Immunogenetics (2022) 74:455–463
https://doi.org/10.1007/s00251-022-01257-6

ORIGINAL ARTICLE

Identification of G-quadruplex DNA sequences in SARS-CoV2

Amit K. Maiti[1]

Received: 17 January 2022 / Accepted: 3 February 2022 / Published online: 18 March 2022

Abstract
G-quadruplex structure or Putative Quadruplex Sequences (PQSs) are abundant in human, microbial, DNA, or RNA viral genomes. These sequences in RNA viral genome play critical roles in integration into human genome as LTR (Long Terminal Repeat), genome replication, chromatin rearrangements, gene regulation, antigen variation (Av), and virulence. Here, we investigated whether the genome of SARS-CoV2, an RNA virus, contained such potential G-quadruplex structures. Using bioinformatic tools, we searched for such sequences and found thirty-seven (forward strand (twenty-five) + reverse strand (Twelve)) QGRSs (Quadruplex forming G-Rich Sequences)/PQSs in SARS-CoV2 genome. These sequences are dispersed mainly in the upstream of SARS-CoV2 genes. We discuss whether existing PQS/QGRS ligands could inhibit the SARS-CoV2 replication and gene transcription as has been observed in other RNA viruses. Further experimental validation would determine the role of these G-quadruplex sequences in SARS-CoV2 genome function to survive in the host cells and identify therapeutic agents to destabilize these PQSs/QGRSs.

Keywords SARS-CoV2 · G-quadruplex · LTR sequences · Therapeutics · Upstream

Introduction

The SARS-CoV2 virus that causes covid-19 disease created pandemic worldwide with considerable mortalities. SARS-CoV2 has a protein coat with spikes and a positive ~ 30 kb (29,903 base) RNA strand (Ren et al. 2020a, Zhou et al. 2020). After attaching to the host cell with its spike proteins, it uses the ACE2 receptor as well as the TMPRSS2 enzyme to enter into host cells. It also uses host machinery to replicate its RNA genome to make thousands of RNA molecules (Hoffmann et al. 2020). A classical retrovirus multiplication consists of several steps (Wessner 2010) that includes reverse transcription to produce cDNA by its own enzyme RDRP (RNA directed DNA polymerase), and it integrates into host genome to replicate and transcribe for producing its own proteins. Unlike retroviruses, SARS-CoV2 as a coronavirus does not integrate into human genome for its survival, but occasional integration could not be ruled out. Current report suggests that SARS-CoV2 indeed integrate in the human genome in cell culture model and in covid-19 patients' nasopharyngeal cells (Zhang et al. 2021).

The SARS-CoV2 is a novel virus, and its molecular characterization of replication mechanism is in the very early stages of investigation. PQSs are known to pause replication forks, delay replication, and play a role in DNA damage and repair. A special helicase, such as Bloom helicase, acts on RNA genome to resolve G-quadruplex structure to load RDRP. In addition, all G-quadruplex helicases use ATP hydrolytic energy to unwind G-quadruplex structures (Chatterjee et al. 2014). G-quadruplexes located upstream or downstream of the Transcription Start Site (TSS) can inhibit or promote transcription. The PQSs recruit translation initiation proteins or block these proteins binding to the Untranslated Region (UTR), and modulate translation (Xu et al. 2021). Host protein nucleolin (NCL) binds with core PQS of HCV virus, and unavailable free NCL substantially increases viral replication.

The PQSs/QGRSs are also required for RNA synthesis and act as initiation sites for transcription of the viral genome (Poltronieri et al. 2015). PQSs also regulate other molecular mechanisms in RNA biology such as splicing, ribosomal frameshifting, mRNA translocations, Repeat-Associated Non-AUG (RAN) translation, and maturation of miRNAs. It also forms intermolecular Hybrid Quadruplex

✉ Amit K. Maiti
akmit123@yahoo.com; amit.maiti@mydnavar.com

[1] Mydnavar, Department of Genetics and Genomics, 2645 Somerset Boulevard, Troy, MI 48084, USA

(HQ) between non-template DNA, nascent mRNA and acts as a transcription-termination signal (Saranathan and Vivekanandan 2019). In addition to gene expression, it also acts as recombination hotspot that involves recombination of genomic segments leading to the production of altered surface proteins to generate antigenic variation (Av) (Deitsch et al. 1997). For EBV (Epstein-Barr Virus) and HIV1, PQSs not only facilitate to generate Av but also contribute to virulence (Harris and Merrick 2015).

PQS often consists of G-quadruplex in the sequences. A quartet of G is the main core of the quadruplex, and the surrounding sequences make it a perfect quadruplex (Fig. 1). G-quadruplex provides stability of the LTR sequences in a stem and loop structure (Butovskaya et al. 2018; Ruggiero et al. 2019). It is observed that promoters of the genes in *Mycobacterium tuberculosis* possess G-quadruplex sequences that enable HIV1 LTR to integrate (Perrone et al. 2017). Putative G-quadruplex in coding and non-coding sequences indeed forms G-quadruplex (Krafčíková et al. 2017; Ruggiero et al. 2019). Human telomeres are mostly made of these G-quadruplex that are resolved and replicated by telomerase (Lin et al. 2001; Mills et al. 2002).

As PQS negatively regulates viral replication, transcription, and translation, it forms the basis of PQS ligands as antiviral agents. Several PQS ligands are being currently tested for HIV1, ZIKA (ZIKV) and HCV etc. (Saranathan and Vivekanandan 2019). However, their specificity is a major concern as these ligands may also bind with many host PQSs. These concerns likely to be resolved as the number of virus particles with their PQSs could outnumber the PQSs of single host cells (Ruggiero et al. 2021). In addition, other strategies such as PQS-specific antibody, anti-oligo against specific PQS, or CRISPR editing to disrupt the quadruplex structure could be important tools to control the virus amplification (Xu et al. 2021).

