journal of biological chemistry*, **270**, 1754-1760.
147. Mazumder, A., Neamati, N., Ojwang, J.O., Sunder, S., Rando, R.F. and Pommier, Y. (1996) Inhibition of the human immunodeficiency virus type 1 integrase by guanosine quartet structures. *Biochemistry*, **35**, 13762-13771.
148. Mukundan, V.T., Do, N.Q. and Phan, A.T. (2011) HIV-1 integrase inhibitor T30177 forms a stacked dimeric G-quadruplex structure containing bulges. *Nucleic acids research*, **39**, 8984-8991.
149. Marchand, C., Maddali, K., Metifiot, M. and Pommier, Y. (2009) HIV-1 IN inhibitors: 2010 update and perspectives. *Current topics in medicinal chemistry*, **9**, 1016-1037.
150. Andreola, M.L., Pileur, F., Calmels, C., Ventura, M., Tarrago-Litvak, L., Toulme, J.J. and Litvak, S. (2001) DNA aptamers selected against the HIV-1 RNase H display in vitro antiviral activity. *Biochemistry*, **40**, 10087-10094.
151. Metifiot, M., Leon, O., Tarrago-Litvak, L., Litvak, S. and Andreola, M.L. (2005) Targeting HIV-1 integrase with aptamers selected against the purified RNase H domain of HIV-1 RT. *Biochimie*, **87**, 911-919.
152. de Soultrait, V.R., Lozach, P.Y., Altmeyer, R., Tarrago-Litvak, L., Litvak, S. and Andreola, M.L. (2002) DNA aptamers derived from HIV-1 RNase H inhibitors are strong anti-integrase agents. *Journal of molecular biology*, **324**, 195-203.
153. Andreola, M.L. (2004) Closely related antiretroviral agents as inhibitors of two HIV-1 enzymes, ribonuclease H and integrase: "killing two birds with one stone". *Current pharmaceutical design*, **10**, 3713-3723.
154. Phan, A.T., Kuryavyi, V., Ma, J.B., Faure, A., Andreola, M.L. and Patel, D.J. (2005) An interlocked dimeric parallel-stranded DNA quadruplex: a potent inhibitor of HIV-1 integrase. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 634-639.
155. Faure-Perraud, A., Metifiot, M., Reigadas, S., Recordon-Pinson, P., Parissi, V., Ventura, M. and Andreola, M.L. (2011) The guanine-quadruplex aptamer 93del inhibits HIV-1 replication ex vivo by interfering with viral entry, reverse transcription and integration. *Antiviral therapy*, **16**, 383-394.
156. Metifiot, M., Faure, A., Guyonnet-Duperat, V., Bellecave, P., Litvak, S., Ventura, M. and Andreola, M.L. (2007) Cellular uptake of ODNs in HIV-1 human-infected cells: a role for viral particles in DNA delivery? *Oligonucleotides*, **17**, 151-165.
157. Liang, Y., Zhang, Z., Wei, H., Hu, Q., Deng, J., Guo, D., Cui, Z. and Zhang, X.E. (2011) Aptamer beacons for visualization of endogenous protein HIV-1 reverse transcriptase in living cells. *Biosensors & bioelectronics*, **28**, 270-276.
158. Alvarado-Mora, M.V., Locarnini, S., Rizzetto, M. and Pinho, J.R. (2013) An update on HDV: virology, pathogenesis and treatment. *Antiviral therapy*, **18**, 541-548.
159. Beaudoin, J.D. and Perreault, J.P. (2008) Potassium ions modulate a G-quadruplex-ribozyme's activity. *Rna*, **14**, 1018-1025.
160. Reymond, C., Beaudoin, J.D. and Perreault, J.P. (2009) Modulating RNA structure and catalysis: lessons from small cleaving ribozymes. *Cellular and molecular life sciences : CMLS*, **66**, 3937-3950.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

161. Hull, R., Covey, S.N. and Maule, A.J. (1987) Structure and replication of caulimovirus genomes. *Journal of cell science. Supplement*, **7**, 213-229.

162. Qiu, B., Zhang, Y.S., Lin, Y.B., Lu, Y.J., Lin, Z.Y., Wong, K.Y. and Chen, G.N. (2013) A novel fluorescent biosensor for detection of target DNA fragment from the transgene cauliflower mosaic virus 35S promoter. *Biosensors & bioelectronics*, **41**, 168-171.

