G-Quadruplexes in Pathogens: A Common Route to Virulence Control?

Lynne M. Harris, Catherine J. Merrick*

Centre for Applied Entomology and Parasitology, School of Life Sciences, Keele University, Keele, Staffordshire, United Kingdom

* c.merrick@keele.ac.uk

Abstract

DNA can form several secondary structures besides the classic double helix: one that has received much attention in recent years is the G-quadruplex (G4). This is a stable four-stranded structure formed by the stacking of quartets of guanine bases. Recent work has convincingly shown that G4s can form in vivo as well as in vitro and can affect both replication and transcription of DNA. They also play important roles at G-rich telomeres. Now, a spate of exciting reports has begun to reveal roles for G4 structures in virulence processes in several important microbial pathogens of humans. Interestingly, these come from a range of kingdoms—bacteria and protozoa as well as viruses—and all facilitate immune evasion in different ways. In particular, roles for G4s have been posited in the antigenic variation systems of bacteria and protozoa, as well as in the silencing of at least two major human viruses, human immunodeficiency virus (HIV) and Epstein-Barr virus (EBV). Although antigenic variation and the silencing of latent viruses are quite distinct from one another, both are routes to immune evasion and the maintenance of chronic infections. Thus, highly disparate pathogens can use G4 motifs to control DNA/RNA dynamics in ways that are relevant to common virulence phenotypes. This review explores the evidence for G4 biology in such processes across a range of important human pathogens.

What Are G-Quadruplexes and Why Are They Important?

Over a hundred years ago it was reported that concentrated guanylic acid can self-assemble [1], but it was not until the 1960s that the structural basis for this phenomenon, the G4, was elucidated [2]. G4s were initially considered a structural curiosity; however, it has since become clear that they are involved in a number of key biological functions. This has led to the emergence of G4s as a hot topic in nucleic acids research with the vast majority of this research thus far undertaken in highly tractable model systems such as *Saccharomyces cerevisiae* or human cell lines. However, there is now a rapidly developing literature on the roles of G4s in human pathogens. This review will briefly outline the known roles for G4s in the cell biology of model systems, and then explore how these can map onto pathogen biology, particularly to facilitate immune evasion.

In terms of structure, the basic unit of the G4 is the G-tetrad, a planar array of four Hoogsteen-bonded guanine bases, which when stabilised by monovalent cations, stack on top of one another to form the G4 structure itself (Fig. 1). The stacked G-tetrads are connected by loops
formed from intervening mixed-sequence nucleotides: these loops vary in both size and sequence from one G4 to another. The four strands comprising the G4 may originate from one, two, or four separate strands of DNA or RNA. G4s can therefore be described as either intramolecular or intermolecular (Fig. 1C, D). In addition, there is directionality to the strands, which can be described as running from the 5' end to the 3' end. G4s can therefore exist as a number of topological variants (Fig. 1C, D). The conformation of glycosidic bonds of guanine bases in G-tetrads, the cations present and the number of stacked G-tetrads further contribute to the myriad of topologies found amongst G4s [3–6].

Predictive algorithms, such as G4P Calculator [7] and QuadParser [8], have been developed to identify putative quadruplex sequences (PQS) within nucleic acid sequences. Use of these algorithms in whole genome sequences has revealed that PQS are not randomly located throughout genomes but are overrepresented in gene regulatory regions and repetitive regions such as telomeres [9,10]. RNA G4s are present in transcripts associated with telomeres, in noncoding regions of primary transcripts, and also in mature transcripts.
The regions in which PQS occur are linked to the specific functions of G4s at these locations. For example, at telomeres, G4 structures are proposed to be involved in telomere maintenance at both the RNA and DNA level. Eukaryotic telomeric DNA consists of long stretches of tandemly repeated G-rich sequences, such as GGGTTA in humans, which end in a 3’ single-stranded DNA overhang. A protein complex caps these overhangs in order to prevent them being identified by cellular surveillance mechanisms as unwanted DNA breaks. These G-rich telomeric repeats can form G4s both in vitro and in vivo and can protect telomeres: in *S. cerevisiae* telomeric G4s may provide an alternative form of telomere capping when natural capping is compromised [11]. In addition, telomeric G4s protect the telomeric 3’-overhang from being recognised by telomerase, thereby regulating telomerase activity. Human telomeres are transcribed to produce long, noncoding telomeric repeat-containing RNAs (TERRAs), which consist of UUAGGG repeats and adopt a G4 RNA structure [12,13]. TERRAs interact with the telomere binding protein TRF2 to promote telomere heterochromatinisation [14,15].