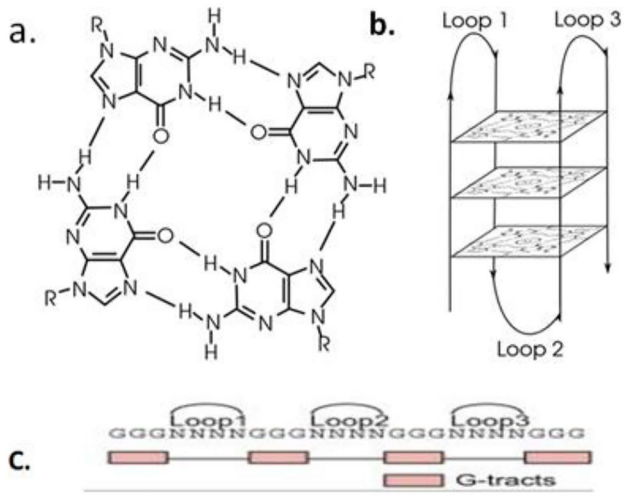

Fig. 1 Structure of G-quadruplex. **a.** QGRS sequences contain G tract **b.** loop structures and **c.** loop sequences that may form G-quadruplex

Here, we searched for PQS sequences in silico in SARS-CoV2 genome and located their position in respect to twenty-nine genes of SARS-CoV2. We discussed their possible role in chromatin rearrangements and gene regulation in which genes they belong and that could have potential impact in its replication, transcription, and survival. We also hypothesized further investigative approach of known and existing PQS ligands to inhibit SARS-CoV2 amplification and gene regulation.

Materials and methods

G-quadruplex mapping

G-quadruplex sequences are mapped on a standard consensus SARS-CoV2 genome (MN908947) in the website QGRS mapper (https://bioinformatics.ramapo.edu/QGRS/analyze.php) (Kikin et al. 2006). The search criteria were of maximum 30 base length and loop consisting of any length from 0 to 30 base sequences. This software gives output as lists of QGRS/PQS sequences for unique and overlapping sequence stretch. After scoring QGRS/PQS from forward strand, same SARS-CoV2 genomic sequences are reverse complemented and again analyzed to find QGRS/PQS sequences. The probability of forming G-quadruplex structure of each sequence is obtained as G-scores as an output of the software.

The position of each forward and reverse strand QGRS/PQS sequences in the SARS-CoV2 genome is listed (software output) and represented in a map using the position of each gene of SARS-CoV2.

Results

QGRS/PQS sequences are dispersed in SARS-CoV2 genome

The ability of a DNA or RNA stretch to form G-quadruplex structure depends on the constituents of the sequences comprising stem and a loop. The composition of sequences determines the stability of the G-quadruplex. In silico methods comprise a score system that determines the stability of potential sequence stretch to form a G-quadruplex structure (Kikin et al. 2006). To search for PQS sequences in SARS-CoV2 genome, we used in silico methods using a standard software that provides G-score to identify G-quadruplex sequences in the genome. We identified several sequences in the SARS-CoV2 genome

Table 1 QGRS/PQS sequences in forward strand (overlaps not included)

Base position in SARS-CoV2 genome	Length	QGRS/PQS	G-Score
353	25	GGCTTTGGAGACTCCGTGGAGGAGG	16
644	20	GGTAATAAAGGAGCTGGTGG	15
1463	26	GGTGGTCGCACTATTGCCTTTGGAGG	6
1574	26	GGTGTTGTTGGAGAAGGTTCCGAAGG	18
2714	29	GGCGGTGCACCAACAAAGGTTACTTTTGG	10
3467	17	GGAGGAGGTGTTGCAGG	15
4162	27	GGTTATACCTACTAAAAAGGCTGGTGG	6
4261	29	GGGTTTAAATGGTTACACTGTAGAGGAGG	10
8687	23	GGATACAAGGCTATTGATGGTGG	14
10,261	30	GGCTGGTAATGTTCAACTCAGGGTTATTGG	9
13,385	20	GGTATGTGGAAAGGTTATGG	19
14,947	28	GGTTTTCCATTTAATAAATGGGGTAAGG	4
15,208	27	GGAACAAGCAAATTCTATGGTGGTTGG	6
15,448	29	GGCGGTTCACTATATGTTAAACCAGGTGG	3
18,296	23	GGATTGGCTTCGATGTCGAGGGG	9
22,316	29	GGTGATTCTTCTTCAGGTTGGACAGCTGG	10
24,215	20	GGTTGGACCTTTGGTGCAGG	17
24,268	24	GGCTTATAGGTTTAATGGTATTGG	19
25,197	22	GGCCATGGTACATTTGGCTAGG	17
25,951	29	GGTGGTTATACTGAAAAATGGGAATCTGG	8
26,746	30	GGATCACCGGTGGAATTGCTATCGCAATGG	7
28,781	29	GGCTTCTACGCAGAAGGGAGCAGAGGCGG	9
28,903	15	GGCTGGCAATGGCGG	18
29,123	19	GGAAATTTTGGGGACCAGG	14
29,234	30	GGCATGGAAGTCACACCTTCGGGAACGTGG	11

that have potential to form G-quadruplex structure (Tables 1 and 2). Although G-scores are not high, thirty-seven sequence (forward (twenty-five) + reversed (twelve)) stretches (without overlapping sequences) have potential to form G-quadruplex structure. When overlapping sequences are considered, total 75 (45 forward and 30 reverse strand) QGRS/PQS are observed (Supplementary Table 1 and 2).