163. Neidle, S. and Balasubramanian, S. (2006) *Quadruplex nucleic acids*. RSC.

164. Neidle, S. (2011) *Therapeutic Applications of Quadruplex Nucleic Acids*. 1 ed. Academic Press.

165. Crooks, G.E., Hon, G., Chandonia, J.M. and Brenner, S.E. (2004) WebLogo: a sequence logo generator. *Genome research*, **14**, 1188-1190.

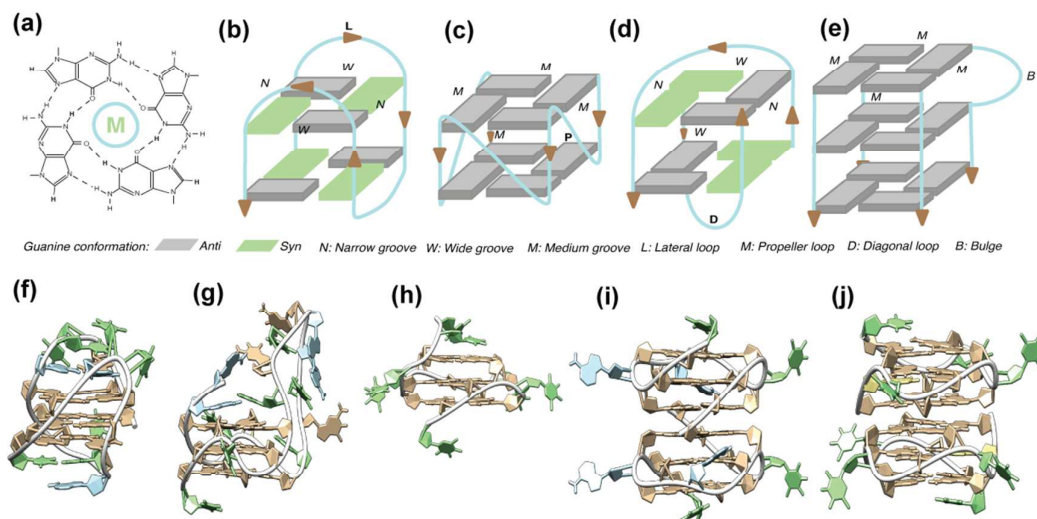


Figure 1: (a-e) Schematic representation of G-quadruplex topologies. (a) A guanine tetrad stabilized by eight Hoogsteen hydrogen bonds and a central monovalent cation (M). (b) Intramolecular antiparallel G4 topology with two tetrads, wide and narrow grooves and only lateral loops. (c) Intramolecular parallel G4 topology with two tetrads, medium grooves and only propeller loops. (d) Dimeric antiparallel G4 topology with two tetrads, wide and narrow grooves and diagonal loops. (e) Tetramolecular parallel G4 topology with three tetrads, only medium grooves and no loops. (f-j) Examples of G-quadruplex structures. (f) Intramolecular anti-parallel G4 structure with two tetrads for the telomeric sequence (PDB ID = 2KF8). (g) Intramolecular parallel G4 structure with three tetrads and a nine nucleotide central loop for the human CEB25 mini-satellite sequence (PDB ID= 2LPW). (h) Intramolecular parallel G4 structure with three tetrads for the T30177 anti-HIV aptamer (PDB ID= 2M4P). (i) Interlocked bimolecular parallel G4 structure with six tetrads for the 93del anti-HIV aptamer (PDB ID= 1Y8D). (j) Two stacked parallel G4 structures with three tetrads each observed for the T30923 anti-HIV aptamer (PDB ID= 2LE6).