In promoter regions, the dynamic behaviour of G4s may be directly involved in gene regulation at the level of transcription. One of the best studied systems for a role of G4s in transcriptional regulation is the human *c-MYC* locus. *c-MYC* is a transcription factor whose expression is linked to cell proliferation and tumourigenesis. Nuclease hypersensitive element III (NHE III), a major regulator of *c-MYC* transcription, contains a PQS that forms a G4 structure in vitro [16]. A gene containing wild-type NHE III is less expressed than one containing a mutated version that cannot form a G4 structure, so the PQS in NHE III represses transcription [17]. Additionally TMPyP4, a G4-stabilising ligand, reduces *c-MYC* expression in lymphoma cell lines [17].

Where G4s occur in gene bodies, they can form steric roadblocks to the DNA transcriptional machinery, and certain helicases, such as those of the RecQ and Pif1 families, possess G4-resolving activity to aid transcription through these four-stranded snags. Disruption of G4-resolving helicases in *Caenorhabditis elegans* [18], *S. cerevisiae* [19] or human cells lines [20] results in genetic instability, and G4s are a hallmark of fragile sites in the human genome [21]. In RNA transcripts, G4s may play roles in pre-mRNA processing, translation and RNA turnover.

Given their propensity to induce genomic instability and DNA damage, it is hypothesised that G4s could be used alongside inhibitors of DNA repair or associated pathways to inhibit tumourigenesis. For example, the G4-interactive compound Quarfloxin (CX-3543, Cylene Pharmaceuticals) is a first-in-class drug that progressed to Phase II clinical trials for cancer. Quarfloxin selectively disrupts the interaction of ribosomal DNA G4s with nucleolin, thereby inhibiting RNA polymerase I and inducing apoptosis in cancer cells [22]. Furthermore, stabilisation of G4s may be an effective approach by which to inhibit telomerase activity in tumour cells [23]. The field of G4 biology, so far largely motivated by this potential for developing novel anti-cancer therapeutics, has provided us with both a conceptual basis and a molecular toolset (Box 1) with which to approach the burgeoning field of pathogen G4 biology.

### Roles for G4s in Recombination-Mediated Antigenic Variation

Antigenic variation (Av) is the process by which pathogens express different versions of their surface epitopes in order to evade detection by the host immune system. In general, the pathogen possesses a bank of genes encoding possible antigenic variants but expresses only one of these genes at any time. Switching can be mediated by a variety of sophisticated genetic and epigenetic systems, but in several well-characterised pathogens—including *Neisseria gonorrhoeae*, the bacterium responsible for gonorrhoea, and *Trypanosoma brucei*, the causative
Box 1. The G4 Biology Toolbox

Many tools have been developed to analyse G4s at both genome-wide and sequence-specific levels.

Predictive algorithms

Intramolecular G4s form from short runs of guanine bases separated by short runs of other bases. Intermolecular G4 structures form from G-runs on two or more nucleic acid strands. Therefore, by searching for sequences matching these criteria, we can identify genomic regions that may form G4s, termed putative quadruplex sequences (PQS) or G4 motifs. Predictive algorithms (reviewed in [70]), including quadparser [8], G4P calculator [7], QGRS mapper [71] and QuadBase [72], have been developed to identify PQS. These algorithms can rapidly analyse large amounts of data, including whole genome sequences.

Biophysical techniques

Several experimental techniques can determine whether the biophysical features of a synthetic PQS-containing oligonucleotide are consistent with G4 structure formation. These include dimethylsulfate footprinting, thermal denaturation, ultraviolet spectroscopy and circular dichroism spectroscopy. Complete structure determination necessitates the use of X-ray crystallography or nuclear magnetic resonance structure determination.

G4-interactive compounds

Diverse families of small molecules preferentially bind to G4 DNA over other types of DNA. These include the fluoroquinolone quarfloxin [22], the acridine derivative BRACO19 [73], telomestatin [74], and triazines [75], among many others (reviewed in [4]). Compounds that stabilise and/or induce G4 formation can be used to investigate the biological roles of G4s. In addition, the RNA G4-selective ligand carboxyPDS has recently been described [69].