Potential G-quadruplex sequences are abundant in the upstream of SARS-CoV2 genes

QGRSs/PQSs are demonstrated to be present at the TSS of many genes (Kikin et al. 2006; Perrone et al. 2017) although their exact functions in transcription are not well understood. QGRSs/PQSs are shown to play critical role in forming a

Table 2 QGRS/PQS sequences in reverse strand (overlaps not included)

Base position in SARS-CoV2 genome	Length	QGRS/PQS	G-Score
165	22	GGCCTCGGTGAAAATGTGGTGG	13
2987	29	GGTCTGGTCAGAATAGTGCCATGGAGTGG	9
6822	30	GGTTGGTAACCAACACCATTAGTGGGTTGG	6
11,440	25	GGCGGTGGTTTAGCACTAACTCTGG	7
13,136	19	GGTTAAGTGGTGGTCTAGG	16
13,963	18	GGATCTGGGTAAGGAAGG	19
16,623	27	GGATTTGGATGATCTATGTGGCAACGG	14
19,865	25	GGTGATAGAGGTTTGTGGTGGTTGG	19
23,877	17	GGATATGGTTGGTTTGG	19
25,003	16	GGTGGAATGTGGTAGG	17
27,432	24	GGGGCTTTTAGAGGCATGAGTAGG	13
29,867	23	GGTTGGTTTGTTACCTGGGAAGG	13

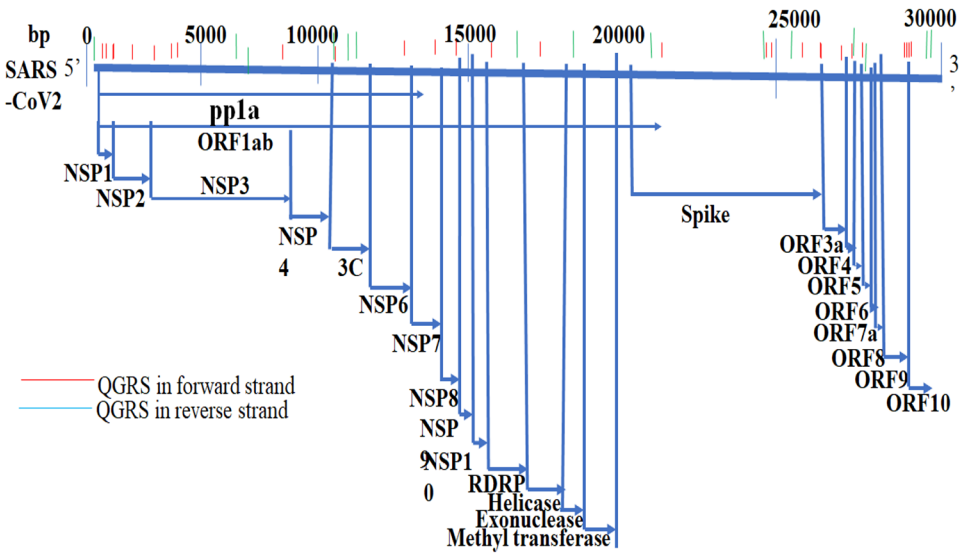

Fig. 2 Location of QGRS/PQS sequences in SARS-CoV2 genome. Most of the QGRSs/PQSs are located upstream of genes

dynamic nucleosome positioning or an RNA–DNA hybrid during transcription (Ohmori and Tsuruyama 2012; Zhang et al. 2020a, b). I observed the upstream of SARS-CoV2 genes is populated with these sequences. SARS-CoV2 genome codes for 29 functional proteins or peptides (Gordon et al. 2020). The identified QGRSs/PQSs are present at the upstream of (−400 bp) of at least 16 proteins (Fig. 2). Among them, 13 genes have QGRS/PQS within – 250 bp of upstream and likely to act as TSS although exact TSS for each SARS-CoV2 gene is yet to be documented.

Significance of QGRS/PQS in upstream of SARS-CoV2 genes

Although QGRS/PQS in any position of a gene can modulate gene expression, and until further characterization of promoters of each gene is documented, I focused those SARS-CoV2 genes that have QGRS/PQS sequences in the upstream at the putative promoter region. The expression of these genes could be modulated by chromatin rearrangements or DNA stability, inhibiting RNA splicing or translation (Machida et al. 2020).

NSP1 ((reverse strand (R) -100 base (b)) NSP1 is a virulence factor of the virus that binds to the human 40S ribosomal complex to inhibit the mRNA entry for translation by the host proteins (Schubert et al. 2020). The inhibition of translation completely shuts down the innate immune system of human (Thoms et al. 2020). NSP1 is also known to interact with POLA1, POLA2, PRIM1, PRIM2, PKP, and COLGALT1 (Gordon et al. 2020). Thus, modulation of expression of NSP1 gene should have tremendous impact for SARS-oV2 survival in the human body.

NSP2 ((forward strand (F)) -47b, -56b) NSP2 contains an endosome-associated protein–like domain that accounts for highly contagious properties of SARS-CoV2 (Angeletti et al. 2020). NSP2 interacts with EIF4A2, SLC27A2, FKBP14, and WASHC4 (Gordon et al. 2020). NSP2 also binds with STAU2, HNRNPL1, ATP6V1B2, RAP1GDS and is involved in several biological processes such as endosome transport and translation (Zheng et al. 2021).

NSP3 ((F) +69b, -268b(R)) NSP3 is a large protein encoded by a SARS-CoV2 gene and consists of eight domains, such as the ubiquitin-like domain 1 (Ubl1), the Glu-rich acidic domain, a macrodomain (Mac1 or X domain), the ubiquitin-like domain 2 (Ubl2), the papain-like protease 2 ($PL2^{pro}$), the Nsp3 ectodomain (3Ecto, "zinc-finger domain"), and the domains Y1 and CoV-Y (Lei et al. 2018). Ubl1 in CoV virus binds with single-strand (ssRNA) and interacts with the nucleocapsid (N) protein (Hurst et al. 2013). Papain-like domain is demonstrated to provide cleavage site in NSP3 for ubiquitination (Wojdyla et al. 2010). X domain (Mac1) is indispensable for RNA replication in the context of a SARS-CoV replicon (Kusov et al. 2015). Evidences showed that the X domain played a role in counteracting the host innate immune response (Fehr et al. 2016).

