



HAL
open science

Unlocking G-Quadruplexes as Antiviral Targets

Ardavan Abiri, Marc Lavigne, Masoud Rezaei, Sanaz Nikzad, Peyman Zare,
Jean-Louis Mergny, Hamid-Reza Rahimi

► **To cite this version:**

Ardavan Abiri, Marc Lavigne, Masoud Rezaei, Sanaz Nikzad, Peyman Zare, et al.. Unlocking G-Quadruplexes as Antiviral Targets. *Pharmacological Reviews*, 2021, 73 (3), pp.897-923. 10.1124/pharmrev.120.000230 . pasteur-03700000

HAL Id: pasteur-03700000

<https://pasteur.hal.science/pasteur-03700000>

Submitted on 20 Jun 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

ASSOCIATE EDITOR: RHIAN TOUYZ

Unlocking G-Quadruplexes as Antiviral Targets

Ardavan Abiri, Marc Lavigne, Masoud Rezaei, Sanaz Nikzad, Peyman Zare,  Jean-Louis Mergny, and  Hamid-Reza Rahimi

Department of Medicinal Chemistry, Faculty of Pharmacy, Kerman University of Medical Sciences, Kerman, Iran (A.A., S.N.); Institut Pasteur, Department of Virology, UMR 3569 CNRS, Paris, France (M.L.); Faculty of Medicine, Kerman University of Medical Sciences, Kerman, Iran (M.R.); Dioscuri Center of Chromatin Biology and Epigenomics, Nencki Institute of Experimental Biology, Polish Academy of Sciences, Warsaw, Poland (P.Z.); Faculty of Medicine, Cardinal Stefan Wyszyński University in Warsaw, Warsaw, Poland (P.Z.); Laboratoire d'Optique et Biosciences, Ecole Polytechnique, CNRS, INSERM, Institut Polytechnique de Paris, Palaiseau cedex, France (J.-L.M.); Pharmaceutics Research Center, Institute of Neuropharmacology, Kerman University of Medical Sciences, Kerman, Iran (H.-R.R.); and Department of Pharmacology and Toxicology, Faculty of Pharmacy, Kerman University of Medical Sciences, Kerman, Iran (H.-R.R.)

Abstract.....	898
Significance Statement	898
I. Introduction.....	898
1. G-Quadruplexes and Pathogens	899
2. A G-Quadruplex/Virus Timeline	901
II. G-Quadruplexes and Viruses	902
1. DNA and RNA G-Quadruplexes	902
2. Searching for G-Quadruplexes.....	903
3. Viral G-Quadruplexes	904
III. Roles of G-Quadruplexes in Viral Replication and Host Response to Viral Infection	905
A. Viral and Cellular G-Quadruplex–Binding Proteins.....	905
1. Viral G-Quadruplex–Binding Proteins	905
2. Cellular G-Quadruplex–Binding Proteins	905
B. G-Quadruplex Ligands as Antiviral Compounds	905
C. G-Quadruplexes Interfering with Different Viral Steps: Human Immunodeficiency Virus-1 as a Model.....	907
D. Diversity of Regulatory Mechanisms	908
E. G-Quadruplexes in the Host Cell as Alternative Targets?.....	910
1. Immune Modulation	910
2. Oncogenic Viruses	912
3. Cellular Pathways	913
4. Telomeric Integration	914
F. Summary of G-Quadruplex Roles in Viruses and Host	914
IV. Discussion: Challenges	914
A. Specific Studies on G-Quadruplexes in Coronaviruses: Response to Coronavirus Disease 2019 Pandemic.....	914
B. Assets and Drawbacks in the Development of Antiviral Therapies Targeting G-Quadruplexes.....	916
C. Hit-to-Lead Optimization for G-Quadruplex Ligands.....	918
D. What Remains Untouched?	918

Address correspondence to: Dr. Jean-Louis Mergny, Laboratoire d'Optique et Biosciences, Institut Polytechnique de Paris, CNRS, INSERM, Université Paris-Saclay, 91128 Palaiseau cedex, France. E-mail: jean-louis.mergny@inserm.fr; or Dr. Hamid-Reza Rahimi, Department of Pharmacology and Toxicology, Faculty of Pharmacy, Kerman University of Medical Sciences, Kerman, Iran. E-mail: hamidrtr@yahoo.com; or h_rahimi@kmu.ac.ir

We acknowledge funding from Institut Pasteur Paris (to M.L.); Agence Nationale de la Recherche. Flash Covid 2020 (to M.L. and J.-L.M.); and recurrent funding from Ecole Polytechnique, CNRS, and Inserm (2020–2021 to J.-L.M.).

Financial disclosure: No author has an actual or perceived conflict of interest with the contents of this article.

<https://doi.org/10.1124/10.1124/pharmrev.120.000230>.

V. Conclusions	919
Acknowledgments	920
References	920

Abstract—Guanine-rich DNA and RNA sequences can fold into noncanonical nucleic acid structures called G-quadruplexes (G4s). Since the discovery that these structures may act as scaffolds for the binding of specific ligands, G4s aroused the attention of a growing number of scientists. The versatile roles of G4 structures in viral replication, transcription, and translation suggest direct applications in therapy or diagnostics. G4-interacting molecules (proteins or small molecules) may also affect the balance between latent and lytic phases, and increasing evidence reveals that G4s are implicated in generally suppressing viral processes, such as replication, transcription, translation, or reverse transcription. In this review, we focus on the discovery of G4s in viruses and the role of G4 ligands in the antiviral drug discovery process. After assessing the role of viral G4s, we argue that host G4s participate in immune modulation, viral tumorigenesis, cellular

pathways involved in virus maturation, and DNA integration of viral genomes, which can be potentially employed for antiviral therapeutics. Furthermore, we scrutinize the impediments and shortcomings in the process of studying G4 ligands and drug discovery. Finally, some unanswered questions regarding viral G4s are highlighted for prospective future projects.

Significance Statement—G-quadruplexes (G4s) are noncanonical nucleic acid structures that have gained increasing recognition during the last few decades. First identified as relevant targets in oncology, their importance in virology is now increasingly clear. A number of G-quadruplex ligands are known: viral transcription and replication are the main targets of these ligands. Both viral and cellular G4s may be targeted; this review embraces the different aspects of G-quadruplexes in both host and viral contexts.

I. Introduction

After giving a general overview of G-quadruplexes in the world of viruses, we will discuss these structures in more detail and the algorithms developed to predict their folding along with nucleic acid molecules. This section will be enriched by several examples of DNA and RNA viral genomes enriched in G-quadruplexes. In the following section, we will describe the partners, ligands, and molecular mechanisms associated with G-quadruplex (G4)-dependent regulation of viral replication and present the development of specific G4 ligands as antiviral compounds. A special focus will be on the role of G-quadruplexes at different steps of the human immunodeficiency virus (HIV)-1 cycle and on recent studies of G4s and severe acute respiratory syndrome coronavirus (SARS-CoV-2) replication. We will conclude by presenting the assets and drawbacks of G4-based medicinal research, especially the one targeting viral infections. In this review, the words “*G-quadruplex(es)*” and “*G4(s)*” are used interchangeably.

Although two G-quadruplex ligands, Quarfloxin (CX-3543) and CX-5461, reached clinical trial stages against

cancer (Carvalho et al., 2020), neither of them reached the clinic for antiviral applications. Regrettably, their antiviral activity in vivo is poorly explored. This probably arises from the fact that cancer has been much more the center of attention for both the public and scientists, as can be seen by the 7-fold gap between the amounts of work dedicated to G-quadruplexes in cancer compared with viral G-quadruplexes (Fig. 1). This figure nevertheless illustrates the growing number of papers published in the viral G-quadruplex field, which clearly demonstrates the remarkable attraction of G-quadruplexes from both the pharmacological and microbiological points of view.

Almost all current antiviral therapies focus on targeting proteins (and mainly viral proteins). A few biologic products are designed to target the host receptors, such as interferons (IFNs). The major advantage of targeting host proteins by therapeutic interferons is that they inhibit the replication of a wide range of viruses, reducing the risk of the development of antiviral resistance. In the case of G-quadruplex ligands, they share some characteristics of both interferons with a broad range of activity (suppressing both host and

ABBREVIATIONS: c-exNDI, core extended naphthalene diimide; COVID-19, coronavirus disease 2019; cPPT, central polypurine tract; ds-DNA, double-stranded DNA; EBNA, Epstein-Barr virus–encoded nuclear antigen; EBV, Epstein-Barr virus; encapsidation, Gag; G/C content, G4; G-quadruplex, HBV; hepatitis B virus, HCMV; human cytomegalovirus, HCV; hepatitis C virus, HHV; human herpesvirus, HIV; human immunodeficiency virus, HPV; human papillomavirus, HSV; herpes simplex virus, IE; immediate early, IFN; interferon, ISG; IFN-stimulated gene, KSHV; Kaposi’s sarcoma-associated herpes virus, LANA; latency-associated nuclear antigen, LTR; long terminal repeat, m⁶A; N⁶-methyladenosine, MDV; Marek disease virus, MERS; Middle East respiratory syndrome, NMM; N-methyl mesoporphyrin IX, NPM1; nucleophosmin, Nsp3; nonstructural protein 3, nuclear factor κB; OAS3, 2’,5’-oligoadenylate synthetase; ORF, open reading frame; pac, packaging; PDS, pyridostatin; PQS, putative quadruplex sequence; QGRS, G-Quadruplex forming G-Rich Sequences; RT, reverse transcription; RVFV, Rift Valley fever virus; SAR, structure-activity relationship; SARS, severe acute respiratory syndrome; SARS-CoV-2, SARS coronavirus; Sp1, specificity protein 1; ss-RNA, single-stranded RNA; SUD, SARS-unique domain; SV40, simian virus 40; TF, transcription factor; TMPyP4, 5,10,15,20-tetrakis(1-methylpyridinium-4-yl)porphyrin tetra(p-toluenesulfonate); TMR, telomeric repeat; TR, terminal repeat.

target proliferation) and traditional antiviral drugs (having a rather small size and better pharmacokinetics), which makes them suitable to be used as convenient and noninvasive drugs.

1. G-Quadruplexes and Pathogens. Guanine-rich nucleic acid sequences can form a variety of structures, such as the well-known tetrahelical G-quadruplex (G4) motif, the G-triplex, G-rich hairpins, and other motifs (Deng et al., 2019). G-quadruplexes are noncanonical but ubiquitous nucleic acid structures that can be formed by G-rich DNA and RNA sequences. Interest for G-quadruplexes is steadily growing, and G4 motifs have been shown or proposed to play important roles in key biologic functions, such as transcription or replication.

G4-prone motifs are found in all domains of life and may contribute to the pathogenicity of disease-causing agents. For example, among pathogens, *Plasmodium falciparum*, the protozoan eukaryote causing malaria, contains a number of G4-forming sequences [putative quadruplex sequences (PQSs)]. These motifs are found near *var* genes, which encode *P. falciparum* erythrocyte membrane protein 1, a group of variant immunodominant surface antigens esteemed as crucial virulence factors (Gage and Merrick, 2020; Gazanion et al., 2020). *Var* genes are expressed in a mutually exclusive manner, and *P. falciparum* periodically switches on or off different genes to express various products. In addition, *var* genes repeatedly recombine to engender new gene variants (Kyriacou et al., 2006; Guizetti and Scherf, 2013). Prokaryotic pathogens also contain intramolecular G-quadruplex motifs in key regions. For example, a short, conserved G-rich motif in the pilin expression locus of *Neisseria gonorrhoeae* plays a critical role in antigenic variation (Cahoon and Seifert, 2009; Prister et al., 2020). The exact role of these elements in

genetic variation and subsequent evasion of the host immune system is an ongoing endeavor.

Among pathogens, viruses have caught the attention of researchers interested in G4s for at least 25 years (Wyatt et al., 1994). Viruses are infectious agents that only replicate inside a host cell, and all require the translational machinery of the host to produce their proteins. They infect organisms from the three domains of life (bacteria, archaea, and eukaryota), but this review will be focused on human viruses. Viruses adopt a variety of shapes, and their genome may be either linear or circular, composed of single-stranded or double-stranded RNA or DNA nucleic acids (Flint et al., 2020). Recent reviews illustrate the interplay between G4 structures and virus functions (Métifiot et al., 2014; Ruggiero and Richter, 2018, 2020; Puig Lombardi et al., 2019). The high density of G-quadruplex sites in some viruses, such as herpesviruses, suggests that these motifs may be strategic components for their pathogenesis, virulence, and life cycle (Biswas et al., 2018; Ravichandran et al., 2018). There is a need for a critical assessment of G-quadruplexes as drug targets in viruses infecting humans. Bioinformatics studies led to the conclusion that noncanonical nucleic acid secondary structures, such as G-quadruplexes, are co-gent elements in the pathogenicity and viral proliferation of both RNA and DNA viruses. G4s have been detected using *in silico* approaches or stricter *in vitro* observations in several viral genomes and transcriptomes, such as HIV-1 (Amrane et al., 2014; Ruggiero et al., 2019; Tassinari et al., 2020) and HIV-2 (Krafcíková et al., 2017a), Epstein-Barr virus (EBV) (Norseen et al., 2009), hepatitis B virus (HBV) (Murat et al., 2014) and hepatitis C virus (HCV) (Jaubert et al., 2018), herpes simplex virus (HSV)-1 [also known as

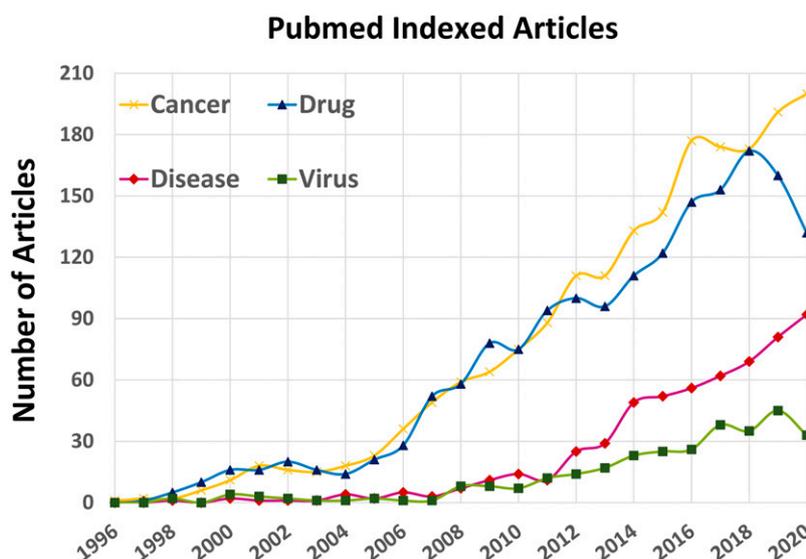


Fig. 1. Time-dependent increase in the number of published articles each year on quadruplexes (1996–2020 period). A dramatic increase is observed during the last decade, especially from the drug-design and discovery perspective. Each curve is drawn by using “G-quadruplex” AND (Boolean term) one of the four keywords listed in the top-left corner of the figure (e.g., G-quadruplex AND drug).

TABLE 1
G4-virus “Hall of fame”

Examples of G4 motifs in viral genomes (aptamers are not considered here) with relevance in antiviral research. A motif found in a host-cell promoter is also shown in this table because of its role in viral replication (Shen et al., 2020). In articles in which multiple sequences were reported, only one sequence was chosen among the highest-scoring motifs. Some motifs shown here correspond to a fragment of a larger DNA/RNA sequence studied in the corresponding article. Unless marked with an asterisk (*), the sequences below have been verified to form G4 structures experimentally. Sequences are ranked according to their G4-Hunter score (G4H): Different sequences from the same virus may be listed from different works.

Chem	Virus/Gene	Example of G4-Prone Sequence ^a	Length	G4H	Reference
DNA	KSHV	GGGGCGGGGGACGGGGGAGGGG	(22)	3.18	Madireddy et al. (2016)
DNA	HSV-1	GGGGTTGGGGTTGGGGTTGGGG	(22)	2.91	Artusi et al. (2015)
RNA	RVFV	GGGGTTGGGGGGAAGGGGAGTTGGGG	(27)	2.85	Charley et al. (2018)
DNA	HSV-2	GGGGACGGCGGGGCGGGGG	(20)	2.85	Frasson et al. (2019)
DNA	HSV-1	GGGGGGTGTGTTTTGGGGGGG	(21)	2.57	Biswas et al. (2018)
DNA	HSV-1	GGGGAGGGGAAAGGCGTGGGG	(21)	2.48	Frasson et al. (2019)
DNA ^b	TMPRSS2 (P)	GGAGGGCGGGGGGGCGGGGGCGGGCGGG	(29)	2.41	Shen et al. (2020)
DNA	HCMV	GGGGTCGGGAATGGGGGACAGAGCAGGGGGTATGGGG	(37)	2.16	Ravichandran et al. (2018)
RNA	HCV	GGGAGGGGGGUCCUGGAGG	(20)	2.05	Jaubert et al. (2018)
DNA	HIV-2	GGGGGGAGGACATGGGGCGGGAGGG	(25)	2.00	Krafčíková et al. (2017a)
DNA	Marburg	GGGGACTGGTTGGGGTCTGGGTGG	(24)	1.96	Krafčíková et al. (2017b)
DNA	VZV	GGGCGGGCGACGGGCGGG	(18)	1.83	Frasson et al. (2019)
DNA	HPV	GGCAGGGGACACAGGGTAGGG	(22)	1.82	Tlučková et al. (2013)
DNA	HTLV-2	GGGGAAGTGGGTAAGGGTGAGG	(22)	1.82*	Ruggiero et al., (2019)
RNA	HCV	GGGCGGGGUGGGCGGGA	(19)	1.79	Wang et al., (2016a)
DNA	Ebola	GGGGTGGTGTGGGGGTTGGG	(23)	1.74	Krafčíková et al. (2017b)
DNA	HPV	GGGTAGGGGACGGGACACAGGGT	(23)	1.74	Marušić and Plavec (2019)
DNA	Ebola	GGGGTCATATGGGAGGGATTGAAG	(25)	1.52	Wang et al. (2016b)
DNA	EBV	GGGGCAGGAGCAGGAGGA	(18)	1.50	Murat et al. (2014)
RNA	Zika	GGAUGGGCAGAGGGGGCUGGAG	(23)	1.43	Fleming et al. (2016)
DNA	HIV-1	GGGAGGCGTGGCTGGGCGGG	(21)	1.43	Perrone et al. (2013a)
DNA	HIV-1	TGGCCTGGGCGGGACTGGG	(19)	1.32	Amrane et al. (2014)
DNA	HBV	GGGAGTGGGAGCATTGGGGCCAGGG	(25)	1.28	Biswas et al. (2017)
RNA	SARS-CoV-2	CCCCAAAUCAGCGAAAUGCACCCC	(25)	1.28 ^{c*}	Bartas et al. (2020)
DNA	HPV	GGGTGGGTACAGGCGGACGACTGGG	(27)	1.11	Marušić et al. (2017)
RNA	Nipah	GUGCGGGGAGGUAAGAGGAGGCCAGG	(27)	1.11	Majee et al. (2020)
RNA	SARS-CoV-2	GGAUUGGCUUCGUAUGUCGAGGGG	(23)	1.04*	Panera et al. (2020)
RNA	SARS-CoV-2	GGCUGGCAAUGGCGG	(15)	0.87	Zhao et al. (2021)
RNA	SARS-CoV-2	GGUAUGUGGAAAGGUUAUGG	(20)	0.85 ^d	Ji et al. (2020)

^aSequences for which structural information is available are provided in the next table.

^bDNA sequence from a promoter in the human host cell.

^cCalculated for the complementary strand (this is a C-rich motif).

^dG4 formation was confirmed in vitro, but stability was low (Tm below physiologic temperature).