G4 structure-specific antibodies

G4-structure specific antibodies have been exploited to visualise G4s in both genomic DNA and cytoplasmic RNA. High-affinity single-chain antibodies, generated by ribosome display, have been used to visualise G4s at the telomeres of the ciliate Stylonychia lemnae [76,77]. More recently, a monoclonal single-chain antibody generated by phage display has been used to visualise G4s in the genomic DNA and cytoplasmic RNA of a range of mammalian cells [68,69].

G4-interacting proteins

Many proteins have been identified which interact physically and/or functionally with G4s in a variety of organisms. Some G4-interacting proteins, such as the β subunit of the Oxytricha nova telomere-binding protein, promote the formation of G4s from PQS [78,79]. The binding of some proteins to extant G4s increases the stability of the G4 structure [80,81]. Conversely, a group of G4 destabilising proteins has been discovered. These include the highly conserved human telomeric protein POT1 [82] and a number of single-strand binding proteins of the heterogenous nuclear ribonucleoprotein (hnRNP) family [83–85]. In addition a number of helicases preferentially bind and disrupt DNA G4s. Experimentally modulating the expression or stability of these proteins offers a means by which to perturb G4 stability.
agent of sleeping sickness—it is genetic recombination into a single gene expression site that mediates antigenic switching.

G4s have now been implicated in Av in several pathogens, the best characterized example being *N. gonorrhoeae*, which uses a G4-mediated system to switch the expression of its cell-surface pilin proteins (Fig. 2). During pilin antigenic variation, only the pilin gene residing at the active *pilE* locus is expressed and the resident gene is frequently replaced with a gene from a pool of silent *pilS* loci. The recombination initiation site has been mapped to a 16 base pair (bp) G-rich segment that forms a parallel intramolecular G4 structure in vitro and is located upstream of the *pilE* locus [24]. Disruption of this structure through site-directed mutagenesis prevented the formation of single-stranded nicks that can initiate recombination, and thus suppressed pilin Av [24]. Furthermore, recombinant RecA specifically bound to the *pilE* G4 to stimulate strand exchange in vitro, suggesting that the G4 structure may recruit recombination factors to the *pilE* locus [25]. Finally, mutation of the *N. gonorrhoeae* RecQ helicase (a structure-specific helicase which can unwind the *pilE* G4 structure in vitro) resulted in defective pilin Av [26].

---

**Fig 2. Schematic model of the role of the *pilE* G-quadruplex (G4) in *N. gonorrhoeae* pilin antigenic variation.** A putative quadruplex sequence (PQS) and a small RNA (sRNA) promoter are located upstream of the *pilE* locus (A). The initiation of transcription from the sRNA promoter (B) provides the single-stranded conditions required for G4 formation (C). The *pilE* G4 recruits RecA (D) and potentially other recombination factors, which stimulates non-reciprocal recombination between a *pilS* locus and the *pilE* locus (E).

doi:10.1371/journal.ppat.1004562.g002
A long-standing question in the field of G4 biology has been whether these structures actually form and persist in vivo from the inherently stable DNA double helix. For this to occur, G4 structure formation must be preceded or accompanied by localised unwinding of the double helix. Active transcription results in the spread of negative superhelicity behind the transcriptional machinery [27,28], and this can provide the torsional stress required for unwinding of the double helix and the formation of G4s [29]. At the pilE locus, active transcription of a small noncoding RNA (sRNA), which initiates within the pilE G4-forming sequence, is required for pilin Aβ (Fig. 2B). The expression of this sRNA in trans cannot complement a loss-of-function mutant in the sRNA promoter, implying that active in situ transcription may be providing the torsional stress required for helix unwinding and G4-structure formation [30].

It is interesting to speculate on whether this well-characterised system for promoting recombinational Aβ can be generalized to other pathogens. A recent analysis of >600 Neisseria meningitidis isolates found PQS adjacent to some, but not all, pilE gene sequences; specifically, they were found 5' or 3' of class I, but not class II, pilE loci. [31]. Given that the former, and not the latter, are antigenically unstable, we speculate that G4s may also play a role in N. meningitidis class I pilin Aβ.