NSP4 ((F)-238b) NSP4 interacts with mitochondrial proteins TIMM9, TIMM10, TIMM10B, TIMM29, and TIMM (Gordon et al. 2020). NSP4 bindings are unique to each strain, such as E3 ubiquitin ligase complexes for SARS-CoV-1 and ER homeostasis factors for SARS-CoV-2. Common NSP4 interactors include *N*-linked glycosylation machinery, unfolded protein response (UPR)-associated proteins, and anti-viral innate immune signaling factors. NSP4 interactors are strongly enriched in proteins localized at

mitochondrial-associated ER membranes suggesting a new functional role for modulating host processes, such as calcium homeostasis at these organelles contact sites (Davies et al. 2020).

RDRP ((F) +45b) The RDRP is the RNA polymerase essential for SARS-CoV2 replication. The structure of RDRP complexes consists of non-structural protein NSP12, NSP8, NSP7 and more than two turns of RNA template–product duplex. The active-site cleft of NSP12 binds to the first turn of RNA (Hillen et al. 2020). Evolutionary studies of whole-genome sequences of SARS-CoV2 represent high similarities (> 90%) with other SARS viruses. Targeting the RDRP active sites, residues ASP760 and ASP761, by antiviral drugs could be a potential therapeutic option for inhibition of SARS-CoV2 RDRP and thus viral replication (Aftab et al. 2020).

NSP10 ((R) -112b) NSP10 is a critical co-factor and major regulator of SARS-CoV2 replicase function. Reverse genetics experiments supported an essential role of the NSP10 that interacts with NSP14 in SARS-CoV1 replication, as several mutations that abolished the interaction in vitro yielded a replication negative viral phenotype. In contrast, mutants in which the NSP10-NSP16 interaction was disturbed, they proved to be crippled but viable. These experiments imply that the NSP10 interacts with NSP14 and NSP16 and may be a target for the development of antiviral compounds against pathogenic coronaviruses (Bouvet et al. 2014). SARS-CoV2 NSP10/NSP14 also functions as an exoribonuclease with both structural and functional integrity (Lin et al. 2021).

ORF3A ((F)-84b, -90b) ORF3a is a viroporin, an integral membrane protein able to function as an ion channel that may promote virus release (Issa et al. 2020). This protein interacts with host caveolin and regulates different phases of viral cycle. ORF3a has also a TRAF3-binding motif that activates the NLRP3 inflammasome and acts as a potent stimulator of pro–IL-1β gene transcription. In animal models of SARS-CoV1, genomic deletion of ORF3a reduced virus replication (Castaño-Rodriguez et al. 2018). Importantly, in SARS-CoV2-infected individuals, significant $CD4^+$ and $CD8^+$ T cell responses were naturally directed against ORF3a. ORF3a also induces apoptosis in host cells (Ren et al. 2020b).

ORF4 (E protein) ((F) -208b) This protein is mostly confined around the ER (endoplasmic reticulum), Golgi complex, and ER-Golgi intermediate compartment (ERGIC), where it is involved in assembly of virus particles and budding. E-protein may act as an ion channel, specifically like viroporin, which then oligomerizes and enhances membrane permeability. It is crucial for apoptosis and possesses an N-terminal, a transmembrane, and a C-terminal domain. Furthermore, the transmembrane region of the protein is responsible for homotypic interactions of the protein as analyzed by Molecular Dynamics (MD) simulations (Kaur et al. 2021). NSP4 interacts with NSP3 and host proteins to confer membrane rearrangement in SARS CoV1. Moreover, the interaction between NSP4 and NSP3 is essential for viral replication (Sakai et al. 2017).

ORF7A ((R) -39b) This protein has biochemical properties consistent with that of an integral membrane protein (Kaur et al. 2021). ORF7A localizes in the Golgi compartment. The ORF7A protein is not only an accessory protein but also a structural component of the SARS-CoV2 virion (Hassan et al. 2021). ORF7A is demonstrated to interact with MDN1 and HEATR3 (Gordon et al. 2020).

ORF10 ((F)-218b) SARS-CoV2 ORF10 are predicted to alter binding affinity to respective HLA alleles and correspondingly changes the immunogenicity of SARS-CoV2 invasion. ORF10 predominantly decreases the affinity of epitopes to escape the host-immune system (Hassan et al. 2021). ORF10 interacts with CUL2 complex comprising CUL2, RBX1, ELOB, ELOC, and ZYG11B (Gordon et al. 2020).

Therapeutic aspect of PQS ligands to inhibit amplification and gene expression of SARS-CoV2

Despite considerable achievements in antiviral research, new and emerging viruses constantly threat human. Novel and innovative therapies are needed to control these viruses. Small molecules that serve as PQS ligands have been developed to target G-quadruplex structures. The common structural characters of these ligands have a heterocyclic core that binds with G-quadruplex and have physiological stabilities, higher binding affinities with low cytotoxic activities. Some of such ligands are mentioned here that could be tested for controlling SARs-CoV2 growth to develop therapeutics.

BRACO-19 (the N, N0-(9-((4-(dimethylamino) phenyl)amino) acridine-3,6-diyl)bis(3-(pyrrolidin-1-yl)propanamide) It is a trisubstituted acridine molecule widely used as PQS ligands for its high specificity and low cytotoxicity. Apart from DNA virus, it was used against PQS of HIV1 and ZIKV, and in both cases, it was effective to stall the replication and decreased virus titer (Ruggiero et al. 2021). Braco-19 is currently being tested against SUD (SARS Unique Domain) PQS in NSP3 gene of SARS-COV1 and SARS-CoV2, and promising results are obtained to inhibit the viral amplification (Zhang et al. 2020b; Cui and Zhang 2020).