*The sequence is only computationally predicted to form G4 structure.

human herpesvirus (HHV)-1] (Artusi et al., 2015), human cytomegalovirus (HCMV) (Ravichandran et al., 2018); filoviruses, such as Ebola or Marburg (Krafčíková et al., 2017), Nipah virus (Majee et al., 2020), Zika virus (Fleming et al., 2016), simian virus 40 (SV40) (Tuesuwan et al., 2008), Kaposi's sarcoma-associated herpes virus (KSHV) (Madireddy et al., 2016), influenza viruses (Glazko and Kosovsky, 2013), Rift Valley fever virus (RVFV) (Charley et al., 2018); and coronaviruses, such as severe acute respiratory syndrome (SARS), Middle East respiratory syndrome (MERS) (Johnson et al., 2010), or SARS-CoV-2 (Bartas et al., 2020; Panera et al., 2020; Zhao et al., 2021). Examples of G-rich viral sequences are provided in Table 1. The recent COVID-19 outbreak evidently illustrates that more efforts are needed for the management of viral infections. The SARS-CoV-2 has ended the lives of more than 2 million individuals worldwide as of March 2021, and there is an urgent need for effective and valid treatment options for this pandemic.

As a few reviews already address the relevance of G4s in virology (Métifiot et al., 2014; Ruggiero and Richter, 2018, 2020), we have chosen to organize the manuscript in a different manner, which is centered around G4 type rather than virus classification (and only viruses

infecting humans are considered). In addition, G4s may be used as *probes*, *drugs*, *carriers*, or *targets* for antiviral studies (Fig. 2). Most of this review will be dedicated to the “G4 as *targets*” aspect, but we briefly illustrate here some of the three other uses.

- As *probes*, G4 structures, alone or complexed with Hemin—an iron-containing porphyrin structure—form DNzyme complexes that can be used to sense the presence of various ligands (Mergny and Sen, 2019).
- As *drugs*, G4s have often been used as aptamers, interacting with biomolecules, such as proteins, and interfering with their functions [for a recent review on aptamers, visit Romanucci et al. (2019)]. Some of the short nucleic acid sequences derived from the hexanucleotide TGGGAG motif, which is commonly recognized as “Hotoda's sequence,” are strong anti-HIV inhibitors (EC₅₀ = 14 nM). Such short sequences are also active on other viruses, as found for the hexanucleotide GGGGGT that forms a tetramolecular G4 structure, attaches to the C-terminal domain of hepatitis A virus protease, and is a strong inhibitor of hepatitis A virus 3C protease (Métifiot et al., 2014). One

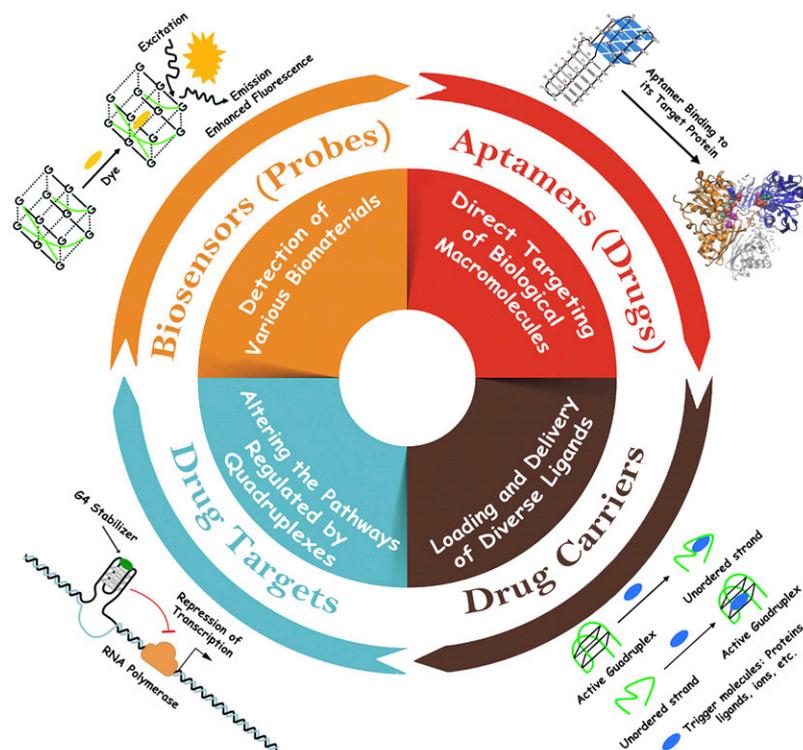


Fig. 2. Four main potential applications of G4s in antiviral research. See the main text for an illustration of these four aspects. Most of this review will be focused on the “G4 as drug targets” aspect, with a strong emphasis on human viruses.

notable argument is that these short sequences are actually too short to be specific and may also act on other cellular components of the host, which bind to noncanonical DNA secondary structures. Thus, their utilization as “drug-like” molecules should be further evaluated by selectivity and specificity tests.

- As *carriers*, one can exploit the ability of G4 structures to sequester various ligands and be used to deliver their cargo inside cells, thereby acting as drug delivery agents [for a recent illustration in cancerology, refer to Figueiredo et al. (2019)].
- Finally, as *targets*, one can exploit the ability of G4s to interact with specific ligands, which may perturb critical functions if the G4 is located in essential regions of the virus or host-cell genome. Whether G4 targets can be druggable for antiviral drug discovery is the main subject of this review.

2. A G-Quadruplex/Virus Timeline. Fig. 3 presents a timeline of G4 discoveries. Although the formation of supramolecular assemblies by guanines has been known for over a century (Bang, 1910), DNA G4s were first proposed by Gellert et al. (1962). After an initial spike of interest, G4s were not the focus of many studies until the ‘90s, during which several publications unraveled the presence of G4s in the human genome (Tasset et al., 1997; Fletcher et al., 1998). In addition, it was

found that one could design *structure*-specific rather than *sequence*-specific DNA ligands, and this result obtained on triplex DNA was later shown to be valid for quadruplexes as well (Mergny et al., 1992). As an added benefit, compounds able to interact with G4s were able to inhibit a key enzyme for the proliferation of cancer cells: telomerase (Sun et al., 1997). G4s were therefore viewed as novel targets for drug design (Mergny and Hélène, 1998).

The first study in which exogenous G-quadruplex structures were shown to inhibit viral infectivity was performed by Hotoda et al. (1998). Two years later, Tamura and colleagues (2000) found that a G, T-rich phosphorothioate oligonucleotide capable of forming a G4 structure prevented HIV infectivity by interfering with virus entry, reverse transcription (RT), and viral genome integration. The first viral protein shown to interact with these G4s was identified in 2008 by Tuesuwan et al. (2008) in SV40. They indicated the interplay between SV40 helicase and genomic G-quadruplexes, which are indeed necessary for proper unwinding and continuation of viral replication (Tuesuwan et al., 2008).

Research on G-quadruplexes in viruses has blossomed since then, finding relevance in a growing number of systems. Tan and colleagues (2009) found that the SARS-unique domain (SUD) of the nonstructural protein 3 (Nsp3) protein interacts with DNA and RNA G4s. Perrone et al. (2013a) and Murat et al. (2014) described G4s in HIV-1 and EBV as topological modulators of viral transcription and translation, respectively. Interestingly, the human nucleolin protein was found to have

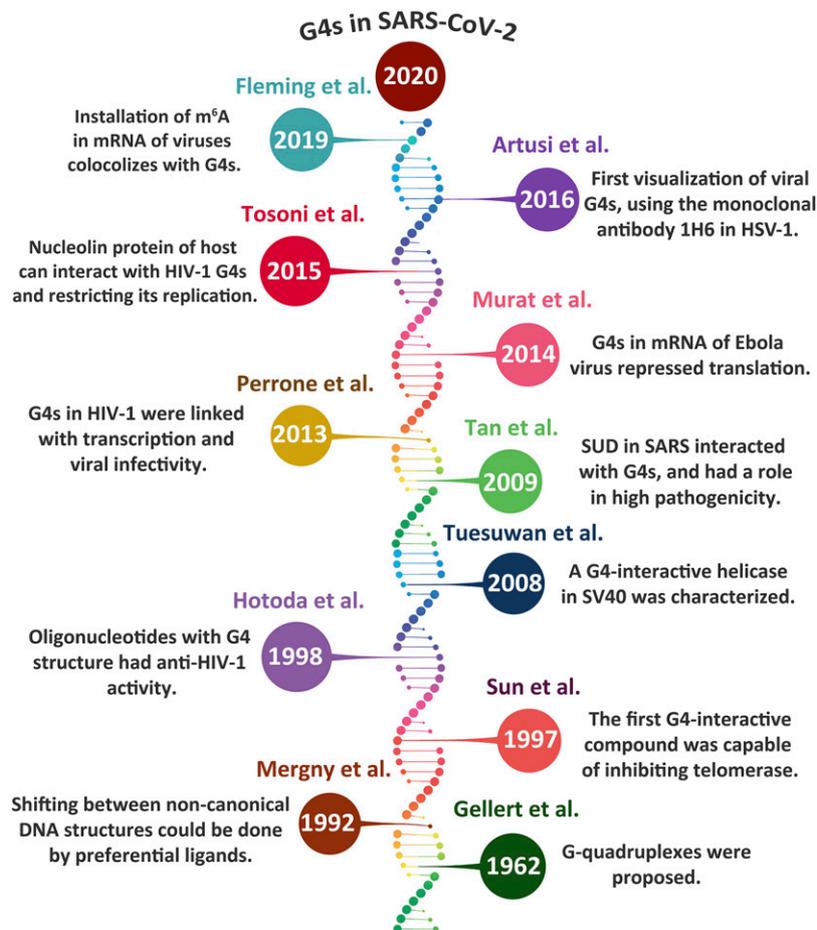


Fig. 3. G-quadruplexes have gone viral, as illustrated by a few key discoveries. Interest for G-quadruplex in viruses has soared remarkably after the first verification of these sequences in 2008 in SV40. Host proteins that interact with viral G-quadruplex structures are increasingly recognized, and the versatile roles of G-quadruplexes are being determined inside the infected cells.

anti-HIV-1 activity by binding to viral G-quadruplex structures. Tosoni et al. (2015) explained that this binding to long-terminal repeat promoter of HIV-1 constricts the transcription level and, ultimately, viral pathogenicity. By using the 1H6 antibody, Artusi and coworkers (2016) succeeded in visualizing the first G4 structure in viruses (HSV-1). Perhaps one of the most intriguing discoveries about the functional implications of G4s lies in the hand of post-transcriptional modifications, which are elaborated in a review article by Fleming et al. (2019). They argued that the sites of post-transcriptional methylation of adenosine residues to yield *N*⁶-methyladenosine (m⁶A) tend to be linked with G-quadruplex-prone regions. Unsurprisingly, enormous efforts are now being made to establish G4 relevance in SARS-CoV-2 (Fig. 3).

II. G-Quadruplexes and Viruses

1. DNA and RNA G-Quadruplexes. As introduced earlier, G-quadruplexes are relevant both at the DNA and RNA levels, with RNA G4s being often more stable than DNA G-quadruplexes. When a DNA sequence can fold into a G4 structure, its corresponding RNA motif

may also adopt a G4 topology (Krafčíková et al., 2017b). Consequently, a sequence involved in the regulation of transcription may also be involved in the regulation of translation or other mechanisms related to mRNA functions if the sequence is transcribed. Overall, the formation of G4 structures by both DNA and RNA sequences makes G-quadruplexes relevant for most viruses as long as their genome contains G-rich regions. Although this feature is attractive for antiviral agents—G4 motifs being targetable both as DNA and RNA—it may complicate the deconvolution of biologic effects of G4-based antiviral agents.

Over 200 G4 structures are currently available in the Protein Database (PDB) structure database. Unfortunately, only a few of them correspond to viral G-quadruplexes. Two solution NMR structures of DNA G4 motifs found in the HIV-1 long terminal repeat (LTR) are known, and one example is presented (Fig. 4) (De Nicola et al., 2016; Butovskaya et al., 2018). This HIV G-quadruplex offers specific epitopes for recognition and can be considered as “druggable.” The other solved structure by NMR belongs to the DNA motif of human papillomavirus type 52 (Marušić and Plavec, 2019).

G-quadruplex stability depends on a number of factors. Intracellular ionic conditions are favorable for G4 formation, and molecular crowding further stabilizes G4s (Matsumoto et al., 2020). Long G-runs allow the formation of more quartets, and this is often correlated with high stability.

Stability is not the only key parameter affected by the primary sequence. G-quadruplexes are inherently polymorphic, and different topologies can be formed (Phan et al., 2006), such as parallel, antiparallel, or hybrid conformations. Even left-handed G-quadruplexes may be observed (Bakalar et al., 2019). Additional levels of variability may be provided by loop-loop interactions, capping base triplets or base pairs, bulges, and non-canonical quartets [for example, a CGCG quartet (Lim et al., 2009)]. As a consequence, although most G-quadruplexes mostly or exclusively rely on the formation of the same elementary bricks, the G-quartets, they can each adopt a unique fold. This druggability feature is due to the intricate assemblage of guanine residues, which provide a patchwork of aromatic surfaces, clefts, and valleys as well as unique electrostatics. These properties enable the G-quadruplex to interact specifically with proteins or even smaller ligands, which can compete with proteins or mimic some of their roles. Obviously, nucleic acids, including G4s, do not offer equivalent targets as protein binding sites. Still, the observation that cellular and viral proteins as well as artificial and natural compounds selectively interact with these structures suggests that these G-quadruplexes are attractive targets for drug design.

2. Searching for G-Quadruplexes. Initial attempts to identify potential G4 motifs in a relatively small genome involved the manual search of GG-, GGG-, and/or GGGG-islands on the same strand. A variety of bioinformatics tools are now available to predict motifs susceptible to adopt a G4 structure. Putative G-quadruplexes were initially searched in the human genome by using computational approaches employing the sequence motif $G_{3+N_{1-7}G_{3+N_{1-7}G_{3+N_{1-7}G_{3+}}$, in which

G represents the guanine residues in each guanine tract, usually directly involved in G-tetrad formation, and N indicates any combination of loop residues (Huppert and Balasubramanian, 2005; Todd et al., 2005). The so-called Quadparser algorithm was first applied to the human genome with the abovementioned sequence motif as a query, and it was allowed to pick up over 300,000 candidate sequences. Since then, alternative search tools, such as G-Quadruplex forming G-Rich Sequences (QGRS) Mapper (Kikin et al., 2006) and G4-Hunter (Bedrat et al., 2016), have been designed [for a recent review, see Lombardi and Londoño-Vallejo (2020)]. Computational bioinformatics studies indicate that the density of G4s in viral genomes is moderately but not exclusively related to the G/C content of the viruses. A global analysis of human versus viral PQSs suggests that short-looped PQSs are more frequent and have a similar composition across viral taxonomic groups. Besides, there is a higher number of pyrimidine loops in viruses infecting animals irrespective of the viral genome type. Computational and statistical studies by Puig Lombardi et al., (2019) advocate the idea that the genome of viruses is rich in C-looped G-quadruplexes, possibly acting as a transcriptional binding site for the host transcription factors. The PQS density is 2- to 3-fold greater in viruses infecting vertebrate hosts than any other host (Puig Lombardi et al., 2019), implying a possible coevolution of these structures, as vertebrates also tend to be G4-rich.

Table 1 presents typical examples of G4 motifs in the virus genomes. We have chosen to rank these sequences according to the G4-Hunter score, which correlates with G4 propensity and stability, as discussed before. Although this ranking does not reflect the relative strength of G4 sequences in each virus, it illustrates that it is much easier to find a stable G4 motif in the KSHV genome than in the one of SARS-CoV-2. This table illustrates the differences in G-richness between motifs studied: Not all G4s are equal, and although some of them are probably so stable that

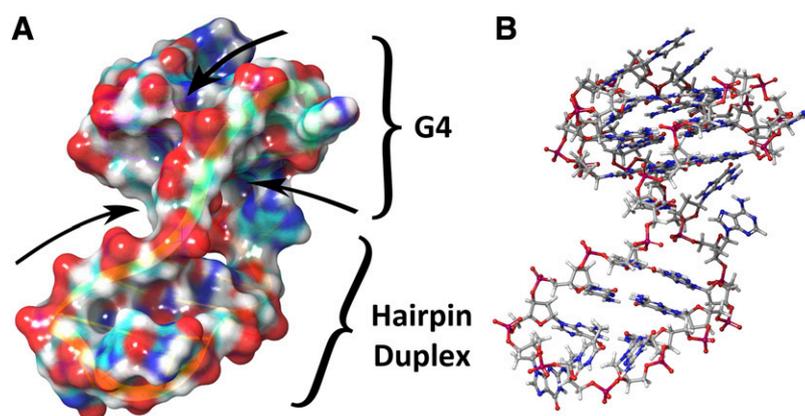


Fig. 4. The structure of HIV-1 LTR G-quadruplex (6H1K), elucidated by the Richter and Phan teams (Butovskaya et al., 2018). This motif involves both a G4 and a duplex part. (A) The surface of G4 displays various cavities, unlike the duplex DNA. (B) A ball-and-stick representation of the same structure.

dedicated helicases are required to unfold them, others are far more labile and may have different functions when folded or unfolded.

3. Viral G-Quadruplexes. Given that both DNA and RNA G-rich sequences are prone to G4 formation, it is not surprising that G4-motifs have been identified in nearly all viruses, including double-stranded DNA (ds-DNA), single-stranded DNA, and single-stranded RNA (ss-RNA) (e.g., *Retroviridae*) (Lavezzo et al., 2018). Among the viruses that contain definitive G4 structures in their genome, characterization of these structures in cancer-related viruses like KSHV, EBV, HBV, HCV, HIV, and human papillomavirus (HPV) deserves paramount attention (Saranathan and Vivekanandan, 2019). For a full review of the importance of G-quadruplexes in viral pathogenesis, one can refer to a recent review by Ruggiero and Richter (2020).

ds-DNA viruses, such as members of the *Herpesviridae* family, are especially enriched in G4 motifs, and most of the computationally identified G4-prone motifs were proven to be exceedingly stable under physiologic conditions. In DNA viruses, such as adeno-associated viruses and human herpesviruses, G4s modulate DNA replication, whereas G4s in the promoter region of HBV and the mRNA of EBV modulate transcription and translation, respectively (Ravichandran et al., 2018). Human herpesviruses are categorized into three sub-categories: alphaherpesviruses, betaherpesviruses, and gammaherpesviruses. PQS frequencies (predicted by Quadparser) are higher in modulatory regions of immediate early genes compared with early and late genes in most herpesviruses. HSV-2 has the highest number of PQSs ($n = 318$) among human herpesviruses, with PQS densities as high as 1.037/kb (i.e., 7-fold greater than the PQS density in the human DNA genome). The PQS densities of HSV-1 and HSV-2 are currently the highest reported for any genome ever sequenced, with the exception of some Archaea (Brázda et al., 2020). The repeat regions in herpesvirus genomes are important for the maintenance of the episomal form of the genome. PQS densities are significantly enriched within the repeat regions of herpesviruses. The telomeric-like motif “GGGTTAGGGTTAGGGTTAGGG” is repeated 81 times within a 3-kb region of the HHV-7 genome and a total of 204 times in the whole genome of this virus (Biswas et al., 2016).

An important feature of G-quadruplexes in some viruses, such as HSV-1, is that regions of the virus genome with higher G4 densities tend to be close to recombination breakpoints. Approximately 11% of breakpoints are located within a G4 motif, proffering these DNA secondary structures as hot spots for recombination regulation in the HSV-1 genome (Saranathan et al., 2019). Intriguing findings of the versatile functions of HSV-1 G-quadruplexes do not end here. Biswas et al. (2018) divulged that the packaging signal (*pac-1* signal) in the termini of HSV-1 genome is crucially involved in

the recognition of protein machinery required for the cleavage of the viral genome, its encapsidation, and virion assembly. Interestingly, the mouse monoclonal antibody 1H6 was shown to visualize G4 structures in host cells infected with HSV-1 by the aid of immunofluorescence and immune-electron microscopy (Artusi et al., 2016). G-quadruplex formation in the viral genome was found to be cell cycle-dependent and peaked at the time of viral DNA replication, traveled to the nuclear membrane at the time of virus nuclear egress (the export of viral capsids from the nucleus to the cytoplasmic area), and was later tracked in HSV-1 immature virions released from the host nucleus. However, this antibody is not truly G4-specific and has been reported to bind to other structures; thus its use for the detection of viral G4 can lead to ambiguous findings (Kazemier et al., 2017). A more specific G4 antibody, such as the now commercially available ScFv version of the BG4 antibody, would be recommended (Biffi et al., 2013). However, a recent study indicates that BG4 can also bind some cytosine-rich sequences on single-stranded DNA (Ray et al., 2020).