An entirely different genus of pathogenic bacteria also evades the host immune response through Aβ: these are the spirochetes Borrelia spp., some of which cause human Lyme disease. Lyme spirochetes evade the host immune response through gene conversion-driven Aβ of a surface lipoprotein, Vmp-like sequence E (VlsE). Runs of guanines capable of forming intermolecular G4 structures in vitro are abundant on the coding strand of the vls locus in several Lyme Borrelia strains and species [32]. It is not yet known whether G4s actually form in vivo at this locus to play an active role in promoting gene conversion. In vitro, however, Borrelia telomere resolvase does appear to disrupt intermolecular G4s formed from this locus [33]. This enzyme has single-strand annealing and strand exchange activities, both of which have been hypothesised to play roles in gene conversion at the vlsE locus.

Further understanding of the initiating events of antigenic switching in this system will help to determine what, if any, role G4s play in Lyme Borrelia Aβ and recombinational switching. More generally, genome-wide scans for the distribution of PQS within bacterial genomes suggest that G4s are widely distributed and may actually have quite broad roles in the regulation of bacterial genes. G4s may therefore affect virulence processes in many more bacterial pathogens, and tractable bacterial species that contain G4s, like Escherichia coli, could prove useful models for future studies [34,35]. Finally, this model may also apply in eukaryotic as well as prokaryotic microbes, since there is an unexplored role for G4s in the antigenic switching of variant surface glycoproteins in T. brucei, where G-rich telomeres—potential G4-forming sites—facilitate the switching of VSG virulence genes [36].

**Other Roles for G4s in Antigenic Variation**

It is clear that G4 motifs have key roles in certain recombination-mediated Aβ systems. There are, however, other systems of Aβ that are mechanistically distinct, such as that mediated by epigenetic silencing in the fungal pathogen Candida glabrata [37] and the protozoan malaria parasite Plasmodium falciparum [38]. Could G4s play a role here too? Although this area remains underexplored, bioinformatic work on P. falciparum suggests that indeed they could. A search of the P. falciparum genome for G4 motifs revealed that outside of the telomeres, PQS are rare in this AT-rich genome [39]. Interestingly, half of the non-telomeric PQS are located in or upstream of a large, multicopy, hypervariable gene family called var [39]. Var genes encode P. falciparum erythrocyte membrane protein 1 (PiEMP1), a family of variant immunodominant surface antigens, which are important virulence factors [40–42]. Var gene expression is
mutually exclusive and *P. falciparum* frequently switches to express different *var* genes. Furthermore, *var* genes regularly recombine to generate new variants. Transcriptional switching is mediated by epigenetic silencing of all but one *var* locus and switching occurs when the current active site becomes heterochromatic while another site becomes euchromatic and actively transcribed—the trigger for this is still unknown [43].

The predominance of G4 motifs in *var* gene regulatory regions suggests that they may play roles in *var* gene recombination and/or switching. G4 motifs upstream of well-characterised genes in model systems can alter levels of transcription, e.g., in the *c-MYC* oncogene [17], and changes to G4 metabolism can also induce switches in epigenetic silencing [44–46]. *Plasmodium* possesses at least some of the enzymes known to metabolise G4s: *P. falciparum* contains two putative RecQ helicases [47], although there is no recognizable Pif1. It is also notable that most of the *var* family (like several other variantly-expressed gene families in this parasite) is located just inside the telomeres, raising the possibility that telomeric G4s may play roles in both telomere maintenance and the regulation of subtelomeric virulence gene families. Synthetic oligonucleotides composed of the degenerate *P. falciparum* telomeric motif GGGTTYA are indeed able to form stable G4 structures in vitro [48], although their biological significance awaits confirmation in vivo.

**Could G4s Play a Role in Viral Latency?**

Recent research suggests that viruses may use G4s as *cis*-acting regulatory elements in gene expression. The retroviral HIV-1 genome contains two copies of single-stranded RNA encoding nine genes. Following infection, the RNA is transcribed by reverse transcriptase into double-stranded DNA, which is integrated into the infected cell genome. The HIV-1 long terminal repeat (LTR) controls viral transcription within the integrated provirus and contains two intramolecular G4s within a 57 bp G-rich tract of its U3 promoter region [49]. This tract contains five cellular transcription factor binding sites: two NF-κB and three Sp1 (Fig. 3A). When the wild-type LTR sequence was placed upstream of a luciferase reporter, G4-disrupting mutations or the addition of a G4-stabilising ligand resulted in increased and decreased promoter activity, respectively, suggesting that G4s may regulate HIV-1 LTR promoter activity [49].