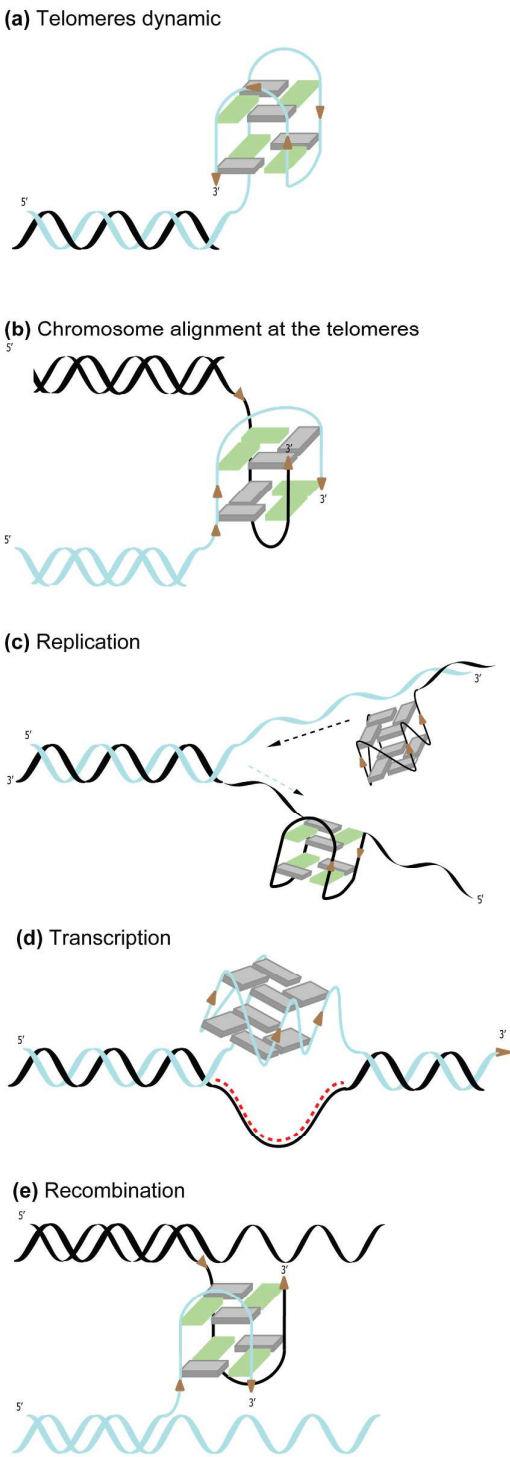


Figure 2: Role of G-quadruplex structures in the genome. **(a)** Formation of a G-quadruplex at the 3' end of a chromosome. **(b)** Formation of an intermolecular G-quadruplex between two telomeres allows for alignment of chromosomes. **(c)** Collision of the replication fork with a G4 motif blocks the leading DNA strand. **(d)** Transcription can lead to formation of a DNA/RNA duplex and G-loops on the non-template strand. **(e)** Intermolecular G-quadruplex formation leads to physical crossing and potential recombination sites at different regions of the DNA.

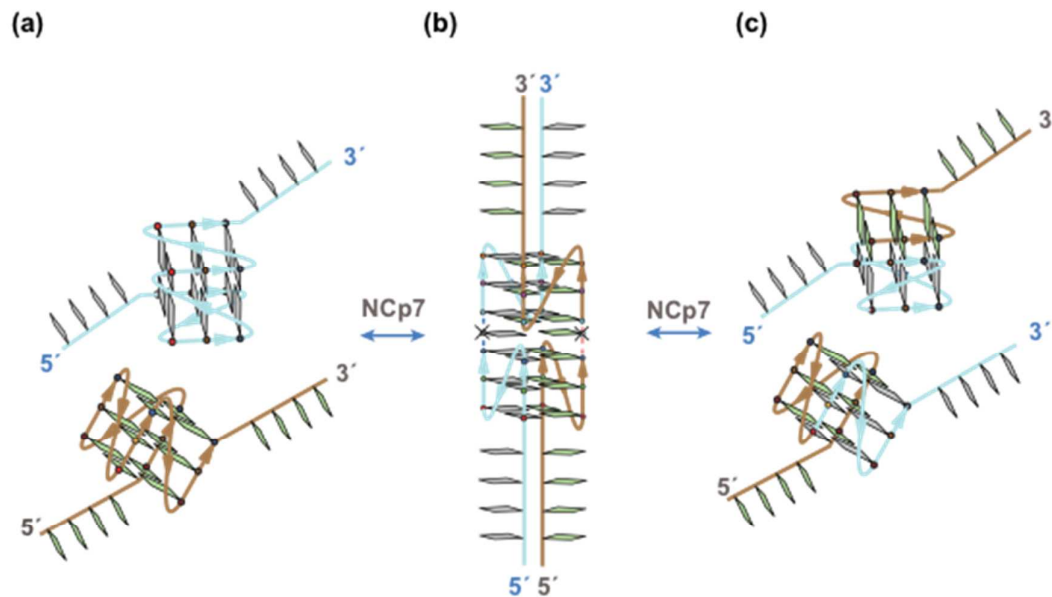


Figure 3: Model of viral RNA:RNA dimerization and recombination involving a pair of monomeric G-quadruplexes mediated by a synaptic dimeric G-quadruplex intermediate [inspired and adapted from a model of intergenic recombination (106)]. **(a)** A pair (blue and brown) of parallel-stranded monomeric G-quadruplexes. **(b)** NCp7 chaperon properties favor formation of an inter-RNA synaptic intermediate involving a parallel-stranded dimeric G-quadruplex. Cleavage, rotation and subsequent rejoining are mediated by a potential, undefined nuclease-topoisomerase-like enzyme complex. **(c)** Following dimerization and/or recombination, NCp7 annealing properties favor formation of a new pair of parallel-stranded monomeric G-quadruplexes.