In RNA viruses, such as retroviruses, flaviviruses, and filoviruses, high variability may be observed, as evidenced by *Coronaviridae*, in which an astonishing intraspecies variance in G4s density was recently found (Lavezzo et al., 2018). In spite of their high genetic variability, all retroviruses besides ϵ -retroviruses contain highly conserved putative G-quadruplex-forming sequences in their promoter regions (Glazko and Kosovsky, 2013; Ruggiero et al., 2019). In fact, the G-rich clusters in LTR regions are strikingly conserved in all primate lentiviruses. Also, the majority of PQSs (~70%) are located in the U3 region just upstream of the transcription start site.

For RNA viruses, it is of quintessential importance to elude the host RNA decay machinery for survival. One of the characterized ways to accomplish this task is to target the host 5'-3' exoribonuclease 1 (XRN1). The 3' untranslated regions (3' UTRs) of several phleboviruses and arenaviruses contain RNA structures that block the ribonuclease activity of XRN1. RVFV, a member of the phlebovirus of the *Bunyaviridae* family, exploits this mechanism of action to evade RNA destruction. The 3'-terminal segment of the nucleocapsid (N) mRNA of RVFV can repress XRN1, and it is likely that this evasion is carried out by employing a G-quadruplex structure (Charley et al., 2018). Among viruses, RVFV is unusual in that it is composed of two negative-sense and one ambisense RNA during infection, generating transcripts from both strands. Zika virus, a member of the *Flaviviridae* family, displays seven PQSs that are markedly conserved within the genomes of >58 flaviviruses. Notwithstanding that these sequences are in general conserved between different species, their exact biologic role has remained nebulous (Göertz et al., 2018).

III. Roles of G-Quadruplexes in Viral Replication and Host Response to Viral Infection

A. Viral and Cellular G-Quadruplex-Binding Proteins

1. *Viral G-Quadruplex-Binding Proteins.* The role of G4s in the viral life cycle is becoming increasingly clear as viral proteins are found to have a G4-binding or unwinding activity. The first example was found in SV40, which bears a large multifunctional protein called T-antigen, which is capable of unwinding G4s (Plyler et al., 2009).

More recently, Rajendran et al. (2013) and Butovskaya et al. (2019) reported that HIV-1 nucleocapsid protein 7 (NCp7) binds to stable G4 structures in viral RNA and may promote its unfolding to assist in viral reverse transcription. The discovery of G4 chaperones or helicases in viral genomes is very promising and also implies that the regulation of viral processes, such as reverse transcription or replication, is so vital that viruses have implemented several regulatory elements like G4 and protein chaperones to supervise these steps. NCp7 has an impressive affinity (K_d around 10 nM) for the central DNA flap region [central polypurine tract (cPPT) region] of HIV-1, which may fold into an intermolecular parallel DNA G4 (Lyonnais et al., 2003). Potassium ions and a dibenzophenanthroline derivative coined as MMQ3 were able to stabilize these G-quadruplexes (Lyonnais et al., 2002), hinting that capsid assembly and packing of the viral genome may also be affected by G4 formation.

2. *Cellular G-Quadruplex-Binding Proteins.* In addition to viral proteins, several host G4-binding proteins have been shown to regulate virus life cycles. More cellular G4-interactive proteins are characterized mainly because of their relevance in various diseases that are linked to G4s, especially cancer.

Nucleolin has been implicated in several pathologic processes, such as tumorigenesis (in which lots of G4 structures in the promoters of the oncogenes are defined) and viral pathogenicity. Nucleolin binds to G4 structures found in the HIV-1 LTR promoter as well as in host promoters, such as *c-myc*. In addition, nucleolin may also be recruited to G4 present in aborted RNA transcripts generated from the hexanucleotide repeat motif (GGGGCC) $_n$ as well as to the Epstein-Barr virus-encoded nuclear antigen (EBNA) 1 mRNA. Indeed, nucleolin acts as a vital host protein capable of binding to the viral RNA G4 structure; this protein can stabilize these structures and suppress viral replication (Bian et al., 2019).

Another case of cellular G4 binding protein is nucleophosmin (NPM1) that augments the infectivity potential of adeno-associated viruses (Satkunanathan et al., 2017). NPM1 stabilizes the G4 structures in the *c-myc* oncogene and is also implicated in the repression of cellular as well as viral replication. It has been suggested that this protein directly interacts with the G4

motifs to mobilize the packaging of viral genome and encapsidation. Our current understanding of the interaction between NPM1 and viral G4 is far from complete, but knockdown of this protein results in the increased replication of viral DNA and localization of packaged adeno-associated virus particles in the cytoplasm. It also seems that NPM1 stabilizes G4 structures in viral genome and thus creates a blockage in the replication process, which is contradictory to its cooperative role in viral packaging and pathogenesis (Satkunanathan et al., 2017).

Heterogeneous nuclear ribonucleoprotein families have also been revealed as unfolding agents of viral G-quadruplex structures. Among this family, heterogeneous nuclear ribonucleoprotein A2/B1 untwists the G4 structure in the LTR of HIV-1 and boosts the level of transcription (Scalabrin et al., 2017). Stabilization of these structures by stabilizing compounds can halt viral transcription without seriously affecting telomere G4 stability (Perrone et al., 2015).

Finally, the SUD domain of SARS Nsp3 protein can interact with G4 structures, and this interaction has been proposed to regulate SARS transcription/replication (Kusov et al., 2015). This regulation and its potential use as a therapeutic target will be discussed in the chapter dedicated to coronaviruses.

B. G-Quadruplex Ligands as Antiviral Compounds

Currently, more than 1000 characterized G4 ligands have been characterized (Li et al., 2013), and we provide examples of G4 ligands tested for their antiviral properties (Fig. 5). Nearly all of them have been reported to stabilize these structures, whereas a few may destabilize them.

5,10,15,20-Tetrakis(1-methylpyridinium-4-yl)porphyrin tetra(p-toluenesulfonate) (TMPyP4) is one of the most studied G4 ligands in antiviral research. Although it binds with a high affinity to G-quadruplexes, it is weakly selective for G4 over ds-DNA (Monsen and Trent, 2018). The same problem may be found for other G4 ligands [e.g., groove-binding ligands, such as distamycin A and netropsin (Randazzo et al., 2002)], which fail to distinguish between G4 and other nucleic acid structures. This failure is a major issue, as it may lead to off-target effects and/or a decrease in the efficiency of the ligand.

More disturbing are results showing a destabilization of G4 structures by G4 ligands. This is, for example, the case of TMPyP4, which demonstrated antiviral activity against pseudorabies virus (the causative agent of Aujeszky disease in pigs), possibly by destabilizing a G4 located in 3'-UTR of *IE180* gene (Zhang et al., 2020). TMPyP4 is alleged to suppress this gene expression and prevent proper replication of the virus, but the causality between this destabilizing effect and the inhibition of viral replication remains to be established. Owing to the fact that TMPyP4 is weakly selective and

able to interact with a large spectrum of nucleic acid structures, stabilization of yet unidentified or uncharacterized targets may also be responsible for the antiviral effect.

Fortunately, most current G4 ligands now exhibit a fair selectivity for G4s over duplexes. The acridine derivative BRACO-19, a potent G4 stabilizer, exerts its anti-HIV-1 activity by stabilizing HIV-1 G-quadruplexes, abating the viral transcription process and repressing reverse transcription (Perrone et al., 2014). Since BRACO-19 also exhibits antiproliferative activity (Zhou et al., 2016), one may speculate that its effects on the host cell contribute to the antiviral effect. The bisquinolinium phenanthroline derivative Phen-DC3 may be considered a prototypical example of a high-affinity G4 ligand with little or no binding to duplexes (De Cian et al., 2007). Even cationic porphyrins may exhibit an exquisite selectivity for G4s (Dixon et al., 2007). It is now common practice to evaluate in parallel G4-binding and duplex-binding activity during ligand screening so that poorly selective compounds are discarded at an early stage. Even then, within the same chemical family one can notice profound differences in selectivity. For example, among core-extended naphthalene diimides, a right trade-off between affinity and selectivity can be found: Mitigating the affinity of the binding core of these naphthalene derivatives could result in increased selectivity for G4s. Instead of screening relatively large molecules, one can also perform a fragment-based methodology, as deployed by Tassinari et al. (2018), to find HIV LTR G4 ligands. The authors identified tetra hetero aryls compounds with an aromatic amidoxime central core that preferential bind to the HIV LTR G4 over human telomeric G4s. The amidoxime functional group is the prodrug of amidine in drug design, further enhancing their drug-likeness potential (Tassinari et al., 2018; Zuffo et al., 2018).

A far more ambitious aim in terms of selectivity would be to bind to *one* G-quadruplex only. This “perfect” ligand would not only be able to distinguish G4 from duplexes but also among G-quadruplexes. Unfortunately, although one can find G4 ligands having a higher affinity for a given topology [e.g., parallel (Zuffo et al., 2018), or antiparallel (Hamon et al., 2011)], reports of compounds binding to one specific G4 sequence/structure are scarce. Most G4 ligands display a very limited selectivity between G-quadruplexes (Tran et al., 2011). As Oskouie and Abiri pointed out (Amjadi Oskouie and Abiri, 2021), a limited selectivity between G4 may still be helpful, as the genes that are regulated by G-quadruplex structures are often involved in related pathways. Thus, broad stabilization of all these processes, which results in the inhibition of all these pathways, may be more effective than selective stabilization of a single G4.

HCMV (commonly referred to as CMV, also known as HHV-5), results in life-threatening infections in newborns and patients who are immunocompromised.

A battery of PQSs have been predicted in its genome, and *in vitro* characterization confirmed their existence. Both TMPyP4 and *N*-methyl mesoporphyrin IX (NMM) stabilize these structures and significantly suppress the transcription of the related promoters, but NMM was found to be clearly superior to TMPyP4 (NMM being a far more selective G4 ligand than TMPyP4). Moreover, NMM unlike TMPyP4 caused suppression in the growth of HSV-1. To complicate things, not all G-quadruplexes in the genome of HCMV are associated with the inhibition of transcription, and suppression of gene expression was not directly related to the thermal stability of G-quadruplexes. The G4s that were responsible for arresting transcription were involved in the infectivity potential of the virus (Ravichandran et al., 2018). This study addresses the presence of highly stable G-quadruplexes in the

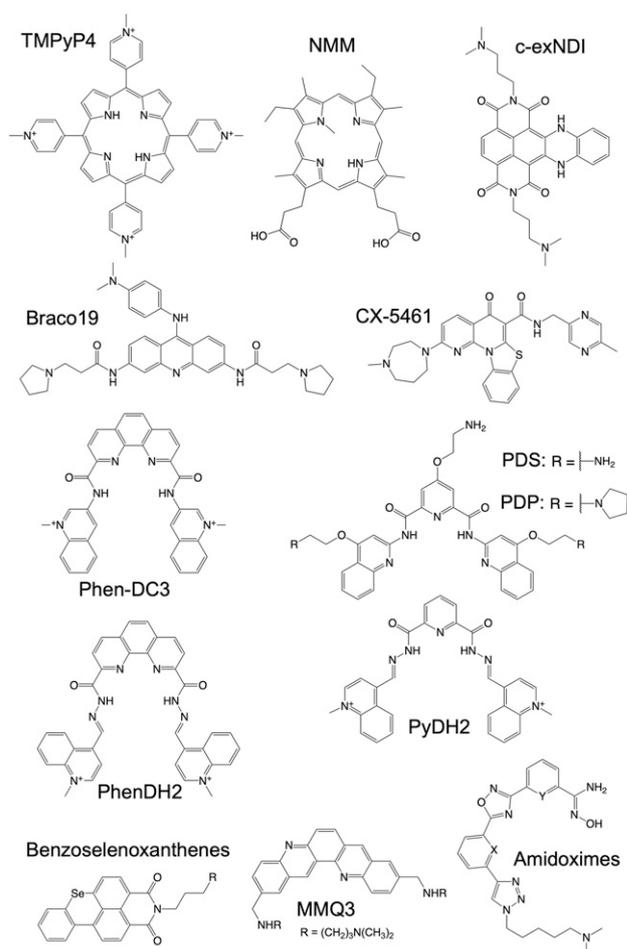


Fig. 5. Structure of G4 ligands tested for virus G4 binding. Two porphyrin derivatives are presented: TMPyP4 and NMM. Braco19 is a trisubstituted acridine, whereas c-exNDI is a core-extended naphthalene diimide. Phen-DC3, PDS, and PDP are bisquinolines/bisquinolinium derivatives. PyDH2 and PheDH2 are cationic bis(acylhydrazones), described by Granzhan and colleagues (Reznichenko et al., 2019). (F) CX 5461 is a G4 ligand with promising antitumoral activity (Xu et al., 2017). Amidoxime derivatives (X = N and Y = CH, or vice versa) were described by Richter and colleagues (Tassinari et al., 2018). Benzoselenoxanthene derivatives were developed by Zhang and colleagues (Shen et al., 2020). MMQ3 is a mono derivative (R = (CH₂)₃N(CH₃)₂) (Lyonnais et al., 2002).

promoter regions of a virus for which stabilization was not quantitatively correlated to the level of transcription suppression.

Even though G4s have more often been proposed to repress the expression of their target genes, a G-quadruplex-mediated increase in promoter activity has sometimes been reported (Fernando et al., 2009; Lam et al., 2013; Wei et al., 2013). At the transcriptome-wide level, G4 ligands can actually lead to an increase in transcription of a number of human genes (Beauvarlet et al., 2019a). For viral genes, the G4-prone motif found in the promoter region of the HBV *preS2/S* gene upregulates its expression. BRACO-19 and pyridostatin (PDS) both exhibited specific binding and stabilization of this G-quadruplex and induced viral gene expression (Biswas et al., 2017).

Finally, and as discussed in later sections, G4 ligands may also have an antiviral effect by interfering with the transcription of genes from the host cell. CX-5461 [an orally bioavailable naphthyridine carboxamide derivative (Drygin et al., 2011)] antagonizes PolII-mediated ribosomal DNA transcription by attaching to G4 DNA structures regulating this process. Interestingly, the addition of CX-5461 at both early and late stages of HCMV infection impedes viral DNA synthesis (Westdorp and Terhune, 2018).

Interestingly, even if not truly specific for one topology, some ligands may have different affinities for some conformations and shift the equilibrium toward the topologies they prefer. This opens up a question: are specific G4 structures or topologies easier to target than others? Some indications come from the binding profile of known G4 ligands: a number of them bind to parallel G4s with higher affinity, and some compounds, such as NMM or core extended naphthalene diimide (c-exNDI), derivatives can act as specific light-up probes for parallel structures (Zuffo et al., 2018).

C. G-Quadruplexes Interfering with Different Viral Steps: Human Immunodeficiency Virus-1 as a Model

In the teeth of evidence defending the proposition that DNA viruses have a stable and functional role in the viral cycle and pathogenesis, most of what we know comes from studies on an RNA virus, HIV-1. A very large fraction (possibly one-third) of articles on G4 in viruses come from HIV for at least four reasons: 1) This virus has been known for almost four decades, giving a strong head start over more recent threats, such as coronaviruses; 2) there is an absence of a true cure for patients with AIDS, arguing for continued efforts to fight this disease, especially “dormant” or latent viruses; 3) there is the presence of attractive and versatile conserved G4-prone motifs in the HIV genome; and finally, 4) some proteins important for HIV pathogenicity are vulnerable to G4 drugs (aptamers and, more generally, G4-forming sequences, such as Hotoda’s hexamer). This interest in HIV has been beneficial for researchers

working on G4 as antiviral strategies, and experience gained in fighting AIDS is invaluable. Nevertheless, one should remember that other viruses, including coronaviruses, may have unique ways of dealing with G-rich sequences that have no equivalent in retroviruses.

The presence of conserved G4 motifs in key regions of viruses, and especially retroviruses, such as HIV, argues that these G4s are relevant either at the RNA or DNA level and interfere with reverse transcription and transcription, respectively. Studies suggest that viral G-quadruplexes can act as *cis*-regulating elements in the vicinity of the genes involved in gene expression regulation to adjust the gene expression of a variety of genes (Li et al., 2013; Puig Lombardi et al., 2019). Notably, all primate lentiviral LTR regions contain conserved binding sites for specificity protein 1 (Sp1) and nuclear factor κ B (Ruggiero et al., 2019). Both transcription factors (TFs) have a consensus binding site involving several consecutive guanines; as a consequence, when multiple TF binding sites are juxtaposed, the sequence becomes compatible with G-quadruplex formation and can be occupied by G4 ligands. It remains to be established whether the near-perfect conservation of these G4-prone motifs is the direct consequence of G4 physiologic roles in HIV transcription, or the indirect consequence of the conservation of TF-binding sites.

The involvement of G4s in the HIV-1 cycle is not limited to transcription or reverse transcription:

- The HIV-1 genome is composed of two identical RNA molecules held together in a parallel orientation at a so-called dimer initiation site. Short synthesized RNA oligonucleotides with HIV-1 sequences containing the Sp1-binding sites can also dimerize, and their interactions have characteristics of the intermolecular G-quadruplex conformation. This suggests that the U3 region is an additional contact point in a multiplying linked genome dimer and together with G-rich sequences in *gag* and cPPT helps to retain interactions of the HIV-1 genomes along their whole sequence.
- Other G4 structures have also been described in HIV-1. The negative regulatory factor (*Nef*) is a small and conserved protein among lentiviruses, which is expressed in the early stage of viral pathogenesis and is necessary for viral replication and infectivity. Apart from replication, *Nef* aids in the intracellular sequestration and degradation of CD4 and major histocompatibility complex I expression on the surface of infected cells, and by this virtue, it restricts its recognition by host immune cells. G-rich-prone sequences in the *Nef* gene have been shown to fold into G4 structures, and BRACO-19, PIPER (N,N'-bis[2-(1-piperidino)ethyl]-3,4,9,10-perylene-tetracarboxylic diimide), and TMPyP4 were all

able to stabilize them (Perrone et al., 2013a). TMPyP4 displayed reduced infectivity in the antiviral assays against HIV-1 in a dose-dependent manner at concentrations (0.1–6 μM) that no significant toxicity was detected for the host cells (Perrone et al., 2013b). All three ligands are positively charged with a large, flat aromatic surface. This explains why these compounds are also able to recognize other G-quadruplexes, such as the ones found in HIV LTR (Perrone et al., 2013a).