**Fig 3. Schematic representation of the mechanisms by which G-quadruplexes (G4) may contribute to the maintenance of viral latency.** (A) In the integrated HIV-1 provirus, the U3 region contains two NF-κB binding sites (black blocks) and three Sp1 binding sites (grey blocks). Here, the formation of a G4 from the G-rich Sp1 binding sites is shown, but several distinct G4s can form in this region. G4 formation inhibits transcription from the transcription start site (TSS). This transcriptional repression may be mediated by the differential affinity of transcription factors for double-stranded and G4 DNA. (B) The glycine-alanine repeat domain of Epstein-Barr virus-encoded nuclear antigen 1 (EBNA1) mRNA contains PQS. G4 formation inhibits translation, thereby reducing EBNA1 protein levels and limiting the presentation of EBNA1 peptides by antigen presenting cells (APC) via the major histocompatibility complex (MHC) pathway.

doi:10.1371/journal.ppat.1004562.g003
There is further evidence that the U3 promoter region may be capable of dynamically adopting various G4 structures with the potential to bind cellular transcription factors. In addition to the two parallel-like intramolecular G4s described above, Amrane and colleagues found that a synthetic sequence derived from the same promoter region could form a hybrid G4 structure consisting of a two-G-tetrad antiparallel G4 with a further Watson-Crick CG bp [50]. Furthermore, Piekna-Przybylska et al. found that synthetic oligonucleotides comprising G-runs from the three Sp1 binding sites could adopt a variety of configurations in vitro, including parallel, antiparallel and hybrid G4 conformations [51]. Pull-down assays indicated that Sp1 could bind to one of the Sp1 sites folded into a G4 structure [51]. Given that Sp1 activity is associated with the maintenance of viral latency [52,53] and G4 structures within the U3 promoter region can suppress transcriptional activity [49], this potential for alternate protein-DNA interactions may play a role in HIV-1 latency [51].

G4s may also play a role in the transcriptional regulation of the HIV-1 provirus itself. Three adjacent, highly conserved PQS, capable of forming stable G4 structures in vitro, are located within the proviral nef gene. When the nef gene was cloned upstream of the reporter gene green fluorescent protein (GFP), the addition of a G4-stabilising drug, but not of its non-G4 binding analogue, reduced GFP expression levels [54]. This G4-mediated transcriptional regulation may have important implications for viral pathogenicity as the HIV-1 Nef protein is an important virulence factor [55]. Finally, stable G4 structures can form in regions of the HIV-1 RNA genome [51,56], where they may be involved in regulating reverse transcription of the genome.

Moving from HIV-1 to EBV, very recent evidence suggests that G4s may also play a role in the latency program of this gammaherpesvirus. Here, however, the G4 acts at the level of RNA translation rather than DNA transcription. Following infection, down-regulation of viral protein synthesis relies heavily upon a class of viral proteins that can inhibit their own synthesis, termed “genome maintenance proteins.” The mRNA for one of these genome maintenance proteins, EBV-encoded nuclear antigen 1 (EBNA1), contains G4 structures within its glycine-alanine repeat domain (GAr)-encoding region. These may regulate its translation, since disruption of these G4 structures reduces ribosome dissociation and increases EBNA1 mRNA translation (Fig. 3B) [57]. Furthermore, destabilisation of these G4 structures, following transfection of a vector expressing native GAr mRNA fused to a sequence encoding an ovalbumin epitope, resulted in enhanced antigen presentation of that epitope by a T-cell hybridoma [57]. These results have been mimicked with the EBNA1 antigen itself in an in vivo mouse model, where EBNA1 mRNA G4 destabilisation resulted in enhanced antigen presentation by dendritic cells and the early priming of CD8+ T cells [58], demonstrating that reducing the translation efficiency of viral proteins is a key part of the latency program of this gammaherpesviruses.