TMPyP4 (*N*-methyl-4-pyridyl) porphyrin) TMPyP4 is the most extensively studied molecule against PQS for its size and hydrophobicity. It has been used against several human herpes viruses (HHVs), human papillomavirus (HPV), and

ZIKV. It is also effective against a PQS that is upstream of miRNAs (miR-K12-1–9,11) and an mRNA associated with latency associated nuclear antigen (LANA) of HSV1. It stabilized G4 structure of ZIKV and inhibited replication of the virus in vivo in dose-dependent manner with viral growth (Ruggiero et al. 2021).

PhenDC3 and bisquinolinium derivatives Bisquinolinium compounds are characterized by an aromatic nucleus substituted with two protonated quinoline moieties. PhenDC3 is the most prominent compound of bisquinolinium. PhenDC3 successfully prevented the NCL binding with G-quadruplex of EBNA mRNA of EBV virus (Ruggiero and Richter 2018). PhenDC3 is considered a prominent player for destabilizing G-quadruplex structures and inhibits viral growth (Reznichenko et al. 2019).

NMM (*N*-methyl mesoporphyrin IX) *N*-core methylated non-planar derivative of mesoporphyrin IX is the first identified PQS destabilizing agent, and it is used against PQS of several viruses, such as HCMV (Human Cyto Megalo Virus) and IAV (Influenza A Virus). In both cases, NMM destabilizes parallel G-quadruplex structures and inhibits viral growth (Ravichandran et al. 2018; Tomaszewska et al. 2021).

Naphthalene diamides These are most versatile G4 binders that are tested on HIV1, HSV1 and provide solid evidences of binding viral PQS over cellular PQS, thus, open up the possibilities as strong antiviral agent (Callegaro et al. 2017). These compounds are currently being tested for several viruses and projected to use as therapeutic agents against a wide range of viruses (Perrone et al. 2015).

Pyridostatin (PDS) and derivatives These PQS binding small molecules are tested against PQS of many viruses, such as HHVs, HPV, HCV, and ZIKV. It targeted PQS of ZIKV and inhibited protease activities suggesting its prominent role as antiviral agent (Zou et al. 2021).

PQS-specific antibody Specific G-quadruplex antibodies are developed in various methodologies and used to visualize PQS in vivo. Recently, their use to disrupt viral PQSs is emerging as an excellent antiviral agent by inhibiting viral gene functions (Abiri et al. 2021).

Anti-oligonucleotides against G-quadruplex Obviously, oligonucleotide-based therapy against specific or combined PQS sequences could be the best antiviral agent for its specificity. Researchers successfully used anti-oligo of PQSs for EBNA mRNA of EBV virus (Murat et al. 2014) and RNA aptamer against SARS-CoV1 helicase (Mashima et al. 2013). In both cases, viral growths are inhibited, and anti-oligotherapy against PQSs have the potential to consider as a strong antiviral therapy.

CRISPR-editing system Li et al. (2018) developed biosensor by disrupting the core G-quadruplex structure of SARS-CoV1 helicase to inhibit the ATP hydrolysis. PQS is mostly broken, and a significant reduction of helicase activity is observed in cells from infected patients.

Apart from these therapeutic ligands, other molecules that are inhibitory to PQSs, such as PQS interacting proteins, PQS degradation proteins, and CX5461 could be tested to break down the G-quadruplex structures to inhibit the viral replication, transcription, and ultimately viral growth. A series of benzoselenoxanthene is shown to destabilize PQS of TMPRSS2 in IAV-infected cells (Shen et al. 2020). As TMPRSS2 is also needed for SARS-CoV2 entry, these ligands could be tested to develop anti-SARS-CoV2 therapeutic agent.

Discussion

SARS-CoV2 causes covid-19 disease and created pandemic worldwide. Subsequent research effort was directed toward developing therapy to control the mortalities. Until date, except limited efficacy of various vaccines, no therapeutic target has been identified. Understanding its molecular mechanism of amplification inside host cells is a prerequisite to find target to develop successful and effective treatment of covid-19.

A stretch of DNA sequences in the genome can adopt abnormal structures except duplex DNA, such as triplex or G-quadruplex. These abnormal structures in DNA are shown to play critical role in transcription, chromatin rearrangement and gene regulation etc. The DNA sequence that can adopt G-quadruplex structures often forms LTR sequences that are abundant in RNA virus genome, such as in HIV. Moreover, HIV undergoes latency period after integrating into human genome, and the integration sites often constitute with LTR sequences. Analyzing RNA-seq data of nasopharyngeal cells of covid-19 patients, (Zhang et al. 2021) identified chimeric sequences carrying SARS-CoV2 viral DNA fragment and human genomic DNA at the integration junction suggesting that SARS-CoV2 may infrequently integrate into human genome. Whether these chimeric sequences contain SARS-CoV2 QGRS/PQS/LTR sequences are not studied. Moreover, they observed integration of SARS-CoV2 occurred in retrotransposons in human cells in culture systems. LTR retrotransposons comprise about 8% of the human genome (Ishak and De carvalho 2020). It remains to be investigated whether SARS-CoV2 integrates into human genome using its LTR sequences and goes to a latency period as observed for HIV (Craigie and Bushman 2012; Cohn et al. 2015).

Our search for G-quadruplex sequences in SARS-CoV2 genome yielded numerous existences of such potential sequences. They are mostly present in the upstream of SARS-CoV2 genes. There are ample evidences that QGRS/PQS sequences in RNA virus genome are involved in controlling host gene regulation and chromatin reorganization (Lin et al. 2001; Ohmori and Tsuruyama 2012; Ruggiero et al. 2019; Machida et al. 2020). However, the role of these QGRS/PQS sequences in SARS-COV2 genome is currently unknown although presumed to be playing similar role in controlling host gene regulation. Also, SARS-CoV2 faces constant challenges to survive in human cells and regulate host immunoregulatory genes for its own survival. The protein–protein interactome map identified numerous host proteins of several biological pathways that could be modulated by SARS-CoV2 (Gordon et al. 2020). The presence of potential G-quadruplex sequences in the upstream of NSP1, NSP2, NSP3, NSP4, RDRP, NSP10, ORF3a, ORF7A, and ORF10 may modulate the expression of these genes by regulating chromatin organization and can have impact in gene regulation. Exploring such roles by designing future experiments with PQSs/QGRSs could be extremely useful to understand whether these sequences modulate the expression of these SARS-CoV2 immunoregulatory genes and help the viral amplification for its own survival in human cells.