D. Diversity of Regulatory Mechanisms

As shown previously with HIV-1, G4s can act at different steps of the viral cycle. Indeed, G4-based modulation can be found during:

- *Transcription*: This is the most prevalent and well known function of G4s in viral context, which is discussed thoroughly in previous parts. The effect of transcriptional regulation can be mediated by other “indirect” means that could have therapeutic relevance. For instance, in human genes, epigenetic modulation can affect the stability of G4s and, therefore, it finally determines the level of gene expression. Epigenetic modifications in DNA can both decrease or increase the stability of G4s (Reina and Cavalieri, 2020). Moreover, G4s are known to be widespread in hypomethylated DNA regions in the host, and they recruit and inactivate DNA methyltransferase enzymes (Mao et al., 2018). Hence, we have two mechanisms to regulate the chromatin packaging that seem to inversely correlate with each other (Reina and Cavalieri, 2020). For viruses, this has not been fully confirmed, but it is tempting to consider the idea that viral G4 structures along with cellular DNA methylation patterns might manipulate cellular gene expression patterns, especially for those viruses that have latent and lytic phases and insert their DNA into the cellular DNA (proviruses) like HIV. Besides, it has been established that G4 structures are intimately linked to chromatin accessibility and its packaging (Jara-Espejo and Line, 2020). This accessibility is closely associated with the probability of viral DNA insertion in a specific site. For example, the lens epithelium-derived growth factor is known to interact with HIV integrase and direct the insertion toward the transcriptionally active areas. Depletion or inhibition of this protein via small molecules led to insertion out of the active transcriptional units. Lens epithelium-derived growth factor interacts with open chromatin by its PWWP domain (a domain containing the conserved Pro-Trp-Trp-Pro amino acids) (Vansant et al., 2020). Although there is no
- solid evidence to sufficiently support this claim, the epigenome and G4 pattern of the cells may either restrict or accelerate the insertion in certain regions (depending on the lysogenic or lytic phase). In this regard, G4s can be excluded from the core structure of nucleosomes (and in a reciprocal fashion exclude nucleosomes), and this provides a possibility of acting as an anchor for the recognition of the host genome by viral proteins, which could ultimately lead to viral genome integration. Thus, manipulation of the chromatin packaging to modify the epigenome or G4 patterns of the host by using drugs may affect viral latency and its survival (Ruggiero and Richter, 2020).
- *Splicing*: HPV, nonenveloped viruses containing an episomal DNA, the presence of G4s in the sequences coding the L2 protein (HPV57), E1 (HPV32, HPV42), and E4 (HPV3, HPV9, HPV25), implies that G4 formation may also modify alternative splicing processes required to produce viral proteins from overlapping open reading frames (ORFs) (Métifiot et al., 2014). G-quadruplex-prone regions are located in the long control region (LCR), L2, E1, and E4 regions of the HPV genome and are likely to be involved in the gene expression by serving as a binding site for host transcription factors (Tlučková et al., 2013; Marušič et al., 2017).
- *Exonucleases*: G4 may also act as a suppressor of the exonucleases of the host in a manner similar to what was proposed for RVFV, but no study has yet uncovered their precise molecular and cellular mechanisms.
- *Extrachromosomal epigenome*: Dabral et al. (2020) recently established the stabilizing effect of TMPyP4 on the latency-associated nuclear antigen (LANA) of KSHV. LANA is the most abundant protein during the latency phase and is crucial for the persistence of KSHV in the host cells. The terminal repeat (TR) region of KSHV is a GC-rich DNA element that embodies a primary origin of latent DNA replication site and is crucial for the persistent maintenance of the viral episome in the proliferating host cells. Sequence analysis of the TR region revealed several potential PQSs. KSHV creates a life-long latent infection preferentially in B-lymphocytes by remaining as an extrachromosomal episome in the infected cells and by sustaining its genome in the dividing cells. KSHV attains this by tethering its epigenome to the host chromosome using LANA, which binds in the TR region of the viral genome. Madireddy et al. (2016) reported that Phen-DC3 is able to stabilize the KSHV G-quadruplex structures, increasing the number of unfinished replication forks in its genome and successful replication and

- hindering the progression of the cell cycle (maintaining latency). At this concentration (10 μ M) TMPyP4 was unable to cause a marked increase in the number of unfulfilled replication forks, and therefore, no significant inhibition of replication was observed (Madireddy et al., 2016). Again, one should remember that TMPyP4 is a G4 ligand of moderate selectivity, which also binds to duplexes, triplexes, and other structures and has been reported to both *stabilize* and *destabilize* G4 structures! For these reasons, extrapolating results obtained with TMPyP4 to more selective G4 ligands may not be straightforward (Fujiwara et al., 2015).
- *RNA replication*: Despite the high genetic variability of riboviruses, bioinformatics investigations unveiled a highly conserved G-rich consensus sequence in the HCV C genomes. G4 ligands were shown to have anti-hepatitis C activity, acting by reducing RNA replication and inhibiting protein translation of intracellular hepatitis C virus. PDP (a bisquinolinium derivative) and TMPyP4 led a decreased expression of the HCV C gene (Wang et al., 2016a). Phen-DC3, a potent and specific G-quadruplex binder (De Cian et al., 2007), can prevent HCV replication in cells in conditions in which no cytotoxicity was observed (Jaubert et al., 2018). RNA synthesis was curtailed in the presence of potassium or Phen-DC3, which both stabilize the RNA G4 structure in HCV. In light of the fact that the last 157 nucleotides of the 3' end of the HCV (–) strand are highly conserved between diverse HCV genotypes, G4 ligands could be of significance for novel (broad-spectrum) anti-HCV medications.
 - *RNA modifications*: One of the modifications in the mRNA structure is methylation at adenosine residues, yielding m⁶A nucleotides. The m⁶A modification in mRNA is implicated in various cellular pathways, including splicing, nuclear export, translation productiveness, and decay of the redundant strands. The presence of m⁶A residues within the loops of two-tetrad PQSs in the RNA genomes of the Zika, HIV, hepatitis B, and SV40 viruses has been established (Fleming et al., 2019). The conserved viral PQSs may provide a framework (based on sequence and/or structure) for m⁶A installation, whereas m⁶A modifications may either favor or disfavor the folding of a sequence into a G-quadruplex depending on the context. Facilitating G4 formation is often linked to a drop in gene expression and latency of viral pathogens. On the opposite end, the unfolding of G4 structures may facilitate the replication of viruses. Based on previous observations, Fleming et al. (2019) thus suggested assessing the role of m⁶A installation in a case-by-case analysis.
 - *Effects on translation*: Viruses can use these G4s at the RNA level to modulate the translation of different proteins and thereby restrain the presentation of these proteins on the major histocompatibility complex molecules to remain latent in the host cells until suitable conditions happen (Harris and Merrick, 2015). EBV can be seen as a prototypical example, as it remains in latent stages in some individuals, and G-quadruplexes can be crucial for this by quashing antigen presentation. EBV is linked to a number of cancers, such as Burkitt lymphoma, nasopharyngeal carcinoma, and 10% of gastric cancers. PDS has been used to investigate the role of G-quadruplexes in EBNA1 mRNA, wherein it led to a decrease in the production of the EBNA1 protein in a concentration-dependent way both in vitro and in vivo (Ruggiero and Richter, 2018). EBNA1 is highly immunogenic (Lista et al., 2017) and implicated in viral replication; the G4 structure in its mRNA is meticulously linked to viral latency (Canaan et al., 2009). Interestingly, cationic bis(acylhydrazones) were also shown to interact with the glycine-alanine repeat-encoding sequence of the EBNA1 mRNA and thereby increase the expression of this mRNA (Reznichenko et al., 2019). The authors attributed this increase to inhibition of nucleolin binding to the EBNA1 mRNA, which ultimately led to the disinhibition of translation. This disinhibition resulted in an increased antigen presentation of EBNA1 and suppression of immune evasion. Yet the significance of this disinhibition in the in vivo context requires further studies. The effect of PDS on EBNA1 expression was interrogated by Lista et al. (2017), who found that this ligand is only capable of weak interactions with the EBNA1 G-quadruplex. They suggest that unlike PDS, which was unable to affect the level of EBNA1, Phen-DC3 prevented nucleolin binding to EBNA1 mRNA and reversed glycine-alanine repeat-mediated repression of EBNA1 expression to restore viral antigen presentation (Lista et al., 2017). Destabilization of the EBNA1 G-quadruplex with targeted mutations led to an increase in antigen presentation resulting in the activation of virus-specific T-cells (Tellam et al., 2014). Additionally, Reznichenko et al. (2019) suggested that chemical scaffolds based upon pyridine, naphthyridine, or phenanthroline are remarkable EBNA G4 binders, whereas pyrimidine-based scaffolds displayed impoverished binding affinity. PyDH2 and PhenDH2 (both are bisquinolinium derivatives) are two newly identified G4 ligands, which induce a marked increase in EBNA1 gene expression and are significantly less cytotoxic than the historically discovered Phen-DC3 compound (Reznichenko et al., 2019).

Another report on the use of BRACO-19 against EBV stated that this compound stabilizes EBNA1 mRNA G4. EBNA1 sequesters the cellular origin recognition complex by using RNA-dependent interactions with two well established domains of EBNA1 called EBNA1 linking region 1 and linking region 2. In this way, BRACO-19 prevents the replication of EBV and exerts its antiviral activity (Norseen et al., 2009).

- *Enhanced genetic variability in some genes:* For example, in influenza, variations in *hemagglutinin* compared with the *neuraminidase* gene may have resulted from the differences in the density of noncanonical DNA structures in these genes (Glazko and Kosovsky, 2013). An increased G4 density in the genes whose products bind to the target cell receptor system may benefit the virus and contribute to the genetic variability required for the competitive interactions in the host-pathogen system (Glazko and Kosovsky, 2013). This raises the interesting possibility of the involvement of G-quadruplexes in mutation enhancement and recombination, which is not yet confirmed.
- *Latency:* HIV-1 latent reservoirs were found to be susceptible to G4 ligands, which could be eliminated by inducing apoptosis without virus reactivation (Piekna-Przybylska et al., 2017). Cells that are infected with latent HIV-1 provirus exhibited altered telomere maintenance mechanism and were vulnerable to G4 ligands. Piekna-Przybylska et al. (2020) revealed that Sp1 binding to HIV-1 promoter was prohibited by the presence of G4 ligand TMPyP4 in vitro but not in cells. In that study, the researchers demonstrated that the G4 ligands TMPyP4 and BRACO-19 can be concomitantly used with latency-reversal agents (i.e., compounds that wake up latently infected cells, such as vorinostat and bryostatins) to observe synergistic effects on the elimination of infected cells with provirus reservoirs. Sp1 in HIV-1 is responsible for basal transcription (Turrini et al., 2015), and inhibition of its binding by G4 stabilization is associated with latency, manifesting the same scenario for G4s in turning on latency as seen for HSV-1, EBV, and KSHV.

A summary of the G-quadruplex roles in the pathogenesis of viral diseases is summarized in Table 2.

E. G-Quadruplexes in the Host Cell as Alternative Targets?

G4-prone motifs are not only found in the genomes of viruses but are also present in the genome of the human infected cells. These “cellular G4s” can be relevant for viral research, as they can be the target of viral processes. This may be especially interesting in

the case of oncogenic viruses, as G4 ligands, which stabilize both viral and host G-quadruplexes, would serve as two-pronged tools in which both the proliferation of the cancerous cells and the replication of viral agents can be neutralized. Some proteins of viral origin bind to G-quadruplexes, and it often remains to be established whether their relevant nucleic acid partners are of cellular or viral origin. In a reciprocal fashion, host-cell proteins binding to viral G-quadruplexes may be relevant for the virus. There are four main aspects that can be investigated for employing host G4s for antiviral therapies: 1) immune modulation, 2) oncogenic viruses, 3) cellular pathways, and 4) telomeric integration. In this category, for the first approach (immune modulation), direct evidence of efficacy is lacking, but for others, G4 ligands are effective at least to some degree in mitigating viral spread and disease symptoms (Fig. 6).

1. Immune Modulation. To address the role of host G-quadruplexes in viral pathogenesis, we first need to clarify the function of these G4s for the host cell. Oncogenic promoters, telomeres, introns, and both 5' and 3' UTRs of mRNAs are the most well characterized locations where G-quadruplexes have been reported in the human genome and transcriptome (Carvalho et al., 2020). Among RNA, G4 can be found in noncoding RNAs as well. Studies have revealed that a type of long noncoding RNAs transcribed at telomeres called “Telomeric Repeat-containing RNA” is actively engaged in the mechanisms orchestrating telomere sustenance and chromosome end sheltering (Bettin et al., 2019). Telomeric RNA/G4-forming sequences suppress the expression [*STAT1* (*signal transducer and activator of transcription 1*), *ISG15*, and 2',5'-oligoadenylate synthetase (*OAS3*)] of the innate immune system in three-dimensional cultures. This suppression was similarly triggered by the nontelomeric G-rich DNA aptamer AS1411. Both TELomeric Repeat-containing RNA and AS1411 fold into G4 structures, which inhibit the induction of specific innate immune genes in cancer cells (Hirashima and Seimiya, 2015). We first need to elaborate the underlying cellular pathways involved in these genes—*STAT1*, IFN-stimulated gene (*ISG*) 15, and *OAS3*—to assess the druggability of these host sequences:

- *STAT1* is a key target of IFN- γ , acting as a transcription factor of various genes associated with antiviral proteins and enzymes, microbicidal molecules, phagocytosis-related receptors, cytokines, chemokines, inflammatory pathways, and also antigen-presenting molecules (Hu and Ivashkiv, 2009).
- *ISG15*, encodes a ubiquitin-like protein induced mainly by type I interferons and viral infections. This protein is conjugated to numerous cellular proteins, a process styled as ISGylation. A plethora of proteins involved in antiviral signaling,

TABLE 2
G-quadruplex roles in the pathogenesis of viral diseases (viruses are sorted according to the Baltimore classification, which is based on the process of mRNA synthesis and could be more relevant for G4 effect)

Virus (Name, Family, Genome)	Related Gene/Protein	Role	Reference
HSV-1 (HHV-1) <i>Herpesviridae</i> ds-DNA	a. 500-bp regions flanking the recombination breakpoints b. Replication compartments c. <i>Pac-1</i> (near genome termini) d. IE genes IE genes IE genes EBNA1 mRNA Various genes (<i>UL34</i> , <i>UL35</i> , <i>UL82</i> , <i>IRS1</i> , <i>US30</i> , <i>UL75</i> and <i>UL76</i> , <i>TRSI</i> , etc.) Direct repeat regions of <i>pac-1</i> (and probably <i>pac-2</i>) and TMRs a. LANA mRNA b. TR region of DNA <i>LCR</i> , <i>L1</i> , <i>E1</i> , and <i>E4</i> A repeat sequence, which plays important role in viral encapsidation and also the coding region of several genes (<i>prM</i> , <i>E</i> , <i>NS1</i> , <i>NS3</i> , <i>NS5</i>) 3' and 5' of UTR and other sites a. Stem-loop Ily' of 3'-end of the negative strand b. HCV <i>core</i> (<i>C</i>) gene ORF1 ab, spike (<i>S</i>), ORF3a, membrane (<i>M</i>), and nucleocapsid (<i>N</i>) genes (in SARS-CoV-2) <i>G</i> and <i>L</i> genes, encoding the cell attachment glycoprotein and RNA-dependent RNA polymerase <i>L</i> gene Hemagglutinin and neuraminidase genes 3'-Terminal of the nucleocapsid (<i>N</i>) mRNA	a. The recombination breakpoints of HSV-1 are placed adjacent to G4 motifs. G4 clusters are recognized as hot spots for recombination in HSV-1 b. G4 structures are likely to be involved in the latent phase of viruses with lytic and latent phases c. Packaging and cleavage of virus and DNA maturation for encapsidation d. Inhibition of transcription Multiple stable G4s in both leading and lagging strand were detected, which inhibit transcription Multiple stable G4s in both leading and lagging strand were detected, which inhibit transcription Regulation of translation. Stabilizing this RNA G4 can be linked with the repression of translation and antiviral activity Inhibition of transcription. NMM ligand stabilized these G4s and prevented gene expression. Not all G4s in the genome of the virus are involved in the regulation of transcription Unknown, probably involved in homologous recombination for TMRs; for <i>pac-1</i> the function is uncharacterized but could be similar to <i>pac-1</i> in HSV-1 a. Inhibition of translation and the protein product of this mRNA is involved in the latency of the viruses in host cells b. Replication inhibition. Phen-DC3 unlike TMPyP4 hampered the replication process Possibly a platform for binding of the host transcription factors and regulation of gene expression Regulation of replication. SV40 bears a large a multifunctional protein called T-antigen, which is capable of unwinding G4s Uncharacterized but probably important for translation or viral replication a. Prevention of RNA synthesis by the RNA-dependent RNA polymerase b. Nucleolin protein of the host acts as an antiviral protein by binding to HCV <i>core</i> gene G-quadruplex and suppressing its replication Not defined but inhibition of transcription/replication or translation is the most probable function These two highly conserved G-quadruplex-forming sequences thwart the replication and may also be of significance for transcription Implicated in the transcription and probably also replication. TMPyP4 stabilizes this structure and inhibits the replication of the virus Although G4 formation is not confirmed experimentally, the presence of these G-rich regions might be related to genetic heterogeneity Evading the host RNA decay ribonucleases by stalling the activity of XRN1	Saranathan et al. (2019) Artusi et al. (2016) Biswas et al. (2018) Frasson et al. (2019) Frasson et al. (2019) Frasson et al. (2019) Murat et al. (2014) Ravichandran et al. (2018) Gilbert-Girard et al. (2017) Dabral et al. (2020) Madireddy et al. (2016) Tlučková et al. (2013); Puig Lombardi et al. (2019) Patel et al. (2000); Tuesuwan et al. (2008); Mishra et al. (2019) Fleming et al. (2016) Jaubert et al. (2018) Wang et al. (2016a); Bian et al. (2019) Bartas et al. (2020); Ji et al. (2020); Panera et al. (2020) Majee et al. (2020) Wang et al. (2016b); Krafčíková et al. (2017b) Glazko and Kosovsky (2013) (Charley et al. (2018)
HSV-2 (HHV-2) <i>Herpesviridae</i> ds-DNA			
VZV ^{ns} (HHV-3) <i>Herpesviridae</i> ds-DNA			
EBV (HHV-4) <i>Herpesviridae</i> ds-DNA			
HCMV (HHV-5) <i>Herpesviridae</i> ds-DNA			
HHV-6 <i>Herpesviridae</i> ds-DNA			
KSHV (HHV-8) <i>Herpesviridae</i> ds-DNA			
HPV <i>Papillomaviridae</i> ds-DNA			
SV40 <i>Polyomaviridae</i> ds-DNA			
ZIKV ^{ns} <i>Flaviviridae</i> (+) ss-RNA			
HCV <i>Flaviviridae</i> (+) ss-RNA			
SARS-CoV <i>Coronaviridae</i> (+) ss-RNA			
NIV ^{ns} <i>Paramyxoviridae</i> (-) ss-RNA			
EBOV ^{ns} <i>Filoviridae</i> (-) ss-RNA			
FLUAV [*] <i>Orthomyxoviridae</i> (-) ss-RNA			
RVFV <i>Phenuiviridae</i> (-) ss-RNA + ambisense RNA			

(continued)

TABLE 2—Continued

Virus (Name, Family, Genome)	Related Gene/Protein	Role	Reference
HIV-1 <i>Retroviridae</i> ss-RNA + RT	a. LTR promoter b. 3' UTR c. Nef gene	a. Silencer of transcription and its stabilization is concurrent with decreased virus production b. Conserved sequences for m ⁶ A installation, presumably important for proper viral replication c. Controls the genes expression of Nef protein and stabilizing the related gene represses the Nef gene expression and restricts viral infectivity d. Inhibition of reverse transcription, which was suppressed further by BRACO-19 e. Along with other G4 structures in the genome, cooperates in the dimerization and adherence of two HIV-1 RNA structures f. Causing virus latency	(Perrone et al. (2015); Butovskaya et al. (2018) Fleming et al. (2019) Perrone et al. (2013b) Perrone et al. (2014)
HBV <i>Orthohepadnaviridae</i> ds-DNA + RT	d. U3 region of the viral RNA e. <i>Gag</i> gene and cPPT f. Sp1 binding region in HIV-1 promoter ^b a. <i>PreS2/S</i> promoter (HBV genotype B or HBV/B) b. ε stem-loop structure of 3' end of RNA (HBV)	a. This G4 adjusts the transcription of HBV surface antigen levels in a positive manner, unlike routine G4 structures b. A location for the installation of m ⁶ A is interleaved between two tetrad P/Qs, which is involved in increased translation and decreased reverse transcription	Piekna-Przybylska et al. (2013) Piekna-Przybylska et al. (2020) Biswas et al. (2017) Inam et al. (2018)

^aEBOV, Ebola virus; FLUAV, influenza A virus; VZV, varicella zoster virus; ZIKV, Zika virus.

^bSp1-binding region is part of LTR in HIV-1.