Finally, G-rich regions from both control and coding regions of some human papillomavirus (HPV) genomes are able to form G4 structures and may also carry out regulatory functions in this virus [59]. The roles of G4s in the silencing of at least two major human viruses provide them with a flexible tool with which to evade immune detection and thereby maintain chronic infections. Thus, although the mechanisms are entirely distinct, the results in terms of virulence are comparable to those of the Av systems described above.

Roles for G4s in Antigenic Diversification

G4 motifs have been proposed to play roles not only in antigenic switching but also in the generation of antigenic diversity. So what are the molecular mechanisms by which these structures might induce recombination or mutation, and thus diversification of PQS-containing antigen-encoding genes? Recent research, using S. cerevisiae as a tractable tool with which to investigate
G4 dynamics, has unearthed some possible mechanisms. Particularly informative is a reporter assay developed by the Nicolas laboratory that determines the frequency of size variants of the human G4-forming CEB1 minisatellite when inserted into a yeast chromosome. This work has established that it is replication fork stalling at unresolved G4s that leads to recombination and contraction or expansion of CEB1 repeats, and this is highly dependent on the Pif1 enzyme, an evolutionarily conserved helicase recently recognized to be a highly-efficient G4 helicase [19].

Interestingly, replication-dependent instability of G4 motifs was affected by the direction of the replication fork; it occurred only when the G4 was in the leading-strand template. The persistence of G4 structures on the leading-strand template led to Rad51/Rad52-dependent repair mechanisms, generating characteristic recombination intermediates and resulting in deletions at CEB1 repeats [60]. Determining the molecular detail of G4-dependent DNA transactions at such exquisite resolution may only be possible in a model organism like *S. cerevisiae*, but such work provides a model of how failure to resolve G4 structures can lead to error-prone recombinational repair that may result in the diversification of antigenic repertoires such as those encoded by *vars* or *vsgs*.

Elegant work in yeast has also provided clues about the possible role of G4s in Av occurring via epigenetic transcriptional switching. Paeschke et al. have showed that unresolved G4s can lead to changes in the epigenetic state of adjacent chromatin [46]. They introduced Pif1-binding G4 motifs into a nonessential arm of the yeast chromosome containing two selectable markers and assayed for gross chromosomal rearrangements or deletions (the term “deletion” normally refers to the loss of both markers, and is in contrast to a point mutation, which would affect only one marker). In wild-type strains, the rare loss of resistance to the selective agents did indeed result from loss of the whole chromosomal region. In Pif1-deficient strains, however, loss of resistance occurred much more frequently, and appeared to occur via epigenetic silencing: the sequences and positions of the selectable markers remained unchanged and their expression could be restored through deletion of *SIR2*, a histone-modifying deacetylase enzyme that is known to enforce epigenetic silencing. Thus, failure of G4 structures to be resolved resulted in epigenetic modifications near the G4 motifs. This may be because DNA synthesis becomes uncoupled from histone recycling mechanisms, leading to dysregulation of epigenetic status [44].

The high recombination rate of the HIV-1 provirus allows the virus to delay immune recognition and clearance, and G4 metabolism may be at play here too. Recombination hot spots and PQS have been found to correlate in the U3 promoter [51], *gag* [61] and cPPT [62] regions of the HIV-1 RNA genome, suggesting that G4s may contribute to this high mutation rate by enabling RT template switching during reverse transcription.

**Perspectives**

G4 motifs offer disparate pathogens—from bacteria, to protozoa, to viruses (Fig. 4 and Table 1)—a means by which to regulate DNA and/or RNA dynamics underlying the common virulence phenotypes of antigenic variation and viral latency. Understanding the G4-mediated regulation of pathogen virulence may open the door to novel therapeutic interventions. Similar to anti-cancer strategies targeting G4s in telomeres and oncogenes, a pathogen’s RNA or DNA could be targeted by G4-specific ligands, and the high topological variety amongst G4 structures suggests that a high level of drug specificity could be achieved [63]. Indeed, the G4-stabilising ligand BRACO-19 is able to inhibit HIV-1 infectivity [49,56] and block the proliferation of EBV-positive cells in vitro [64].