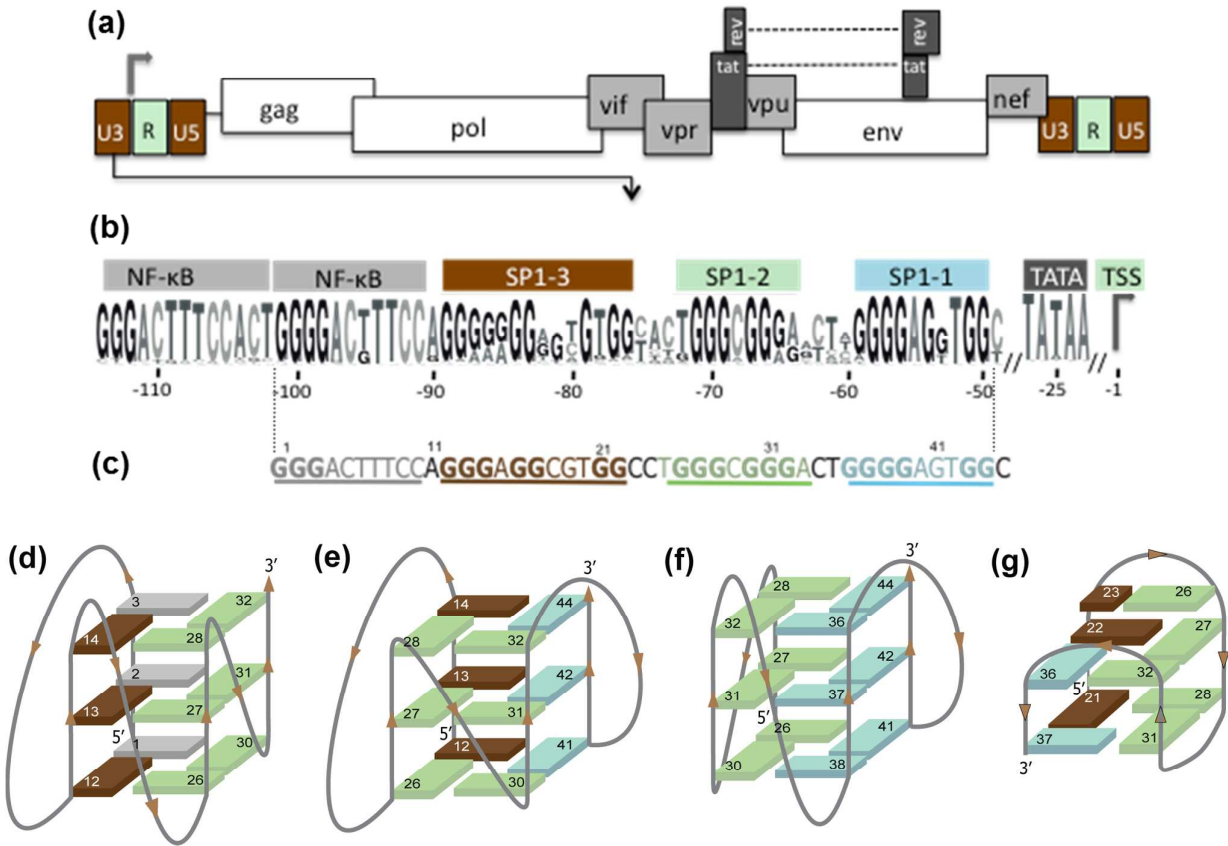


Figure 4: (a) Genomic structure of the HIV-1 provirus and (b) LOGO representation of the G-rich region for the HIV-1 promoter generated using the weblogo software (165) and based on an alignment of 1684 HIV-1 sequences provided by the HIV-1 database (www.hiv.lanl.gov). (c) The U3 region of the LTR from the HXB2_LAI;NC_001802 HIV-1 representative strain. (d,e,g) Topologies of the LTR G-quadruplexes determined using Clerocidin and DMS-mediated footprinting assays (107) for the (d) 1-32 nt segment, (e) 12-44 nt segment and (f) 26-44 nt segment. (g) Topology of the 12-37 nt segments determined using nuclear magnetic resonance (NMR) (108).