*G4 formation is not verified experimentally for FLUAV.

including RIG-I (for retinoic acid-inducible gene I), MDA-5 (for melanoma differentiation-associated protein 5), Mx1 (for Myxovirus resistance protein 1), PKR (protein kinase R), STAT1, and JAK1 (Janus kinase I), have been characterized as target proteins for ISGylation. Some viruses stimulate the production of viral-specific proteins that can deconjugate ISG15 from its target proteins or inhibit the ISGylation of the mentioned proteins, therefore abolishing the antiviral response of the immune system (Jeon et al., 2010).

- OAS proteins are a group of proteins with antiviral activity stimulated by the IFN- α and IFN- β (Lee et al., 2019). They are implicated in the degradation of viral RNA with the aid of ribonuclease L (RNase L) (Choi et al., 2015).

Overall, these telomeric G4 structures inhibit the transcription of antiviral proteins. Therefore, targeting host-cell telomeric G4s by ligands may aid in the body's response to the viral pathogens. This mechanism may explain why some ligands display spurious results in vitro but contradictory or disappointing results in vivo. The selectivity and specificity toward viral G4s should be high enough to override the stabilization caused in the immune-related genes. Apart from this point, this stabilization can also be helpful in viral conditions in which a massive cytokine storm is noted, as found in severe cases of COVID-19 (Ye et al., 2020). In this case, the immune activation by the alluded genes and pathways of the interferons is only aggravating the condition, and a broad-spectrum stabilizer might be more efficient in subjugating the viral disease.

There are some other aspects of the immune system that are interestingly regulated by G4 structures. G-quartet nuclease 1 is a human nuclease capable of cutting G4 DNA. This protein is linked to heavy-chain class-switch recombination in immunoglobulin genes (Sun et al., 2001). Typically, this switch entails a shift between expressing IgM and IgD to the expression of IgG, IgA, or IgE, inducing an upregulation in the humoral response of the immune system that is essential in viral elimination (Stavnezer and Schrader, 2014). Whether the activity of this nuclease can be exploited in antiviral drug development remains an open question.

2. Oncogenic Viruses. A diverse set of viruses known as oncoviruses is known to cause cancer in humans. Some viruses, such as EBV, KSHV, HPV, HBV, and HCV, which are all known to have G4 sequences with functional importance, are classified as type 1 carcinogens by the International Agency for Research on Cancer (Kumar et al., 2020). G4 ligands that stabilize cellular oncogenes suppress the expression of proliferative genes, and they can also prevent the expression of viral genes. Thus, application of such ligands should be explored for preclinical and clinical applications.

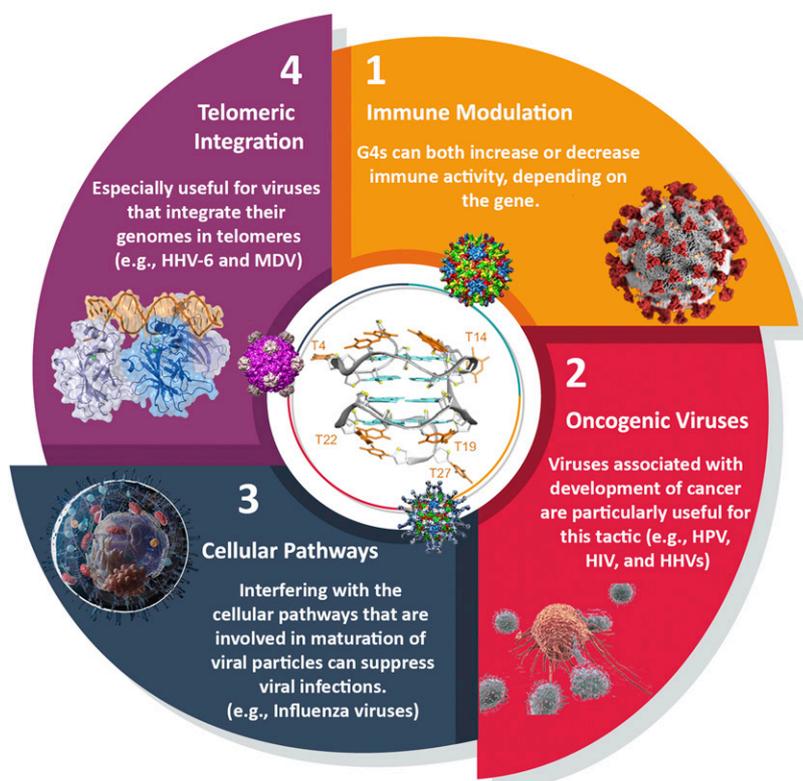


Fig. 6. Host-based G-quadruplexes as targets for antiviral approaches. 1) Immune modulation may be useful for viruses that significantly disturb the immune balance; 2) targeting G-quadruplexes can restrain the growth of oncogenic viruses, by inhibiting both the virus and infected cells that exhibit cancer-like behavior; 3) modulating cellular pathways and enzymes that are essential for viruses like serine proteases by G4 stabilization can hugely impact the virus survival; 4) targeting telomeres can restrict the access of viruses for their DNA integration.

Oncogene promoters are among the best-studied forms of G-quadruplexes. Oncogenes are known to have regulatory effects on the immune-related genes. Here we summarize a brief perspective about the immune significance of these oncogenes to assess whether these oncogenes can have positive or negative effects on the immune system and viral resistance. Depending on the type of oncogene, this effect can be upregulatory or downregulatory. For instance, the *Myc* oncogene is well recognized as one of the proteins involved in the mitigation of immune elicitation by the highly proliferative cells (Casey et al., 2018). VEGF (vascular endothelial growth factor) can cause a defect in the functional maturation of dendritic cells from their progenitors. VEGF also prevents the differentiation and function of different immune cells during hematopoiesis (Li et al., 2016). There is an increasing amount of evidence that advocates the idea that KRAS (Karsten-RAS) is closely tied with the immune evasion of tumor cells and the production of Th1-2-suppressive cytokines (van Maldegem and Downward, 2020). HIF-1 α (hypoxia-inducible factor 1 alpha) activity is stimulated in response to viral pathogens, but increasing evidence indicates that the result of this activation can favor the virus and not the host. Some viruses have developed mechanisms to stabilize HIF-1 α to produce an anti-apoptotic effect that sustains the survival of the infected

cells (Palazon et al., 2014). On the other hand, *c-kit* (tyrosine-protein kinase KIT, CD117) is known to have very complicated signaling pathways in the immune system, and *c-kit* mutation with a gain of function is reportedly related to mast cell proliferation and allergic reactions. *c-kit* has also been implicated in skewing the differentiation of T cells to Th2 and Th17 cells and away from Th1 (Ray et al., 2010). Overexpression of BCL-2 (B-cell lymphoma 2) in the mice is linked with higher immune response and prolonged survival of B cells (Renault and Chipuk, 2013). *RET* (rearranged during transfection), another proto-oncogene, is engaged in the proinflammatory pathways and also homeostasis of the immune system (Rusmini et al., 2013). To conclude, the crosstalk between oncogenes and the immune system is complex and needs to be better understood if one wants to exploit it against viral infections.

3. Cellular Pathways. Recently, researchers discovered that a G-quadruplex is present in the promoter region of the human *TMPRSS2* (transmembrane protease, serine 2) gene, which encodes a type II transmembrane serine protease that can sever hemagglutinin of many subtypes of influenza viruses and spike glycoprotein of coronaviruses (Shen et al., 2020). Benzoselenoxanthene analogs are capable of stabilizing this G-quadruplex, downregulating this gene and manifesting demonstrable antiviral activity commensurate with

the inhibitory activity of oseltamivir, a neuraminidase inhibitor of influenza viruses. It remains to be determined whether this antiviral effect was mediated by Tmprss2 inhibition or via the stabilization of other viral or host-cell G-quadruplexes. In fact, benzothioxanthene derivatives, which are structurally related to benzoselenoxanthenes, also stabilize telomeric G4s (Mergny et al., 1998). In any case, these studies sprout the idea that stabilization of human G4 structures may be beneficial for the treatment of viral infections.

4. Telomeric Integration. HHV-6A and HHV-6B are two distinct types of ds-DNA viruses that belong to the subfamily *Betaherpesvirinae*. HHV-6B is a ubiquitous virus that infects nearly 100% of the human population. HHV-6A infection causes sixth disease (exanthema subitum or roseola infantum) in children, and the presence of HHV-6 in normal brains suggests a latent phase in the central nervous system, which, in some rare cases, can be later reactivated and lead to encephalitis (Limeres Posse et al., 2017; Fida et al., 2019). In the latent phase, human herpesviruses typically maintain their genomes as extrachromosomal nuclear episomes. How HHV-6A/B achieves latency is still enigmatic. The HHV-6A/B genomes consist of a unique sequence that is flanked by G-rich direct repeat regions that harbor the packaging sequences (*pac-1* and *pac-2*) and two arrays of either perfect or imperfect telomeric repeats (TMRs) at the genome termini. The fluidity of telomeres is important for efficient chromosomal integration of HHV-6A, and that interference with telomerase activity negatively affects the generation of cellular clones containing integrated HHV-6A. Gilbert-Girard and colleagues (2017) have examined the effects of a G-quadruplex binding and stabilizing agent, BRACO-19, on HHV-6A chromosomal integration. BRACO-19 reduced the number of clones harboring integrated HHV-6A and negatively affected HHV-6A integration in telomerase-expressing cells (Gilbert-Girard et al., 2017). Although the effects of BRACO-19 on the viral TMRs and *pac-1* G4s remain partially unexplained, their observation (stabilization of host G4s in telomeres) suggests that telomeric G4 ligands can serve as a prevention therapy for those at risk of HHV-6A virus or even a treatment for related viruses that integrate their genome at the telomeric segments of the host like Marek disease virus [MDV, an oncogenic alphaherpesvirus (Previdelli et al., 2019)].

F. Summary of G-Quadruplex Roles in Viruses and Host

At the end of this discussion of host and viral G-quadruplexes, we tried to summarize several decades of findings in a simplified figure (Fig. 7). Currently, there is no evidence that G4 structures in the viral genome are involved in mRNA splicing, DNA repair induction or abrogation, post-translational modifications, and shifting the host cells into a programmed or

unordered form of cell death. For instance, pyroptosis is a form of programmed inflammatory cell death that is triggered by pathogens like viruses. It has been documented that some G-quadruplex-bearing viruses like the influenza virus are able to stimulate pyroptosis or apoptosis, but there is no report for the role of G-quadruplex structures in these processes (Lee et al., 2018). Although there are some pieces of evidence that the telomeric host G-quadruplex can serve as an integration site for viral DNA, no valid evidence has yet unveiled the role of viral G4 in their genome integration.

IV. Discussion: Challenges

A. Specific Studies on G-Quadruplexes in Coronaviruses: Response to Coronavirus Disease 2019 Pandemic. The recent COVID-19 outbreak stimulated an unprecedented research effort to find new targets against the SARS-CoV-2 virus. Knowing that G-quadruplex-prone motifs were predicted in the genome of the *Coronaviridae* family, including the MERS and SARS, it was interesting to determine whether putative G-quadruplex sequences are also present in the SARS-CoV-2 genome. Screening with QGRS Mapper, a web-based tool for identifying quadruplex-prone sequences, suggested that 25 such sequences are present but only with G2 motifs, meaning that they correspond to relatively unstable G4s (Ji et al., 2020; Panera et al., 2020). Two of these short RNA sequences were experimentally confirmed to form G4 structures in vitro but with low thermal stability (melting temperature, or $T_m < 37^\circ\text{C}$; meaning these structures are predominantly unfolded at physiologic temperature) (Ji et al., 2020). The same authors suggested that PQSs may be present in the open reading frames of several SARS-CoV-2 genes, such as spike glycoprotein, membrane, and nucleocapsid genes.

Bartas et al. (2020) analyzed the occurrence of putative G4 motifs within the 109 genomes of all known *Nidovirales*, such as coronaviruses, including SARS-CoV, MERS-CoV, and SARS-CoV-2. Rather than QGRS Mapper, they used G4Hunter to predict G4-prone motifs, keeping a minimal threshold of 1.2 (not a single motif with a G4-Hunter score of 1.6 or above was found in all *Nidovirales*). This threshold value, which has previously been shown to maximize accuracy (Bedrat et al., 2016), will miss relatively unstable G4 with lower scores – in other words, it is more stringent than the QGRS studies discussed above, explaining why fewer hits would be found. The lowest G4 density was found in *Nidovirales* infecting humans (0.06), including SARS-CoV-2, for which a single motif is present on the negative strand and none are present on the positive strand, with a G4-Hunter score above 1.2. The G4 density in this virus is lower than expected by chance, arguing for a counter-selection evolution against G-prone motifs

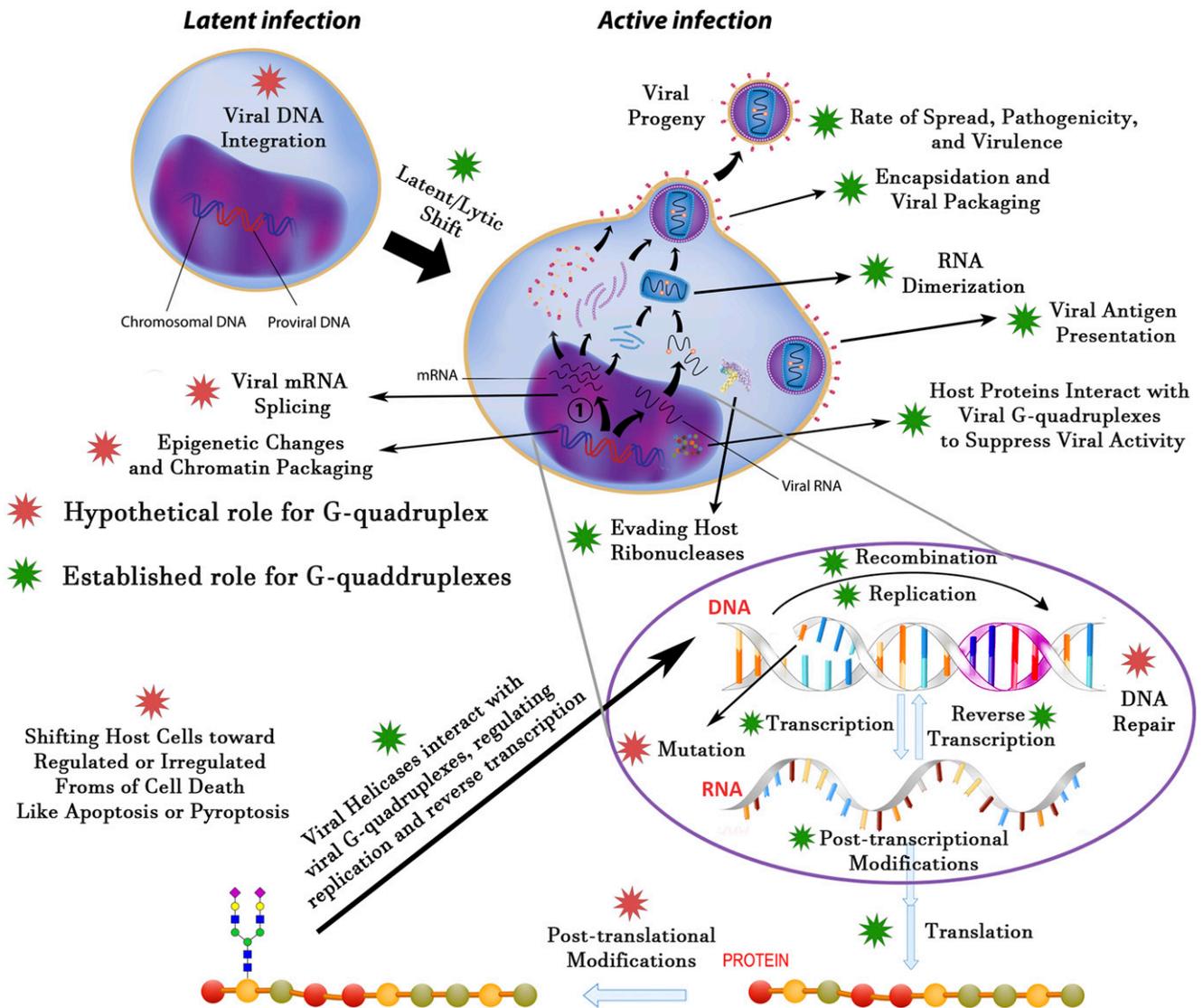


Fig. 7. A depiction of the established and hypothetical roles of viral G-quadruplexes for a virus that displays both latent and active phases (taking HIV-1 as a prototypical virus; some but not all of these principles could be applied to other viruses). Transcription can be regarded as the most-studied process in viral pathogenesis. The hypothetical roles are well established roles for G-quadruplexes that are verified in other organisms but not yet approved in viral processes and may be unraveled with further research works.

in SARS-CoV-2. The 25 motifs reported by Ji et al. (2020) or Panera et al. (2020) all have lower G4-Hunter scores.

Interestingly, some coronaviruses code for G4-binding proteins. For example, the Nsp3 of SARS-CoV possesses a SUD, indispensable for the replication/transcription procedure of viral pathogenesis (Kusov et al., 2015). SUD is a 338-amino-acid domain that is absent in less-deadly coronaviruses and is therefore regarded as one of the domains that enhances the pathogenicity of SARS with respect to other members of the *Coronaviridae* family. SUD is composed of three subdomains dubbed as SUD-N [macrodomain 2], SUD-M (macrodomain 3), and SUD-C (domain preceding Ubl2 (ubiquitin-like protein 2) and PL2^{pro} (papain-like protease 2)). The SARS-CoV SUD domain binds to G-quadruplexes (Tan et al., 2009; Kusov et al., 2015;

Lei et al., 2018). SUD-NM interacts with RNA G-quadruplexes since Nsp3 has not been found in the nucleus of the cells. Mutations of some lysine residues in SUD-M abolish G4 binding and inhibit viral replication. SUD-C has also been announced as a regulator of the specificity of the RNA-binding activity of SUD-M (Johnson et al., 2010). By altering the pattern of gene expression in infected cells through the interaction of SUD to G4 structures present in mRNAs, SARS-CoV may wield the cellular processes in its desired states. As this SUD domain is conserved between SARS-CoV and SARS-CoV-2 (M. Lavigne et al., manuscript in preparation), this mechanism may contribute to the high pathogenicity of these two viruses.

The fact that the SARS virus codes for a G4-binding protein while being itself relatively G4-poor may seem surprising. However, besides viral G4, one should not

forget that the host cell also contains G4-prone nucleic acids that may well be relevant for viral infections. In addition, other host-cell proteins important for infection, such as TMPRSS2 may possibly be regulated by G4 ligands. Both aspects will be discussed in the next section.

B. Assets and Drawbacks in the Development of Antiviral Therapies Targeting G-Quadruplexes. Nearly all currently used drugs are known to target protein structures, with a few exceptions, such as antibiotics, aimed at ribosomal RNA. Our understanding of how proteins behave far exceeds that for nucleic acids. There are countless ligands with versatile structures that are capable of modulating the behavior of different sets of proteins. Protein cavities are adroitly designed to interact with a panoply of other proteins, cellular metabolites, and xenobiotics. Hence, they are more druggable than their nucleic acid cousins, which are often considered to be less intricate than proteins and large peptides. The unique folding of proteins—with the notable exception of intrinsically disordered ones—makes them appropriate targets for medicinal chemists, especially when binding or catalytic sites are known. This can be exploited by a variety of ligands, including small molecules, peptides, aptamers, or nucleotides. In contrast, this subtlety is not the same for nucleic acids, as they offer relatively limited diversity in terms of “residue” properties (all nucleotides have identical charges and relatively similar hydrophilicities). In addition, RNA tends to be inherently flexible, and apart from noncanonical structures, it often lacks a specific binding pocket that can be exploited for specific recognition.

These drawbacks do not mean that nucleic acid targeting is impossible: for example, a number of antibiotics work by binding to the prokaryotic ribosome. In addition, G-quadruplexes constitute original targets among nucleic acids with unique advantages: 1) They are often well defined structures; 2) their electrostatics are fundamentally different from other DNA/RNA fold because of the presence of four sugar-phosphate backbones and a central spine of positively charged ions; 3) terminal quartets offer a stacking platform for planar aromatic ligands; and 4) loops, grooves, bulges, and flanking sequences offer additional epitopes for specific recognition. Overall, intramolecular G-quadruplexes should be considered as globular shapes rather than linear DNA or RNA polymers. Their folding landscape and timescale are also reminiscent of proteins.