The wide range of pathogens using G4s in regulatory processes suggests that G4s may regulate similar processes in other pathogens. For example, in addition to *Neisseria* and *Borrelia*...
there are a number of other prokaryotic and eukaryotic pathogens that use controlled DNA-recombination-associated Av systems. Investigation of G4 dynamics in these pathogens, which include, among others, species of *Mycoplasma* [65], *Trypanosoma* [66] and *Babesia* [67],
would merit an interesting line of investigation. Amongst viruses, the PQS found in the U3 promoter region of the HIV-1 proviral genome were found to be conserved in HIV-2 and simian immunodeficiency virus [51], indicating that G4s may affect the latency program of these viral species. Furthermore, in addition to EBV EBNA1 mRNA, PQS are present in the mRNA of a number of other gammaherpesviral genome maintenance proteins [57], which may also therefore be subject to G4-mediated translational control.

A highly G4-specific monoclonal antibody generated by phage display has recently been used to demonstrate the presence of G4s in the genomic DNA and cytoplasmic RNA of a range of human cells [68,69]. This antibody, or others similarly developed, could provide a useful tool with which to investigate pathogen G4 dynamics. Finally, it is worth noting that research on pathogen G4 dynamics is relevant to those pursuing G4s as anticancer targets. A number of cancers are associated with HPV and gammaherpesviruses such as EBV; therefore the development of new therapeutic targets based on G4s in these viruses also has the potential to reduce the burden of these malignancies.

Acknowledgments

We are grateful to Dr. James Edwards-Smallbone for critical reading of the manuscript.

References

1. Bang (1910) Untersuchungen über die Guanylsäure. Biochemische Zeitschrift 26: 293–311.
2. Gellert M, Lipsett MN, Davies DR (1962) Helix formation by guanylic acid. Proc Natl Acad Sci USA 48: 2013–2018. PMID: 13947099
3. Qin Y, Hurley LH (2008) Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. Biochimie 90: 1149–1171. doi: 10.1016/j.biochi.2008.02.020 PMID: 18355457
4. Yang D, Okamoto K (2010) Structural insights into G-quadruplexes: towards new anticancer drugs. Future Med Chem 2: 619–646. PMID: 20563318
5. Phan AT, Kuryavyi V, Patel DJ (2006) DNA architecture: from G to Z. Curr Opin Struct Biol 16: 288–298. PMID: 16714104
6. Burge S, Parkinson GN, Hazel P, Todd AK, Neidle S (2006) Quadruplex DNA: sequence, topology and structure. Nucleic Acids Research 34: 5402–5415. PMID: 17012276
7. Eddy J, Maizels N (2006) Gene function correlates with potential for G4 DNA formation in the human genome. Nucleic Acids Research 34: 3887–3896. PMID: 16914419
8. Huppert JL, Balasubramanian S (2005) Prevalence of quadruplexes in the human genome. Nucleic Acids Research 33: 2908–2916. PMID: 15914667
9. Huppert JL, Balasubramanian S (2007) G-quadruplexes in promoters throughout the human genome. Nucleic Acids Research 35: 406–413. PMID: 17169996
10. Verma A, Halder K, Halder R, Yadav VK, Rawal P, et al. (2008) Genome-wide computational and expression analyses reveal G-quadruplex DNA motifs as conserved cis-regulatory elements in human and related species. J Med Chem 51: 5641–5649. doi: 10.1021/jm800448a PMID: 18767830
11. Smith JS, Chen Q, Yatsunyk LA, Nicoludis JM, Garcia MS, et al. (2011) Rudimentary G-quadruplex-based telomere capping in Saccharomyces cerevisiae. Nat Struct Mol Biol 18: 5485–5486. doi: 10.1038/nsmb.2033 PMID: 21399640
12. Martadinata H, Heddi B, Lim KW, Phan AT (2011) Structure of long human telomeric RNA (TERRA): G-quadruplexes formed by four and eight UUAGGG repeats are stable building blocks. Biochemistry 50: 6455–6461. doi: 10.1021/bi200569f PMID: 21671673
13. Luke B, Lingner J (2009) TERRA: telomeric repeat-containing RNA. EMBO J 28: 2503–2510. doi: 10.1038/emboj.2009.166 PMID: 19629047
14. Deng Z, Norseen J, Wiedmer A, Riethman H, Lieberman PM (2009) TERRA RNA binding to TRF2 facilitates heterochromatin formation and ORC recruitment at telomeres. Mol Cell 35: 403–413. doi: 10.1016/j.molcel.2009.06.025 PMID: 19716786
15. Biffi G, Tannahill D, Balasubramanian S (2012) An intramolecular G-quadruplex structure is required for binding of telomeric repeat-containing RNA to the telomeric protein TRF2. J Am Chem Soc 134: 11974–11976. doi: 10.1021/ja300573x PMID: 22780456
16. Simonsson T, Pecinka P, Kubista M (1998) DNA tetraplex formation in the control region of c-myc. Nucleic Acids Research 26: 1167–1172. PMID: 9469822
17. Siddiqui-Jain A, Grand CL, Beans DJ, Hurley LH (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. Proc Natl Acad Sci USA 99: 11593–11598. doi: 10.1073/pnas.102105699 PMID: 12195017
18. Kruisselbrink E, Guryev V, Brouwer K, Pontier DB, Cuppen E, et al. (2008) Mutagenic capacity of endogenous G4 DNA underlies genome instability in FANCJ-defective C. elegans. Nat Struct Mol Biol 15: 146–154. doi: 10.1038/nsmb.1372 PMID: 19739048
19. Kouzine F, Sanford S, Elisha-Feil Z, Levens D (2008) The functional response of upstream DNA to dynamic supercoiling in vivo. Nat Struct Mol Biol 15: 146–154. doi: 10.1038/nsmb.1372 PMID: 18193062
20. Kouzine F, Levens D (2007) Supercoll-driven DNA structures regulate genetic transactions. Front Biosci 12: 4409–4423. PMID: 17485385
29. Sun D, Hurley LH (2009) The importance of negative superhelicity in inducing the formation of G-quadruplex and i-motif structures in the c-Myc promoter: implications for drug targeting and control of gene expression. J Med Chem 52: 2863–2874. doi: 10.1021/jm900055s PMID: 19385599