Lastly, we discussed several PQS/QGRS destabilizing compounds that are known to stall replication forks, inhibit transcription and viral growth in both DNA and RNA virus. These strategies especially the anti-oligonucleotides specific to each PQS or CRISPR-editing of PQS could be useful to develop therapeutic agents to control SARS-CoV2 amplification.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00251-022-01257-6.

Acknowledgements I am indebted to Dr. S. Gupta for his critical reading of this manuscript.

Data availability All data in this manuscript will be available publicly.

Declarations

Conflict of interest The author declares no competing interests.

References

Abiri A, Lavigne M, Rezaei M, Nikzad S, Zare P, Mergny JL, Rahimi HR (2021) Unlocking G- quadruplexes as antiviral targets. Pharmacol Rev 73(3):897–923

Aftab SO, Ghouri MZ, Masood MU, Haider Z, Khan Z, Ahmad A, Munawar N (2020) Analysis of SARS-CoV-2 RNA-dependent RNA polymerase as a potential therapeutic drug target using a computational approach. J Transl Med 18(1):275

Angeletti S, Benvenuto D, Bianchi M, Giovanetti M, Pascarella S, Ciccozzi M (2020) COVID-2019: The role of the nsp2 and nsp3 in its pathogenesis. J Med Virol 92(6):584–588

Bouvet M, Lugari A, Posthuma CC, Zevenhoven JC, Bernard S, Betzi S, Imbert I, Canard B, Guillemot JC, Lécine P, Pfefferle S, Drosten C, Snijder EJ, Decroly E, Morelli X (2014) Coronavirus Nsp10, a critical co-factor for activation of multiple replicative enzymes. J Biol Chem 289(37):25783–25796

Butovskaya E, Heddi B, Bakalar B, Richter SN, 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(42):13654–13662

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

Castaño-Rodriguez C, Honrubia JM, Gutiérrez-Álvarez J, DeDiego ML, Nieto-Torres JL, Jimenez-Guardeño JM, Regla-Nava JA, Fernandez-Delgado R, Verdia-Báguena C, Queralt-Martín M, Kochan G (2018) Role of severe acute respiratory syndrome coronavirus viroporins E, 3a, and 8a in replication and pathogenesis. mBio 9(3):e02325-17

Chatterjee S, Zagelbaum J, Savitsky P, Sturzenegger A, Huttner D, Janscak P, Hickson ID, Gileadi O, Rothenberg E (2014) Mechanistic insight into the interaction of BLM helicase with intra-strand G-quadruplex structures. Nat Commun 5:5556

Cohn LB, Silva IT, Oliveira TY, Rosales RA, Parrish EH, Learn GH, Hahn BH, Czartoski JL, McElrath MJ, Lehmann C, Klein F, Caskey M, Walker BD, Siliciano JD, Siliciano RF, Jankovic M, Nussenzweig MC (2015) HIV-1 integration landscape during latent and active infection. Cell 160(3):420–432

Craigie R, Bushman FD (2012) "HIV DNA integration." Cold Spring Harb Perspect Med 2(7):a006890

Cui H, Zhang L (2020) "G-Quadruplexes are present in human coronaviruses including SARS-CoV-2." Front Microbiol 11:567317

Davies JP, Almasy KM, McDonald EF, Plate L (2020) "Comparative multiplexed interactomics of SARS-CoV-2 and homologous coronavirus non-structural proteins identifies unique and shared host-cell dependencies." ACS Infect Dis 6(12): 3174–3189.

Deitsch KW, Moxon ER, Wellems TE (1997) Shared themes of antigenic variation and virulence in bacterial, protozoal, and fungal infections. Microbiol Mol Biol Rev 61(3):281–293

Fehr AR, Channappanavar R, Jankevicius G, Fett C, Zhao J, Athmer J, Meyerholz DK, Ahel I, Perlman S (2016) "The conserved coronavirus macrodomain promotes virulence and suppresses the innate immune response during severe acute respiratory syndrome coronavirus infection." mBio **7**(6):e01721-16

Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, White KM, O'Meara MJ, Rezelj VV, Guo JZ, Swaney DL, Tummino TA, Hüttenhain R, Kaake RM, Richards AL, Tutuncuoglu B, Foussard H, Batra J, Haas K, Modak M, Kim M, Haas P, Polacco BJ, Braberg H, Fabius JM, Eckhardt M, Soucheray M, Bennett MJ, Cakir M, McGregor MJ, Li Q, Meyer B, Roesch F, Vallet T, Mac Kain A, Miorin L, Moreno E, Naing ZZC, Zhou Y, Peng S, Shi Y, Zhang Z, Shen W, Kirby IT, Melnyk JE, Chorba JS, Lou K, Dai SA, Barrio-Hernandez I, Memon D, Hernandez-Armenta C, Lyu J, Mathy CJP, Perica T, Pilla KB, Ganesan SJ, Saltzberg DJ, Rakesh R, Liu X, Rosenthal SB, Calviello L, Venkataramanan S, Liboy-Lugo J, Lin Y, Huang XP, Liu Y, Wankowicz SA, Bohn M, Safari M, Ugur FS, Koh C, Savar NS, Tran QD, Shengjuler D, Fletcher SJ, O'Neal MC, Cai Y, Chang JCJ, Broadhurst DJ, Klippsten S, Sharp PP, Wenzell NA, Kuzuoglu-Ozturk D, Wang HY, Trenker R, Young JM, Cavero DA, Hiatt J, Roth TL, Rathore U, Subramanian A, Noack J, Hubert M, Stroud RM, Frankel AD, Rosenberg OS, Verba KA, Agard DA, Ott M, Emerman M, Jura N, von Zastrow M, Verdin E, Ashworth A, Schwartz O, d'Enfert C, Mukherjee S, Jacobson M, Malik HS, Fujimori DG, Ideker T, Craik CS, Floor SN, Fraser JS, Gross JD, Sali A, Roth BL,