These properties were exploited by a number of teams who identified hundreds of compounds binding to G-quadruplexes. Unfortunately, many of these molecules are not drug-like and violate some of Lipinski's rule of five. Even though Lipinski's rule of five was initially applied to oral drugs and has been challenged these years, at least for some parameters, this rule provides useful guidelines for optimizing pharmacokinetic

tolerance and efficacy (Doak et al., 2014). For example, most G4 ligands tend to have a relatively large size; Quarfloxin (CX-3543), the first G4-targeting drug to reach the clinic [it went in phase II clinical trials for the treatment of cancer (Buket et al., 2014)] violates this rule in terms of molecular weight (Buket et al., 2014). In addition, while G4 offer unique advantages among nucleic acid targets, they also pose specific problems: 1) Some, but not all, G4s tend to be polymorphic; this structural variability is illustrated with the human telomeric motif (or, to a lesser extent, to viral sequences, such as the one present in the HIV LTR region) for which multiple folds are known. Identifying which ones are relevant in a physiologic environment is not straightforward; 2) although distinguishing G4 from non-G4 structures is relatively simple, making the distinction between various G4s is harder, especially when planar aromatic compounds are considered and most will bind to all G4 offering an accessible terminal tetrad for stacking. Given the high number of potential G4 sequences in the human genome, it may be well advised to limit off-target effects by selectively target a subset of G4 motifs; 3) the physiologic and biochemical pathways altered by G4 ligands are still largely elusive, in part because many biologic studies reported their effects on a few genes only and not at the transcriptome or genome-wide level; and 4) related to that, and in contrast with a widely accepted belief, G-quadruplex structures are not always acting as transcriptional repressors, and predicting the impact of a G4 ligand on the expression profile of a given gene is not straightforward.

These opportunities and drawbacks apply to all G4-based strategies, aimed at fighting cancer or transmittable diseases. There are, however, additional hurdles in antiviral research, as the biologic roles of viral G4s only start to be uncovered. It took decades to unambiguously prove the formation of G-quadruplexes in human cells (Lipps and Rhodes, 2009). Not all G-rich motifs found in viruses may adopt a G4 fold, and even then, not all G4 may be appropriate targets (i.e., they may not be essential for viral survival or pathogenesis).

Importantly, whenever viral G4 targets are considered, the genome and transcriptome of the host cell offer thousands of potential off-sites that can divert G4 ligands from their intended viral target. These cellular sites may be present in large molar excess, even when the virus is actively replicating. Searching for highly efficient binders with low interference with the host-cell functions may therefore be the way to go to design potent antiviral agents capable of fighting the current and future viral outbreaks. Some G-quadruplex ligands may have a higher affinity for viral G4 structures than for human structures. A c-exNDI induced significant virus inhibition with no cytotoxicity by inhibiting viral DNA replication, with consequent impairment of viral genes transcription. c-exNDI preferentially targeted

HSV-1 G4s over cellular telomeric G4s, one of the most well established G4s within host cells, whereas other less abundant cellular G4s were also recognized in the host (Callegaro et al., 2017). Naphthalene diimide derivatives also displayed significant antiviral activity at nanomolar concentrations by targeting HIV-1 LTR promoter G4s (Perrone et al., 2015). The regulation of transcription banks on the binding of host-cell transcription factors and associated regulatory proteins to the LTRs. Treatment of HSV-1-infected cells with BRACO-19 caused a pronounced halt of virus production and viral DNA synthesis (Artusi et al., 2015; Frasson et al., 2019). On the other hand, one can propose that the broad stabilization of both host and viral G4s may have synergistic effects and constitute an alternative strategy for antiviral drug discovery.

To make progress, the physiologic role of host and pathogen G-quadruplexes ought to be uncovered, and to this end, we should first establish which sequences of viral genomes fold into these ordered structures. It may be possible that, as it is increasingly accepted for human G4s, only a small fraction of potential G4-forming motifs adopts a G4 fold under physiologic conditions. The same can be true for viral candidate sequences. The current number of putative G4 viral structures massively outnumber the number of well characterized ones. Many of these sequences are annotated as putative G4-forming sequences (PQSs), pending exact *in vitro* and *in vivo* characterization. Identifying and validating a new G4 structure in the genome of a virus is a laborious task. These PQSs have been proposed in viral genomes that have not been fully investigated, such as the Zika virus. Moreover, HSV-1 genome indicates multiple clusters of repeated sequences creating highly stable G-quadruplexes involved in the replication of viral DNA (Puig Lombardi et al., 2019). Thus,

our identification techniques should be further improved for a better understanding of G-quadruplexes.

We summarize the major drawbacks in utilizing G4 targetability in the antiviral context (Fig. 8). The first step is to identify appropriate targets. A complete *in silico* mapping of the known viral genomes has already been performed, and the currently available prediction tools allow testing every new virus genome in a few minutes. The recent example of the SARS-CoV-2 situation demonstrates that even this analysis is not straightforward, as teams may disagree on what a candidate sequence should look like and which algorithm should be used. As a consequence, the number of candidate sequences oscillates between 0 and 25. The next step will be to validate these candidate motifs, first by showing their formation on short DNA/RNA oligonucleotides *in vitro* (which may be harder than it seems) and then on longer fragments and ideally on the full transcript or the whole viral genome. The use of electronic circular dichroism allows an easy and straightforward technique to infer *in vitro* not only the presence of G4s but also their specific topology (parallel, antiparallel, hybrid, etc.) thanks to the emergence of specific spectral signatures (Cheng et al., 2018; del Villar-Guerra et al., 2018).

Once validated, structural determination, although not indispensable, would definitely facilitate drug design but seems hard to achieve, as we do not have a single X-ray crystallographic structure of a viral G-quadruplex structure. Although one can argue that G4 structures can manifest themselves into a variety of topologies, their “targetability” or “ligandability” may not vary as much as for protein structures.

A number of screening approaches may then be proposed to identify G4 ligands binding to the target(s). Initial screening, either *in silico* with virtual libraries or

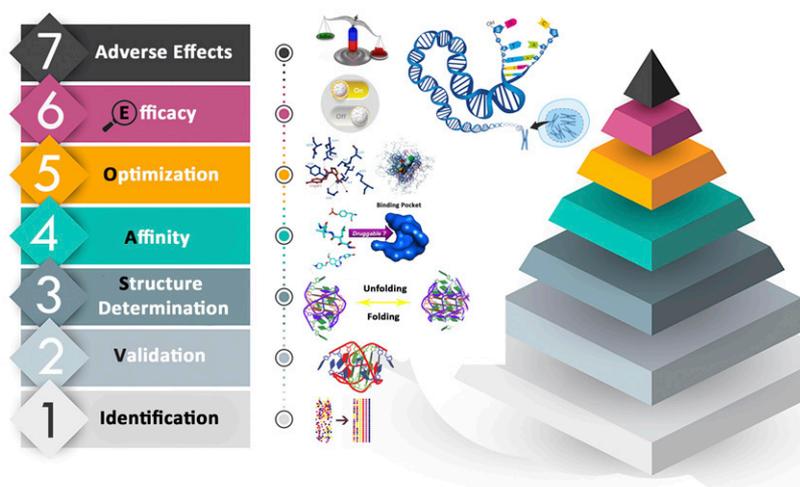


Fig. 8. The general scheme of the obstacles of G-quadruplexes, as a mean to direct drug-like molecules toward them. For the development of a new ligand molecule in drug design and discovery, first we need to identify, validate, and determine the structure of a relatively stable G-quadruplex in the viral (or host) cells. This G-quadruplex should have pertinent role to viral infectivity. With the aid of medicinal chemistry, the affinity of the ligand should be optimized to reach the desired *in vitro* and *in vivo* efficacy. Certain precautions are needed to avoid genomic instability and toxicological damage to the normal cells.

directly with chemical libraries, is used to identify hit molecules. Using structure-activity relationship (SAR) between the previously characterized compounds allows establishing a correlation between activity and structure. This process can be also quantitatively accomplished using quantitative SAR. The next step will be to convert hits into leads and then drug-like compounds. This part should be familiar to any medicinal chemist, as it is in large part identical to any other drug development program. It may prove harder than for other targets though, as most G4 ligands display mediocre drug-likeness potential when compared. A series of drug-likeness evaluations should be implemented to modify and improve structure toward molecules with a balanced lipophilicity/hydrophilicity, appropriate size and molecular weight, and fair permeability to enter the cells. The penultimate step should be to minimize off-target effects and toxicity toward the host, which is crucial given the presence of multiple potential targets in normal cells. Long-term genomic instability studies have to be considered in this stage, since these compounds may interact with DNA (even if preferentially aimed at RNA). G-quadruplex ligand PDS, for example, was thought to cause double-strand breaks in some studies (Moruno-Manchon et al., 2017), whereas in some others it was found to even mildly mitigate the formation of such DNA damage (Kumari et al., 2019). The evaluation of the *in vivo* effect of the compound, along with its bioavailability, biodistribution, and pharmacokinetic profile is necessary, and the antiviral effect of the compound can first be determined in cellular models and then *in vivo*.

C. Hit-to-Lead Optimization for G-Quadruplex Ligands. In the rational drug design, hit-to-lead optimization can be accomplished by comparing the effects of compounds belonging to the same family but differing by chemical groups or atoms at specific locations. This allows to determine SARs, and this is often coupled with structural studies, wherein the impact of such modification on the binding to its target is determined thanks to high-resolution structural methods coupled to molecular dynamics simulations. Indeed, recently Hognon et al. (2020) assessed the exact binding interactions that favors the dimerization of SUD in cooperation with G-quadruplexes via molecular dynamics simulations and free-energy profiles. In their inspiring review about the application of molecular modeling and simulations in antiviral drug discovery, Francés-Monerris and colleagues (2020) state that the “open inactive” structure of SUD was shown to be disfavored by RNA G-quadruplexes. Their work suggests that specific ligands that perturb this mutual interaction could be employed in therapeutic approaches. Combined quantum mechanics/molecular mechanics studies are gaining attention despite their computationally intensive nature. The works by Batista group are a suitable example of application of these hybrid

methods in high-quality prediction of PQS structures (Ho et al., 2014), which should be extended to study viral G-quadruplexes. The coupling of computational molecular simulation with experimental spectroscopy (Gattuso et al., 2016) provides a useful tool to unravel the detailed folding of G-quadruplexes and their interactions with other proteins and ligands, which are crucial in understanding viral infections, and can also reveal the drug mechanism of actions.

Alternative attractive strategies are based on fragment-based or scaffold-based hit optimization. Amiodoxime, cationic bis(acylhydrazones), benzoxalenoxanthene, naphthalene diimides, dibenzophenanthroline, bisquinolines/quinolinium, acridine, and porphyrin are the scaffolds that have been tested for stabilization of viral G-quadruplexes. The relative paucity of medicinal chemistry studies on G4 ligands against viruses makes it difficult to provide a general core or scaffold for G-quadruplex ligands [for a review on G-quadruplex ligand structure, please refer to Sun et al. (2019)]. Hit optimization also uses *in silico* toxicological profile software and databases to ensure the relative safety of the compounds and their lack of mutagenic (and possibly carcinogenic) potential.

Unfortunately, in the G-quadruplex field, 1) the target is not always known, 2) some ligands may bind to multiple targets, and 3) either the three-dimensional structure of the target is poorly known, especially in a physiologic environment (topology may be affected by (macro)molecular crowding (Amjadi Oskouie and Abiri, 2021), for example), or no data are available on ligand-binding mode: although there are over 250 G4 structures in the Protein Database (PDB), only a few high-resolution structures of drug-G4 complexes are known.

There is, therefore, considerable work to be done to progress from hit to lead, and from lead to drug, and a G4 ligand has yet to reach clinical trial against a viral infection. The knowledge accumulated on G4 agents active against cancer should be precious, as some of the issues anticipated for a G4-against-virus strategy are shared for anticancer agents.

D. What Remains Untouched? There are some untouched areas that may boost the significance of viral G4s in future with further experiments. For instance:

1. It remains to be investigated whether virulence and pathogenicity are related to the density of G4s in viral genome or not. Owing to the ability of G4s to shift viral cycle to lytic/latent phases, there could be a relationship with acuteness or chronicity of viral infections (Bohálová et al., 2021). According to results found for EBNA1 of EBV (Lista et al., 2017; Reznichenko et al., 2019), replication compartments of HSV-1 (Artusi et al., 2016), LANA of KSHV (Madireddy et al., 2016), and Sp1 binding-site region in HIV-1 LTR (Piekna-Przybylska et al.,

2020), we can assume that G4s are more likely to switch the active phase to the latent phase, but the correlation of this latency with the duration of diseases (acuteness or chronicity) requires further analysis. Viral latency is involved in some poorly diagnosed or quiescent forms of cancers like cervical cancer caused by HPV. G4s are one of the pivotal players in viral latency, and their targeting can serve as both therapeutic and diagnostic goals.

2. Autophagy is a cellular mechanism that is mainly used to emancipate the cells from the junk products as well as external pathogens like viruses. There is some evidence that indicates the role of G4s in modulating autophagy (Beauvarlet et al., 2019a,b, 2020; Lejault et al., 2020). This modulation could impede viral elimination by possible drug molecules but has not yet been studied.
3. The role of G4s in chromatin packaging for viruses that integrate their genome in the host DNA requires further investigation. Also, the interaction between the G4s and epigenome is scantily investigated.
4. The safety of G4 ligands in animal models with viral infections is not assessed and demands scrutiny.
5. Should we selectively target the viral G4s or not? This is probably the hottest unexplored conundrum regarding the use of drugs that target G4s. The evidence seems to suggest that this kind of selectivity may not be necessary.
6. Some viruses seem to be devoid of stable G4 structures. Among the most notable examples of these viruses are measles virus, mumps virus, poliovirus, and hantavirus. Whether this lack of G4 occurrence is due to insufficient studies or is due to the real absence of these elements requires further explorations.
7. Higher-order multimerization of G4 structures is being investigated for genomic sequences. Some mutations have also been characterized to facilitate the formation of such higher-order structures (Kolesnikova et al., 2017). It would be interesting to study whether such dimeric or tetrameric assembly (multimerization) really exists for viral G4 motifs and to study whether they have some functional role in viral pathogenesis, which may be further used for therapeutic strategies.

V. Conclusions

Most if not all viruses contain G4-prone sequences and/or express proteins that interact with G-quadruplexes. These structures may have a conspicuous role in disease manifestation and progression. A large part of what we

know about viral G4 comes from research on HIV-1, for which we find multiple examples of the involvement of G4 in the virus life cycle.

In this review, we give an overview of how G-quadruplexes might be relevant in antiviral drug design and discovery. The conserved nature of viral G4s in various strains along with their proposed significance in replication, transcription, translation, post-transcriptional modifications, and viral latency indicates that these elements are felicitous but unexplored armamentaria in the treatment of viral infections. With a few exceptions, G4 structures in viral genes are thought to be negative (repressive) elements for replication, transcription, or translation, and their stabilization is therefore linked with slumped viral propagation. It then remains to be established why some viruses are enriched in G4 motif (e.g., herpesviruses, which have an up to 7-fold higher PQS density than the host-cell genome), whereas others are not (e.g., SARS-CoV-2). By acting as suppressors of the expression of host growth-related genes or by trapping essential elements like proteins in viral life cycle, host G4 structures may also be considered as potential—but possibly less straightforward—targets for G4 stabilizers. The presence of conserved G4 motifs in viruses implies that these G-rich sequences play important roles and participate in pathogenesis. The EBNA1 results illustrate a situation wherein G4 may provide a regulatory function, as overexpression of a viral protein may be harmful to the virus, especially if it is immunogenic. G4 elements may be used to deflect host exonucleases, lower the rate of antigen presentation, allow an appropriate recognition of the host-protein machinery for viral packaging and encapsidation, and shift between the latent and lytic phases.

Still, their conserved and ubiquitous nature implies their importance in the pathogenesis should be seeded in other biologic roles rather than simple regulation of those processes. For instance, we observed that these elements are necessary for the deflection of host exonucleases, lowering the rate of antigen presentation, appropriate recognition of host protein machinery for viral packaging and encapsidation, and timely shift of the latent to lytic phase in viral cycles. Thus, these indispensable items in viral genomes can be regarded as targetable in drug design and discovery approaches.

Whether viral G4s should selectively be targeted over host-cell G4s remains a controversial question, as their selectivity for a given fold is often limited, and both are targeted. Interestingly, active replication or transcription of a virus may create a hundred or thousand copies of a DNA/RNA viral target (as shown by immunostaining by a G4-specific antibody), which may compensate for a limited intra-G4 specificity of a ligand. Whether DNA or RNA G4s should be targeted also awaits future investigation. RNA G-quadruplexes are often more stable than their DNA counterparts. In addition, when embedded in a double-stranded DNA region, one needs

to pay an energetic cost to open the double helix and form a DNA G-quadruplex. This cost may be lower in the case of RNA, at least in regions that are not heavily structured. This is not always the case, and RNA G4-prone motifs may be in equilibrium with a local hairpin structure, as found for HCV (Jaubert et al., 2018); neither viral RNA nor the host-cell target mRNA should be considered as single-stranded! Bottlenecks in the process of exploiting G-quadruplexes as antiviral targets are discussed in this review.

To the best of our knowledge, targeting host G-quadruplex structures has not yet been directly used in antiviral research. Some notable and interesting cases for studying the antiviral activity of the host G4 stabilizers are those viruses that integrate their genome to the host telomeric or G-rich regions like MDV and HHV-6.

In contrast to G4s in cancerology, we have yet to witness clinical trials on agents fighting viral infection via a G4-based mechanism. The fact that some derivatives can halt the replication of different, unrelated viruses (e.g., HSV-1 and HIV-1) indicates that G4 ligands can display broad-spectrum antiviral activity. The threat posed by the new coronavirus infections offers a new opportunity to test “out-of-the-box” approaches, such as G4 ligands. Time will tell whether these approaches will soon prove to be successful.

Acknowledgments

We wish to thank our past and present colleagues for stimulating discussions. J.-L.M. thanks V. Brazda (Institute of Biophysics, Brno) for sharing unpublished work.

Authorship Contributions

Participated in research design: Abiri, Lavigne, Mergny, Rahimi.
Conducted experiments: Abiri, Mergny.
Contributed new reagents or analytic tools: Rezaei.
Performed data analysis: Abiri, Lavigne, Mergny.
Wrote or contributed to the writing of the manuscript: Abiri, Nikzad, Zare, Mergny, Rahimi.

References

Amjadi Oskouie A and Abiri A (2021) Refining our methodologies for assessing quadruplex DNA ligands; selectivity or an illusion of selectivity? *Anal Biochem* **613**:113744.

Amrane S, Kerkour A, Bedrat A, Vialet B, Andreola M-LL, and Mergny J-LL (2014) Topology of a DNA G-quadruplex structure formed in the HIV-1 promoter: a potential target for anti-HIV drug development. *J Am Chem Soc* **136**:5249–5252.

Artusi S, Nadai M, Perrone R, Biasolo MA, Palù G, Flamand L, Calistri A, and Richter SN (2015) 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.

Artusi S, Perrone R, Lago S, Raffa P, Di Iorio E, Palù G, and Richter SN (2016) Visualization of DNA G-quadruplexes in herpes simplex virus 1-infected cells. *Nucleic Acids Res* **44**:10343–10353.

Bakalar B, Heddi B, Schmitt E, Mechulam Y, and Phan AT (2019) A minimal sequence for left-handed G-quadruplex formation. *Angew Chem Int Ed Engl* **58**:2331–2335.

Bang I (1910) Untersuchungen über die Guanylsäure. *Biochem Z* **26**:293–311.