Ruggero D, Taunton J, Kortemme T, Beltrao P, Vignuzzi M, García-Sastre A, Shokat KM, Shoichet BK, Krogan NJ (2020) A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature 583(7816):459–468

Harris LM, Merrick CJ (2015) "G-quadruplexes in pathogens: a common route to virulence control?" PLoS Pathog 11(2):e1004562

Hassan SS, Attrish D, Ghosh S, Choudhury PP, Uversky VN, Aljabali AAA, Lundstrom K, Uhal BD, Rezaei N, Seyran M, Pizzol D, Adadi P, Soares A, Abd El-Aziz TM, Kandimalla R, Tambuwala MM, Azad GK, Sherchan SP, Baetas-da-Cruz W, Lal A, Palù G, Takayama K, Serrano-Aroca Á, Barh D, Brufsky AM (2021) Notable sequence homology of the ORF10 protein introspects the architecture of SARS-CoV-2. Int J Biol Macromol 181:801–809

Hillen HS, Kokic G, Farnung L, Dienemann C, Tegunov D, Cramer P (2020) Structure of replicating SARS-CoV-2 polymerase. Nature 584(7819):154–156

Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, Schiergens TS, Herrler G, Wu NH, NitscheA, Müller MA, Drosten C, Pöhlmann S (2020) "SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor." Cell 181(2):271-280.e8

Hurst KR, Koetzner CA, Masters PS (2013) Characterization of a critical interaction between the coronavirus nucleocapsid protein and nonstructural protein 3 of the viral replicase-transcriptase complex. J Virol 87(16):9159–9172

Ishak CA, De carvalho DD (2020) Reactivation of endogenous retroelements in cancer development and therapy. Annual Review of Cancer Biology 4:159–176

Issa E, Merhi G, Panossian B, Salloum T, Tokajian S (2020) "SARS-CoV-2 and ORF3a: nonsynonymous mutations, functional domains, and viral pathogenesis." mSystems 5(3):e00266-20

Kaur M, Sharma A, Kumar S, Singh G, Barnwal RP (2021) SARS-CoV-2: insights into its structural intricacies and functional aspects for drug and vaccine development. Int J Biol Macromol 179:45–60

Kikin O, D'Antonio L, Bagga PS (2006) "QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences." Nucleic Acids Res 34(Web Server issue):W676–682

Krafčíková P, Demkovičová E, Halaganová A, Víglaský V (2017) Putative HIV and SIV G-quadruplex sequences in coding and noncoding regions can form G-quadruplexes. J Nucleic Acids 2017:6513720

Kusov Y, Tan J, Alvarez E, Enjuanes L, 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

Lei J, Kusov Y, Hilgenfeld R (2018) Nsp3 of coronaviruses: structures and functions of a large multi-domain protein. Antiviral Res 149:58–74

Li SY, Cheng QX, Liu JK, Nie XQ, Zhao GP, Wang J (2018) CRISPR-Cas12a has both cis- and trans-cleavage activities on single-stranded DNA. Cell Res 28(4):491–493

Lin S, Chen H, Chen Z, Yang F, Ye F, Zheng Y, Yang J, Lin X, Sun H, Wang L, Wen A, Dong H, Xiao Q, Deng D, Cao Y, Lu G (2021) Crystal structure of SARS-CoV-2 nsp10 bound to nsp14-ExoN domain reveals an exoribonuclease with both structural and functional integrity. Nucleic Acids Res 49(9):5382–5392

Lin YC, Shih JW, Hsu CL, Lin JJ (2001) Binding and partial denaturing of G-quartet DNA by Cdc13p of Saccharomyces cerevisiae. J Biol Chem 276(50):47671–47674

Machida S, Depierre D, Chen HC, Thenin-Houssier S, Petitjean G, Doyen CM, Takaku M, Cuvier O, Benkirane M (2020) Exploring histone loading on HIV DNA reveals a dynamic nucleosome positioning between unintegrated and integrated viral genome. Proc Natl Acad Sci U S A 117(12):6822–6830

Mashima T, Nishikawa F, Kamatari YO, Fujiwara H, Saimura M, Nagata T, Kodaki T, Nishikawa S, Kuwata K, Katahira M (2013) Anti-prion activity of an RNA aptamer and its structural basis. Nucleic Acids Res 41(2):1355–1362

Mills M, Lacroix L, Arimondo PB, Leroy JL, François JC, Klump H, Mergny JL (2002) Unusual DNA conformations: implications for telomeres. Curr Med Chem Anticancer Agents 2(5):627–644

Murat P, Zhong J, Lekieffre L, Cowieson NP, Clancy JL, Preiss T, Balasubramanian S, Khanna R, Tellam J (2014) G-quadruplexes regulate Epstein-Barr virus-encoded nuclear antigen 1 mRNA translation. Nat Chem Biol 10(5):358–364

Ohmori R, Tsuruyama T (2012) "In vitro HIV-1 LTR integration into T-cell activation gene CD27 segment and the decoy effect of modified-sequence DNA." PLoS One 7(11):e49960

Perrone R, Doria F, Butovskaya E, Frasson I, Botti S, Scalabrin M, Lago S, Grande V, Nadai M, Freccero M, Richter SN (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(24):9639–9652

Perrone R, Lavezzo E, Riello E, Manganelli R, Palù G, Toppo S, Provvedi R, Richter SN (2017) Mapping and characterization of G-quadruplexes in Mycobacterium tuberculosis gene promoter regions. Sci Rep 7(1):5743

Poltronieri P, Sun B, Mallardo M (2015) RNA viruses: RNA roles in pathogenesis, coreplication and viral load. Curr Genomics 16(5):327–335

Ravichandran S, Kim YE, Bansal V, Ghosh A, Hur J, Subramani VK, Pradhan S, Lee MK, Kim KK, Ahn JH (2018) "Genome-wide analysis of regulatory G-quadruplexes affecting gene expression in human cytomegalovirus." PLoS Pathog 14(9):e1007334

Ren LL, Wang YM, Wu ZQ, Xiang ZC, Guo L, Xu T, Jiang YZ, Xiong Y, Li YJ, Li XW, Li H, Fan GH, Gu XY, Xiao Y, Gao H, Xu JY, Yang F, Wang XM, Wu C, Chen L, Liu YW, Liu B, Yang J, Wang XR, Dong J, Li L, Huang CL, Zhao JP, Hu Y, Cheng ZS, Liu LL, Qian ZH, Qin C, Jin Q, Cao B, Wang JW (2020a) "Identification of a novel coronavirus causing severe pneumonia in human: a descriptive study." Chin Med J (Engl) 133(9):1015-1024

Ren Y, Shu T, Wu D, Mu J, Wang C, Huang M, Han Y, Zhang XY, Zhou W, Qiu Y, Zhou X (2020b) The ORF3a protein of SARS-CoV-2 induces apoptosis in cells. Cell Mol Immunol 17(8):881–883

Reznichenko O, Quillévéré A, Martins RP, Loaëc N, Kang H, Lista MJ, Beauvineau C, González-García J, Guillot R, Voisset C, Daskalogianni C, Fåhraeus R, Teulade-Fichou MP, Blondel M, Granzhan A (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

Ruggiero E, Richter SN (2018) G-quadruplexes and G-quadruplex ligands: targets and tools in antiviral therapy. Nucleic Acids Res 46(7):3270–3283

Ruggiero E, Tassinari M, Perrone R, Nadai M, Richter SN (2019) Stable and conserved G-quadruplexes in the long terminal repeat promoter of retroviruses. ACS Infect Dis 5(7):1150–1159

Ruggiero E, Zanin I, Terreri M, Richter SN (2021) "G-quadruplex targeting in the fight against viruses: an update." Int J Mol Sci 22(20):10984

Sakai Y, Kawachi K, Terada Y, Omori H, Matsuura Y, Kamitani W (2017) Two-amino acids change in the nsp4 of SARS coronavirus abolishes viral replication. Virology 510:165–174

Saranathan N, Vivekanandan P (2019) G-quadruplexes: more than just a kink in microbial genomes. Trends Microbiol 27(2):148–163

Schubert K, Karousis ED, Jomaa A, Scaiola A, Echeverria B, Gurzeler LA, Leibundgut M, Thiel V, Mühlemann O, Ban N (2020) SARS-CoV-2 Nsp1 binds the ribosomal mRNA channel to inhibit translation. Nat Struct Mol Biol 27(10):959–966

Shen LW, Qian MQ, Yu K, Narva S, Yu F, Wu YL, Zhang W (2020) Inhibition of influenza A virus propagation by benzoselenoxanthenes stabilizing TMPRSS2 Gene G-quadruplex and hence downregulating TMPRSS2 expression. Sci Rep 10(1):7635

Thoms M, Buschauer R, Ameismeier M, Koepke L, Denk T, Hirschenberger M, Kratzat H, Hayn M, Mackens-Kiani T, Cheng J, Straub JH, Stürzel CM, Fröhlich T, Berninghausen O, Becker T, Kirchhoff F, Sparrer KMJ, Beckmann R (2020) Structural basis for translational shutdown and immune evasion by the Nsp1 protein of SARS-CoV-2. Science 369(6508):1249–1255

Tomaszewska M, Szabat M, Zielińska K, Kierzek R (2021) "Identification and structural aspects of G-quadruplex-forming sequences from the influenza A virus genome." Int J Mol Sci 22(11):6031

Wessner DL (2010) The Origins of Viruses. Nature. Education 3:37

Wojdyla JA, Manolaridis I, van Kasteren PB, Kikkert M, Snijder EJ, Gorbalenya AE, Tucker PA (2010) Papain-like protease 1 from transmissible gastroenteritis virus: crystal structure and enzymatic activity toward viral and cellular substrates. J Virol 84(19):10063–10073

Xu J, Huang H, Zhou X (2021) G-quadruplexes in neurobiology and virology: functional roles and potential therapeutic approaches. JACS Au 1(12):2146–2161

Zhang JY, Xia Y, Hao YH, Tan Z (2020a) DNA:RNA hybrid G-quadruplex formation upstream of transcription start site. Sci Rep 10(1):7429

Zhang L, Richards A, Barrasa MI, Hughes SH, Young RA, Jaenisch R (2021) "Reverse-transcribed SARS-CoV-2 RNA can integrate into the genome of cultured human cells and can be expressed in patient-derived tissues." Proc Natl Acad Sci USA 118(21):e2105968118

Zhang R, Xiao K, Gu Y, Liu H, Sun X (2020b) "Whole genome identification of potential G- quadruplexes and analysis of the G-quadruplex binding domain for SARS-CoV-2." Front Genet 11:587829

Zheng YX, Wang L, Kong WS, Chen H, Wang XN, Meng Q, Zhang HN, Guo SJ, Jiang HW, Tao SC (2021) Nsp2 has the potential to be a drug target revealed by global identification of SARS-CoV-2 Nsp2-interacting proteins. Acta Biochim Biophys Sin (shanghai) 53(9):1134–1141

Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, Guo H, Jiang RD, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao GF, Shi ZL (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579(7798):270–273

Zou M, Li JY, Zhang MJ, Li JH, Huang JT, You PD, Liu SW, Zhou CQ (2021) G-quadruplex binder pyridostatin as an effective multi-target ZIKV inhibitor. Int J Biol Macromol 190:178–188