Bartas M, Brázda V, Bohálová N, Cantara A, Volná A, Stachurová T, Malachová K, Jagelská EB, Porubiaková O, Červený J, et al. (2020) In-depth bioinformatic analyses of *Nidovirales* including human SARS-CoV-2, SARS-CoV, MERS-CoV viruses suggest important roles of non-canonical nucleic acid structures in their lifecycles. *Front Microbiol* **11**:1583.

Beauvarlet J, Bensadoun P, Darbo E, Labrunie G, Rousseau B, Richard E, Draskovic I, Londono-Vallejo A, Dupuy JW, Nath Das R, et al. (2019a) Modulation of the ATM/autophagy pathway by a G-quadruplex ligand tips the balance between senescence and apoptosis in cancer cells. *Nucleic Acids Res* **47**:2739–2756.

Beauvarlet J, Mergny J-L, and Djavaheri-Mergny M (2019b) Activation of the Ataxia Telangiectasia Mutated/Autophagy pathway by a G-quadruplex ligand links senescence with apoptosis. *Mol Cell Oncol* **6**:1604047.

Beauvarlet J, Nath Das R, Alvarez-Valadez K, Martins I, Muller A, Darbo E, Richard E, Soubeyran P, Kroemer G, Guillon J, et al. (2020) Triarylpyridine compounds and chloroquine act in concert to trigger lysosomal membrane permeabilization and cell death in cancer cells. *Cancers (Basel)* **12**:1621.

Bedrat A, Lacroix L, and Mergny JL (2016) Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res* **44**:1746–1759.

Bettin N, Oss Pegorar C, and Cusanelli E (2019) The emerging roles of TERRA in telomere maintenance and genome stability. *Cells* **8**:246.

Bian WX, Xie Y, Wang XN, Xu GH, Fu BS, Li S, Long G, Zhou X, and Zhang XL (2019) Binding of cellular nucleolin with the viral core RNA G-quadruplex structure suppresses HCV replication. *Nucleic Acids Res* **47**:56–68.

Biffi G, Tannahill D, McCafferty J, and Balasubramanian S (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat Chem* **5**:182–186.

Biswas B, Kandpal M, Jauhari UK, and Vivekanandan P (2016) Genome-wide analysis of G-quadruplexes in herpesvirus genomes. *BMC Genomics* **17**:949.

Biswas B, Kandpal M, and Vivekanandan P (2017) A G-quadruplex motif in an envelope gene promoter regulates transcription and virion secretion in HBV genotype B. *Nucleic Acids Res* **45**:11268–11280.

Biswas B, Kumari P, and Vivekanandan P (2018) Pac1 signals of human herpesviruses contain a highly conserved G-quadruplex motif. *ACS Infect Dis* **4**:744–751.

Bohálová N, Cantara A, Bartas M, Kaura P, Štátný J, Pečinka P, Fojta M, Mergny JL, and Brázda V (2021) Analyses of viral genomes for G-quadruplex forming sequences reveal their correlation with the type of infection. *Biochimie*, in press.

Brázda V, Luo Y, Bartas M, Kaura P, Porubiaková O, Štátný J, Pečinka P, Verga D, Da Cunha V, Takahashi TS, et al. (2020) G-quadruplexes in the archaea domain. *Biomolecules* **10**:1–23.

Buket O, Clement L, and DanZhou Y (2014) DNA G-quadruplex and its potential as anticancer drug target. *Sci China Chem* **57**:1605–1614.

Butovskaya E, Heddi B, Bakalar B, Richter SN, and Phan AT (2018) Major G-quadruplex form of HIV-1 LTR reveals a (3 + 1) folding topology containing a stem-loop. *J Am Chem Soc* **140**:13654–13662.

Butovskaya E, Soldà P, Scalabrini M, Nadai M, and Richter SN (2019) HIV-1 nucleocapsid protein unfolds stable RNA G-quadruplexes in the viral genome and is inhibited by G-quadruplex ligands. *ACS Infect Dis* **5**:2127–2135.

Cahoon LA and Seifert HS (2009) An alternative DNA structure is necessary for pilin antigenic variation in *Neisseria gonorrhoeae*. *Science* **325**:764–767.

Callegaro S, Perrone R, Scalabrini M, Doria F, Palù G, and Richter SN (2017) A core extended naphthalene diimide G-quadruplex ligand potently inhibits herpes simplex virus 1 replication. *Sci Rep* **7**:2341.

Canaan A, Haviv I, Urban AE, Sulz VP, Hartman S, Zhang Z, Palejev D, Deisseroth AB, Lacy J, Snyder M, et al. (2009) EBNA1 regulates cellular gene expression by binding cellular promoters. *Proc Natl Acad Sci USA* **106**:22421–22426.

Carvalho J, Mergny J-L, Salgado GF, Queiroz JA, and Cruz C (2020) G-quadruplex, friend or foe: the role of the G-quartet in anticancer strategies. *Trends Mol Med* **26**:848–861.

Casey SC, Baylot V, and Felsner DW (2018) The MYC oncogene is a global regulator of the immune response. *Blood* **131**:2007–2015.

Charley PA, Wilusz CJ, and Wilusz J (2018) Identification of phlebovirus and arenavirus RNA sequences that stall and repress the exoribonuclease XRN1. *J Biol Chem* **293**:285–295.

Cheng M, Cheng Y, Hao J, Jia G, Zhou J, Mergny J-L, and Li C (2018) Loop permutation affects the topology and stability of G-quadruplexes. *Nucleic Acids Res* **46**:9264–9275.

Choi UY, Kang J-S, Hwang YS, and Kim Y-J (2015) Oligoadenylate synthase-like (OASL) proteins: dual functions and associations with diseases. *Exp Mol Med* **47**:e144.

Dabral P, Babu J, Zareie A, and Verma SC (2020) LANA and hnRNP A1 regulate the translation of LANA mRNA through G-quadruplexes. *J Virol* **94**:e01508–e01619.

De Cian A, Delemos E, Mergny JL, Teulade-Fichou MP, and Monchaud D (2007) Highly efficient G-quadruplex recognition by bisquinolinium compounds. *J Am Chem Soc* **129**:1856–1857.

De Nicola B, Lech CJ, Heddi B, Regmi S, Frasson I, Perrone R, Richter SN, and Phan AT (2016) Structure and possible function of a G-quadruplex in the long terminal repeat of the proviral HIV-1 genome. *Nucleic Acids Res* **44**:6442–6451.

Del Villar-Guerra R, Trent JO, and Chaires JB (2018) G-quadruplex secondary structure obtained from circular dichroism spectroscopy. *Angew Chem Int Ed Engl* **57**:7171–7175.

Deng H, Gong B, Yang Z, Li Z, Zhou H, Zhang Y, Niu X, Liu S, and Wei D (2019) Intensive distribution of G2-quadruplexes in the pseudorabies virus genome and their sensitivity to cations and G-quadruplex ligands. *Molecules* **24**:774.

Dixon IM, Lopez F, Tejera AM, Estève JP, Blasco MA, Pratiel G, and Meunier B (2007) A G-quadruplex ligand with 10000-fold selectivity over duplex DNA. *J Am Chem Soc* **129**:1502–1503.

Doak BC, Over B, Giordanetto F, and Kihlberg J (2014) Oral druggable space beyond the rule of 5: insights from drugs and clinical candidates. *Chem Biol* **21**:1115–1142.

Drygin D, Lin A, Bliesath J, Ho CB, O'Brien SE, Proffitt C, Omori M, Haddach M, Schwaeb MK, Siddiqui-Jain A, et al. (2011) Targeting RNA polymerase I with an oral small molecule CX-5461 inhibits ribosomal RNA synthesis and solid tumor growth. *Cancer Res* **71**:1418–1430.

Fernando H, Sewitz S, Darot J, Tavaré S, Huppert JL, and Balasubramanian S (2009) Genome-wide analysis of a G-quadruplex-specific single-chain antibody that regulates gene expression. *Nucleic Acids Res* **37**:6716–6722.

Fida M, Hamdi AM, Bryson A, Razonable RR, and Abu Saleh O (2019) Long-term outcomes of patients with human herpesvirus 6 encephalitis. *Open Forum Infect Dis* **6**:ofz269.

Figueiredo J, Lopes-Nunes J, Carvalho J, Antunes F, Ribeiro M, Campello MPC, Paulo A, Paiva A, Salgado GF, Queiroz JA, et al. (2019) AS1411 derivatives as

- carriers of G-quadruplex ligands for cervical cancer cells. *Int J Pharm* **568**: 118511.
- Fleming AM, Ding Y, Alenko A, and Burrows CJ (2016) Zika virus genomic RNA possesses conserved G-quadruplexes characteristic of the Flaviviridae family. *ACS Infect Dis* **2**:674–681.
- Fleming AM, Nguyen NLB, and Burrows CJ (2019) Colocalization of m⁶A and G-quadruplex-forming sequences in viral RNA (HIV, Zika, Hepatitis B, and SV40) suggests topological control of adenosine N⁶-methylation. *ACS Cent Sci* **5**:218–228.
- Fletcher TM, Sun D, Salazar M, and Hurley LH (1998) Effect of DNA secondary structure on human telomerase activity. *Biochemistry* **37**:5536–5541.
- Flint J, Racaniello VR, Rall GF, Hatzioannou T, and Skalka AM (2020) *Principles of Virology, Multi-Volume*, 5th ed, ASM Press, Washington, DC.
- Francés-Monerris A, Hognon C, Miclot T, García-Iriepa C, Iriepa I, Terenzi A, Grandemange S, Barone G, Marazzi M, and Monari A (2020) Molecular basis of SARS-CoV-2 infection and rational design of potential antiviral agents: modeling and simulation approaches. *J Proteome Res* **19**:4291–4315.
- Frasson I, Nadai M, and Richter SN (2019) Conserved G-quadruplexes regulate the immediate early promoters of human *alphaherpesviruses*. *Molecules* **24**:2375.
- Fujiwara N, Mazzola M, Cai E, Wang M, and Cave JW (2015) TMPyP4, a stabilizer of nucleic acid secondary structure, is a novel acetylcholinesterase inhibitor. *PLoS One* **10**:e0139167.
- Gage HL and Merrick CJ (2020) Conserved associations between G-quadruplex-forming DNA motifs and virulence gene families in malaria parasites. *BMC Genomics* **21**:236.
- Gattuso H, Spinello A, Terenzi A, Assfeld X, Barone G, and Monari A (2016) Circular dichroism of DNA G-quadruplexes: combining modeling and spectroscopy to unravel complex structures. *J Phys Chem B* **120**:3113–3121.
- Gazanian E, Lacroix L, Alberti P, Gurung P, Wein S, Cheng M, Mergny JL, Gomes AR, and Lopez-Rubio JJ (2020) Genome wide distribution of G-quadruplexes and their impact on gene expression in malaria parasites. *PLoS Genet* **16**:e1008917.
- Gellert M, Lipssett MN, and Davies DR (1962) Helix formation by guanylic acid. *Proc Natl Acad Sci USA* **48**:2013–2018.
- Gilbert-Girard S, Gravel A, Artusi S, Richter SN, Wallaschek N, Kaufner BB, and Flamand L (2017) Stabilization of telomere G-quadruplexes interferes with human herpesvirus 6A chromosomal integration. *J Virol* **91**:e00402–e00417.
- Glazko VI and Kosovsky GY (2013) Structure of genes coding the envelope proteins of the avian influenza A virus and bovine leucosis virus. *Russ Agric Sci* **39**:511–515.
- Goertz GP, Abbo SR, Fros JJ, and Pijlman GP (2018) Functional RNA during Zika virus infection. *Virus Res* **254**:41–53.
- Guizetti J and Scherf A (2013) Silence, activate, poise and switch! Mechanisms of antigenic variation in Plasmodium falciparum. *Cell Microbiol* **15**:718–726.
- Hamon F, Largy E, Guédin-Beaupaire A, Rouchon-Dagois M, Sidibe A, Monchaud D, Mergny J-L, Riou J-F, Nguyen C-H, and Teulade-Fichou M-P (2011) An acyclic oligoheteroaryle that discriminates strongly between diverse G-quadruplex topologies. *Angew Chem Int Ed Engl* **50**:8745–8749.
- Harris LM and Merrick CJ (2015) G-quadruplexes in pathogens: a common route to virulence control? *PLoS Pathog* **11**:e1004562.
- Hirashima K and Seimiya H (2015) Telomeric repeat-containing RNA/G-quadruplex-forming sequences cause genome-wide alteration of gene expression in human cancer cells in vivo. *Nucleic Acids Res* **43**:2022–2032.
- Ho J, Newcomer MB, Ragain CM, Gascon JA, Batista ER, Loria JP, and Batista VS (2014) MoD-QM/MM structural refinement method: characterization of hydrogen bonding in the oxytricha nova G-quadruplex. *J Chem Theory Comput* **10**:5125–5135.
- Hognon C, Miclot T, Garcé A-Iriepa C, Francés-Monerris A, Grandemange S, Terenzi A, Marazzi M, Barone G, and Monari A (2020) Role of RNA guanine quadruplexes in favoring the dimerization of SARS unique domain in coronaviruses. *J Phys Chem Lett* **11**:5661–5667.
- Hotoda H, Koizumi M, Koga R, Kaneko M, Momota K, Ohmine T, Furukawa H, Agatsuma T, Nishigaki T, Sone J, et al. (1998) Biologically active oligodeoxyribonucleotides. 5. 5'-End-substituted d(TGGGAG) possesses anti-human immunodeficiency virus type 1 activity by forming a G-quadruplex structure. *J Med Chem* **41**:3655–3663.
- Hu X and Ivashkiv LB (2009) Cross-regulation of signaling pathways by interferon-gamma: implications for immune responses and autoimmune diseases. *Immunity* **31**:539–550.
- Huppert JL and Balasubramanian S (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res* **33**:2908–2916.
- Imam H, Khan M, Gokhale NS, McIntyre ABR, Kim GW, Jang JY, Kim SJ, Mason CE, Horner SM, and Siddiqui A (2018) N⁶-methyladenosine modification of hepatitis B virus RNA differentially regulates the viral life cycle. *Proc Natl Acad Sci USA* **115**:8829–8834.
- Jara-Espejo M and Line SR (2020) DNA G-quadruplex stability, position and chromatin accessibility are associated with CpG island methylation. *FEBS J* **287**:483–495.
- Jaubert C, Bedrat A, Bartolucci L, Di Primo C, Ventura M, Mergny JL, Amrane S, and Andreola ML (2018) RNA synthesis is modulated by G-quadruplex formation in Hepatitis C virus negative RNA strand. *Sci Rep* **8**:8120.
- Jeon YJ, Yoo HM, and Chung CH (2010) ISG15 and immune diseases. *Biochim Biophys Acta* **1802**:485–496.
- Ji D, Juhas M, Tsang CM, Kwok CK, Li Y, and Zhang Y (2020) Discovery of G-quadruplex-forming sequences in SARS-CoV-2. *Brief Bioinform* DOI: 10.1093/bib/bbaa114 [published ahead of print].
- Johnson MA, Chatterjee A, Neuman BW, and Wüthrich K (2010) SARS coronavirus unique domain: three-domain molecular architecture in solution and RNA binding. *J Mol Biol* **400**:724–742.
- Kazemier HG, Paeschke K, and Lansdorp PM (2017) Guanine quadruplex monoclonal antibody 1H6 cross-reacts with restrained thymidine-rich single stranded DNA. *Nucleic Acids Res* **45**:5913–5919.
- Kikin O, D'Antonio L, and Bagga PS (2006) QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res* **34**:W676–82.
- Kolesnikova S, Hubálek M, Bednárová L, Cvačka J, and Curtis EA (2017) Multi-merization rules for G-quadruplexes. *Nucleic Acids Res* **45**:8684–8696.
- Krafciková P, Demkovičová E, Halaganová A, and Víglašský V (2017a) Putative HIV and SIV G-quadruplex sequences in coding and noncoding regions can form G-quadruplexes. *J Nucleic Acids* **2017**:6513720.
- Krafciková P, Demkovičová E, and Víglašský V (2017b) Ebola virus derived G-quadruplexes: thiazole orange interaction. *Biochim Biophys Acta Gen Subj* **1861**:1321–1328.
- Kumar M, Seema K, Sharma AK, Prasad A, Sinha N, Ali Bhuttoo Z, and Kumari P (2020) Oncogenic viruses, in *Emerging and Reemerging Viral Pathogens* (Ennaji MM ed), pp 375–409, Elsevier, Amsterdam.
- Kumari N, Vartak SV, Dahal S, Kumari S, Desai SS, Gopalakrishnan V, Choudhary B, and Raghavan SC (2019) G-quadruplex structures contribute to differential radiosensitivity of the human genome. *iScience* **21**:288–307.
- Kusov Y, Tan J, Alvarez E, Enjuanes L, and Hilgenfeld R (2015) A G-quadruplex-binding macrodomain within the “SARS-unique domain” is essential for the activity of the SARS-coronavirus replication-transcription complex. *Virology* **484**:313–322.
- Kyriacou HM, Stone GN, Challis RJ, Raza A, Lyke KE, Thera MA, Koné AK, Doumbo OK, Plowe CV, and Rowe JA (2006) Differential var gene transcription in Plasmodium falciparum isolates from patients with cerebral malaria compared to hyperparasitaemia. *Mol Biochem Parasitol* **150**:211–218.
- Lam EYN, Beraldi D, Tannahill D, and Balasubramanian S (2013) G-quadruplex structures are stable and detectable in human genomic DNA. *Nat Commun* **4**:1796.
- Lavezzo E, Berselli M, Frasson I, Perrone R, Palù G, Brazzale AR, Richter SN, and Toppo S (2018) G-quadruplex forming sequences in the genome of all known human viruses: a comprehensive guide. *PLoS Comput Biol* **14**:e1006675.
- Lee S, Hirohama M, Noguchi M, Nagata K, and Kawaguchi A (2018) Influenza A virus infection triggers pyroptosis and apoptosis of respiratory epithelial cells through the type I interferon signaling pathway in a mutually exclusive manner. *J Virol* **92**:e00396-18.
- Lee WB, Choi WY, Lee DH, Shim H, Kim-Ha J, and Kim YJ (2019) OAS1 and OAS3 negatively regulate the expression of chemokines and interferon-responsive genes in human macrophages. *BMB Rep* **52**:133–138.
- Lei J, Kusov Y, and Hilgenfeld R (2018) Nsp3 of coronaviruses: structures and functions of a large multi-domain protein. *Antiviral Res* **149**:58–74.
- Lejault P, Moruno-Manchon JF, Vemu SM, Honarpisheh P, Zhu L, Kim N, Urayama A, Monchaud D, McCullough LD, and Tsvetkov AS (2020) Regulation of autophagy by DNA G-quadruplexes. *Autophagy* **16**:2252–2259.
- Li Q, Xiang JF, Yang QF, Sun HX, Guan AJ, and Tang YL (2013) G4LDB: a database for discovering and studying G-quadruplex ligands. *Nucleic Acids Res* **41**:D1115–D1123.
- Li YL, Zhao H, and Ren XB (2016) Relationship of VEGF/VEGFR with immune and cancer cells: staggering or forward? *Cancer Biol Med* **13**:206–214.
- Lim KW, Alberti P, Guédin A, Lacroix L, Riou JF, Royle NJ, Mergny JL, and Phan AT (2009) Sequence variant (CTAGGG)_n in the human telomere favors a G-quadruplex structure containing a G.C.G.C tetrad. *Nucleic Acids Res* **37**:6239–6248.
- Limeres Posse J, Diz Dios P, and Scully C, editors (2017) *Viral diseases transmissible by kissing, in Saliva Protection and Transmissible Diseases*, pp 53–92, Elsevier, London.
- Lipps HJ and Rhodes D (2009) G-quadruplex structures: in vivo evidence and function. *Trends Cell Biol* **19**:414–422.
- Lista MJ, Martins RP, Billant O, Contesse M-A, Findakly S, Pochard P, Daskalogianni C, Beauvieux C, Guetta C, Jamin C, et al. (2017) Nucleolin directly mediates Epstein-Barr virus immune evasion through binding to G-quadruplexes of EBNA1 mRNA. *Nat Commun* **8**:16043.
- Lombardi EP and Londoño-Vallejo A (2020) A guide to computational methods for G-quadruplex prediction [published correction appears in *Nucleic Acids Res* (2020) **48**:1–15]. *Nucleic Acids Res* **48**:1603.
- Lyonnais S, Gorelick RJ, Mergny JL, Le Cam E, and Mirambeau G (2003) G-quartets direct assembly of HIV-1 nucleocapsid protein along single-stranded DNA. *Nucleic Acids Res* **31**:5754–5763.
- Lyonnais S, Hounsou C, Teulade-Fichou MP, Jeusset J, Le Cam E, and Mirambeau G (2002) G-quartets assembly within a G-rich DNA flap. A possible event at the center of the HIV-1 genome. *Nucleic Acids Res* **30**:5276–5283.
- Madireddy A, Purushothaman P, Loosbrock CP, Robertson ES, Schildkraut CL, and Verma SC (2016) G-quadruplex-interacting compounds alter latent DNA replication and episomal persistence of KSHV. *Nucleic Acids Res* **44**:3675–3694.
- Majee P, Kumar Mishra S, Pandya N, Shankar U, Pasadi S, Muniyappa K, Nayak D, and Kumar A (2020) Identification and characterization of two conserved G-quadruplex forming motifs in the Nipah virus genome and their interaction with G-quadruplex specific ligands. *Sci Rep* **10**:1477.
- Mao SQ, Ghanbarian AT, Spiegel J, Martínez Cuesta S, Beraldi D, Di Antonio M, Marasco G, Hängel-Hertsch R, Tannahill D, and Balasubramanian S (2018) DNA G-quadruplex structures mold the DNA methylome. *Nat Struct Mol Biol* **25**:951–957.
- Marusić M, Hošnjak L, Krafcikova P, Poljak M, Víglašský V, and Plavec J (2017) The effect of single nucleotide polymorphisms in G-rich regions of high-risk human papillomaviruses on structural diversity of DNA. *Biochim Biophys Acta Gen Subj* **1861**:1229–1236.
- Marusić M and Plavec J (2019) Towards understanding of polymorphism of the G-rich region of human papillomavirus type 52. *Molecules* **24**:1294.
- Matsumoto S, Tateishi-Karimata H, Takahashi S, Ohyama T, and Sugimoto N (2020) Effect of molecular crowding on the stability of RNA G-quadruplexes with various numbers of quartets and lengths of loops. *Biochemistry* **59**:2640–2649.
- Mergny JL, Duval-Valentin G, Nguyen CH, Perrouault L, Faucon B, Rougée M, Montanay-Garestier T, Bisagni E, and Hélène C (1992) Triple helix-specific ligands. *Science* **256**:1681–1684.
- Mergny J-L and Hélène C (1998) G-quadruplex DNA: a target for drug design. *Nat Med* **4**:1366–1367.

- Mergny JL, Phan AT, and Lacroix L (1998) Following G-quartet formation by UV-spectroscopy. *FEBS Lett* **435**:74–78.
- Mergny JL and Sen D (2019) DNA quadruple helices in nanotechnology. *Chem Rev* **119**:6290–6325.
- Métifiot M, Amrane S, Litvak S, and Andreola M-L (2014) G-quadruplexes in viruses: function and potential therapeutic applications. *Nucleic Acids Res* **42**:12352–12366.
- Mishra SK, Jain N, Shankar U, Tawani A, Sharma TK, and Kumar A (2019) Characterization of highly conserved G-quadruplex motifs as potential drug targets in *Streptococcus pneumoniae*. *Sci Rep* **9**:1791.
- Monsen RC and Trent JO (2018) G-quadruplex virtual drug screening: a review. *Biochimie* **152**:134–148.
- Moruno-Manchon JF, Koellhoffer EC, Gopakumar J, Hambarde S, Kim N, McCullough LD, and Tsvetkov AS (2017) The G-quadruplex DNA stabilizing drug pyridostatin promotes DNA damage and downregulates transcription of Breal in neurons. *Aging (Albany NY)* **9**:1957–1970.
- Murat P, Zhong J, Lekieffre L, Cowieson NP, Clancy JL, Preiss T, Balasubramanian S, Khanna R, and Tellam J (2014) G-quadruplexes regulate Epstein-Barr virus-encoded nuclear antigen 1 mRNA translation. *Nat Chem Biol* **10**:358–364.
- Norseen J, Johnson FB, and Lieberman PM (2009) Role for G-quadruplex RNA binding by Epstein-Barr virus nuclear antigen 1 in DNA replication and metaphase chromosome attachment. *J Virol* **83**:10336–10346.
- Palazon A, Goldrath AW, Nizet V, and Johnson RS (2014) HIF transcription factors, inflammation, and immunity. *Immunity* **41**:518–528.
- Panera N, Tozzi AE, and Alisi A (2020) The G-quadruplex/helicase world as a potential antiviral approach against COVID-19. *Drugs* **80**:941–946.
- Patel PK, Bhavesh NS, and Hosur RV (2000) NMR observation of a novel C-tetrad in the structure of the SV40 repeat sequence GGGCGG. *Biochem Biophys Res Commun* **270**:967–971.
- Perrone R, Butovskaya E, Daelemans D, Palù G, Pannecouque C, and Richter SN (2014) Anti-HIV-1 activity of the G-quadruplex ligand BRACO-19. *J Antimicrob Chemother* **69**:3248–3258.
- Perrone R, Doria F, Butovskaya E, Frasson I, Botti S, Scalabrin M, Lago S, Grande V, Nadai M, Freccero M, et al. (2015) Synthesis, binding and antiviral properties of potent core-extended naphthalene diimides targeting the HIV-1 long terminal repeat promoter G-quadruplexes. *J Med Chem* **58**:9639–9652.
- Perrone R, Nadai M, Frasson I, Poe JA, Butovskaya E, Smithgall TE, Palumbo M, Palù G, and Richter SN (2013a) A dynamic G-quadruplex region regulates the HIV-1 long terminal repeat promoter. *J Med Chem* **56**:6521–6530.
- Perrone R, Nadai M, Poe JA, Frasson I, Palumbo M, Palù G, Smithgall TE, and Richter SN (2013b) Formation of a unique cluster of G-quadruplex structures in the HIV-1 Nef coding region: implications for antiviral activity. *PLoS One* **8**:e73121.
- Phan AT, Kuryavvi V, and Patel DJ (2006) DNA architecture: from G to Z. *Curr Opin Struct Biol* **16**:288–298.
- Piekna-Przybylska D, Bambara RA, Maggirwar SB, and Dewhurst S (2020) G-quadruplex ligands targeting telomeres do not inhibit HIV promoter activity and cooperate with latency reversing agents in killing latently infected cells. *Cell Cycle* **19**:2298–2313.
- Piekna-Przybylska D, Sharma G, and Bambara RA (2013) Mechanism of HIV-1 RNA dimerization in the central region of the genome and significance for viral evolution. *J Biol Chem* **288**:24140–24150.
- Piekna-Przybylska D, Sharma G, Maggirwar SB, and Bambara RA (2017) Deficiency in DNA damage response, a new characteristic of cells infected with latent HIV-1. *Cell Cycle* **16**:968–978.
- Plyler J, Jasheway K, Tuesuwan B, Karr J, Brennan JS, Kerwin SM, and David WM (2009) Real-time investigation of SV40 large T-antigen helicase activity using surface plasmon resonance. *Cell Biochem Biophys* **53**:43–52.
- Previdelli RL, Bertzbach LD, Wight DJ, Vychodil T, You Y, Arndt S, and Kaufer BB (2019) The role of Marek's disease virus UL12 and UL29 in DNA recombination and the virus lifecycle. *Viruses* **11**:111.
- Priester LL, Yin S, Cahoon LA, and Seifert HS (2020) Altering the *Neisseria gonorrhoeae pilE* guanine quadruplex loop bases affects pilin antigenic variation. *Biochemistry* **59**:1104–1112.
- Puig Lombardi E, Londoño-Vallejo A, and Nicolas A (2019) Relationship between G-quadruplex sequence composition in viruses and their hosts. *Molecules* **24**:1942.
- Rajendran A, Endo M, Hidaka K, Tran PLT, Mergny JL, Gorelick RJ, and Sugiyama H (2013) HIV-1 nucleocapsid proteins as molecular chaperones for tetramolecular antiparallel G-quadruplex formation. *J Am Chem Soc* **135**:18575–18585.
- Randazzo A, Galeone A, Esposito V, Varra M, and Mayol L (2002) Interaction of distamycin A and netropsin with quadruplex and duplex structures: a comparative 1H-NMR study. *Nucleosides Nucleotides Nucleic Acids* **21**:535–545.
- Ravichandran S, Kim Y-E, Bansal V, Ghosh A, Hur J, Subramani VK, Pradhan S, Lee MK, Kim KK, and Ahn J-H (2018) Genome-wide analysis of regulatory G-quadruplexes affecting gene expression in human cytomegalovirus. *PLoS Pathog* **14**:e1007334.
- Ray P, Krishnamoorthy N, Oriss TB, and Ray A (2010) Signaling of c-kit in dendritic cells influences adaptive immunity. *Ann N Y Acad Sci* **1183**:104–122.
- Ray S, Tillo D, Boer RE, Assad N, Barshai M, Wu G, Orenstein Y, Yang D, Schneekloth JS Jr, and Vinson C (2020) Custom DNA microarrays reveal diverse binding preferences of proteins and small molecules to thousands of G-quadruplexes. *ACS Chem Biol* **15**:925–935.
- Reina C and Cavalieri V (2020) Epigenetic modulation of chromatin states and gene expression by g-quadruplex structures. *Int J Mol Sci* **21**:1–22.
- Renault TT and Chipuk JE (2013) Getting away with murder: how does the BCL-2 family of proteins kill with immunity? *Ann N Y Acad Sci* **1285**:59–79.
- Reznichenko O, Quillévère A, Martins RP, Loaec N, Kang H, Lista MJ, Beauvineau C, González-García J, Guillot R, Voisset C, et al. (2019) Novel cationic bis(acylhydrazones) as modulators of Epstein-Barr virus immune evasion acting through disruption of interaction between nucleolin and G-quadruplexes of EBNA1 mRNA. *Eur J Med Chem* **178**:13–29.
- Romanucci V, Zarrelli A, and Di Fabio G (2019) Hotoda's sequence and anti-HIV activity: where are we now? *Molecules* **24**:1417.
- Ruggiero E and Richter SN (2018) G-quadruplexes and G-quadruplex ligands: targets and tools in antiviral therapy. *Nucleic Acids Res* **46**:3270–3283.
- Ruggiero E and Richter SN (2020) Viral G-quadruplexes: new frontiers in virus pathogenesis and antiviral therapy, in *Annual Reports in Medicinal Chemistry* (Neidle S ed) pp 101–131, Academic Press, Cambridge, MA.
- Ruggiero E, Tassinari M, Perrone R, Nadai M, and Richter SN (2019) Stable and conserved G-quadruplexes in the long terminal repeat promoter of retroviruses. *ACS Infect Dis* **5**:1150–1159.
- Rusmini M, Griseri P, Lantieri F, Matera I, Hudspeth KL, Roberto A, Mikulak J, Avanzini S, Rossi V, Mattioli G, et al. (2013) Induction of RET dependent and independent pro-inflammatory programs in human peripheral blood mononuclear cells from Hirschsprung patients. *PLoS One* **8**:e59066.
- Saranathan N, Biswas B, Patra A, and Vivekanandan P (2019) G-quadruplexes may determine the landscape of recombination in HSV-1. *BMC Genomics* **20**:382.
- Saranathan N and Vivekanandan P (2019) G-quadruplexes: more than just a kink in microbial genomes. *Trends Microbiol* **27**:148–163.
- Satkunanathan S, Thorpe R, and Zhao Y (2017) The function of DNA binding protein nucleophosmin in AAV replication. *Virology* **510**:46–54.
- Scalabrin M, Frasson I, Ruggiero E, Perrone R, Tosoni E, Lago S, Tassinari M, Palù G, and Richter SN (2017) The cellular protein hnRNP A2/B1 enhances HIV-1 transcription by unfolding LTR promoter G-quadruplexes. *Sci Rep* **7**:45244.
- Shen LW, Qian MQ, Yu K, Narva S, Yu F, Wu YL, and Zhang W (2020) Inhibition of Influenza A virus propagation by benzoselenoxanthenes stabilizing TMPRSS2 Gene G-quadruplex and hence down-regulating TMPRSS2 expression. *Sci Rep* **10**:7635.
- Stavnezer J and Schrader CE (2014) IgH chain class switch recombination: mechanism and regulation. *J Immunol* **193**:5370–5378.
- Sun D, Thompson B, Cathers BE, Salazar M, Kerwin SM, Trent JO, Jenkins TC, Neidle S, and Hurley LH (1997) Inhibition of human telomerase by a G-quadruplex-interactive compound. *J Med Chem* **40**:2113–2116.
- Sun H, Yabuki A, and Maizels N (2001) A human nuclease specific for G4 DNA. *Proc Natl Acad Sci USA* **98**:12444–12449.
- Sun Z-Y, Wang X-N, Cheng S-Q, Su X-X, and Ou T-M (2019) Developing novel G-quadruplex ligands: from interaction with nucleic acids to interfering with nucleic acid-protein interaction. *Molecules* **24**:396.
- Tamura Y, Yoshida M, Suzuki J, Hiratou T, Miyano-Kurosaki N, Takai K, and Takaku H (2000) Properties of quadruplex oligonucleotides with anti-HIV-1 activity. *Nucleic Acids Symp Ser* **44**:181–182.
- Tan J, Vonrhein C, Smart OS, Bricogne G, Bollati M, Kusov Y, Hansen G, Mesters JR, Schmidt CL, and Hilgenfeld R (2009) The SARS-unique domain (SUD) of SARS coronavirus contains two macrodomains that bind G-quadruplexes. *PLoS Pathog* **5**:e1000428.
- Tasset DM, Kubik MF, and Steiner W (1997) Oligonucleotide inhibitors of human thrombin that bind distinct epitopes. *J Mol Biol* **272**:688–698.
- Tassinari M, Lena A, Butovskaya E, Pirotta V, Nadai M, Freccero M, Doria F, and Richter SN (2018) A fragment-based approach for the development of G-quadruplex ligands: role of the amidoxime moiety. *Molecules* **23**:1874.
- Tassinari M, Zuffo M, Nadai M, Pirotta V, Sevilla Montalvo AC, Doria F, Freccero M, and Richter SN (2020) Selective targeting of mutually exclusive DNA G-quadruplexes: HIV-1 LTR as paradigmatic model. *Nucleic Acids Res* **48**:4627–4642.
- Tellam JT, Zhong J, Lekieffre L, Bhat P, Martinez M, Croft NP, Kaplan W, Tellam RL, and Khanna R (2014) mRNA Structural constraints on EBNA1 synthesis impact on in vivo antigen presentation and early priming of CD8+ T cells. *PLoS Pathog* **10**:e1004423.
- Tlučková K, Marušić M, Tóthová P, Bauer L, Šket P, Plavec J, and Viglaský V (2013) Human papillomavirus G-quadruplexes. *Biochemistry* **52**:7207–7216.
- Todd AK, Johnston M, and Neidle S (2005) Highly prevalent tetraplex sequence motifs in human DNA. *Nucleic Acids Res* **33**:2901–2907.
- Tosoni E, Frasson I, Scalabrin M, Perrone R, Butovskaya E, Nadai M, Palù G, Fabris D, and Richter SN (2015) Nucleolin stabilizes G-quadruplex structures folded by the LTR promoter and silences HIV-1 viral transcription. *Nucleic Acids Res* **43**:8884–8897.
- Tran PLT, Largy E, Hamon F, Teulade-Fichou MP, and Mergny JL (2011) Fluorescence intercalator displacement assay for screening G4 ligands towards a variety of G-quadruplex structures. *Biochimie* **93**:1288–1296.
- Tuesuwan B, Kern JT, Thomas PW, Rodriguez M, Li J, David WM, and Kerwin SM (2008) Simian virus 40 large T-antigen G-quadruplex DNA helicase inhibition by G-quadruplex DNA-interactive agents. *Biochemistry* **47**:1896–1909.
- Turrini F, Marelli S, Kajaste-Rudnitski A, Lusica M, Van Lint C, Das AT, Harwig A, Berkhout B, and Vicenzi E (2015) HIV-1 transcriptional silencing caused by TRIM22 inhibition of Sp1 binding to the viral promoter. *Retrovirology* **12**:104.
- van Maldegem F and Downward J (2020) Mutant KRAS at the heart of tumor immune evasion. *Immunity* **52**:14–16.
- Vansant G, Chen HC, Zorita E, Trejbalová K, Miklík D, Filion G, and Debyser Z (2020) The chromatin landscape at the HIV-1 provirus integration site determines viral expression. *Nucleic Acids Res* **48**:7801–7817.
- Wang S-R, Min Y-Q, Wang J-Q, Liu C-X, Fu B-S, Wu F, Wu L-Y, Qiao Z-X, Song Y-Y, Xu G-H, et al. (2016a) A highly conserved G-rich consensus sequence in hepatitis C virus core gene represents a new anti-hepatitis C target. *Sci Adv* **2**:e1501535.
- Wang S-R, Zhang Q-Y, Wang J-Q, Ge X-Y, Song Y-Y, Wang Y-F, Li X-D, Fu B-S, Xu G-H, Shu B, et al. (2016b) Chemical targeting of a G-quadruplex region in the Ebola virus L gene. *Cell Chem Biol* **23**:1113–1122.
- Wei PC, Wang ZF, Lo WT, Su MI, Shew JY, Chang TC, and Lee WH (2013) A cis-element with mixed G-quadruplex structure of NPGPx promoter is essential for nucleolin-mediated transactivation on non-targeting siRNA stress. *Nucleic Acids Res* **41**:1533–1543.

- Westdorp KN and Terhune SS (2018) Impact of RNA polymerase I inhibitor CX-5461 on viral kinase-dependent and -independent cytomegalovirus replication. *Antiviral Res* **153**:33–38.
- Wyatt JR, Vickers TA, Roberson JL, Buckheit RW Jr, Klimkait T, DeBaets E, Davis PW, Rayner B, Imbach JL, and Ecker DJ (1994) Combinatorially selected guanosine-quartet structure is a potent inhibitor of human immunodeficiency virus envelope-mediated cell fusion. *Proc Natl Acad Sci USA* **91**:1356–1360.
- Xu H, Di Antonio M, McKinney S, Mathew V, Ho B, O'Neil NJ, Santos ND, Silvester J, Wei V, Garcia J, et al. (2017) CX-5461 is a DNA G-quadruplex stabilizer with selective lethality in BRCA1/2 deficient tumours. *Nat Commun* **8**:14432.
- Ye Q, Wang B, and Mao J (2020) The pathogenesis and treatment of the 'Cytokine Storm' in COVID-19. *J Infect* **80**:607–613 DOI: 10.1016/j.jinf.2020.03.037.
- Zhang Y, Liu S, Jiang H, Deng H, Dong C, Shen W, Chen H, Gao C, Xiao S, Liu ZF, et al. (2020) G₂-quadruplex in the 3'UTR of IE180 regulates Pseudorabies virus replication by enhancing gene expression. *RNA Biol* **17**:816–827.
- Zhao C, Qin G, Niu J, Wang Z, Wang C, Ren J, and Qu X (2021) Targeting RNA G-quadruplex in SARS-CoV-2: a promising therapeutic target for COVID-19? *Angew Chem Int Ed Engl* **60**:432–438.
- Zhou G, Liu X, Li Y, Xu S, Ma C, Wu X, Cheng Y, Yu Z, Zhao G, and Chen Y (2016) Telomere targeting with a novel G-quadruplex-interactive ligand BRACO-19 induces T-loop disassembly and telomerase displacement in human glioblastoma cells. *Oncotarget* **7**:14925–14939.
- Zuffo M, Guédin A, Leriche E-D, Doria F, Pirola V, Gabelica V, Mergny J-L, and Freccero M (2018) More is not always better: finding the right trade-off between affinity and selectivity of a G-quadruplex ligand. *Nucleic Acids Res* **46**:e115.