30. Cahoon LA, Seifert HS (2013) Transcription of a cis-acting, noncoding, small RNA is required for pilin antigenic variation in Neisseria gonorrhoeae. PLoS Pathog 9: e1003074. doi: 10.1371/journal.ppat.1003074 PMID: 23349628

31. Wörmann ME, Horien CL, Bennett JS, Jolley KA, Maiden MCJ, et al. (2014) Sequence, distribution and chromosomal context of class I and class II pilin genes of Neisseria meningitidis identified in whole genome sequences. BMC Genomics 15: 253. doi: 10.1186/1471-2164-15-253 PMID: 24690385

32. Walia R, Chaconas G (2013) Suggested role for G4 DNA in recombinational switching at the antigenic variation locus of the Lyme disease spirochete. PLoS ONE 8: e57792. doi: 10.1371/journal.pone.0057792 PMID: 23469068

33. Mir T, Huang SH, Kobryn K (2013) The telomere resolvase of the Lyme disease spirochete, Borrelia burgdorferi, promotes DNA single-strand annealing and strand exchange. Nucleic Acids Research 41: 10438–10448. doi: 10.1093/nar/gkt832 PMID: 24049070

34. Rawal P, Kummarasetti VBR, Ravindran J, Kumar N, Halder K, et al. (2006) Genome-wide prediction of G4 DNA as regulatory motifs: role in Escherichia coli global regulation. Genome Res 16: 644–655. PMID: 16651665

35. Du X, Wojtowicz D, Bowers AA, Levens D, Benham CJ, et al. (2013) The genome-wide distribution of non-B DNA motifs is shaped by operon structure and suggests the transcriptional importance of non-B DNA structures in Escherichia coli. Nucleic Acids Research 41: 5965–5977. doi: 10.1093/nar/gkt308 PMID: 23620297

36. Glover L, Alsford S, Horn D (2013) DNA break site at fragile subtelomeres determines probability and mechanism of antigenic variation in African trypanosomes. PLoS Pathog 9: e1003260. doi: 10.1371/journal.ppat.1003260 PMID: 23555264

37. Kaur R, Domergue R, Zupancic ML, Cormack BP (2005) A yeast by any other name: Candida glabrata and its interaction with the host. Curr Opin Microbiol 8: 378–384. PMID: 15996895

38. Voss TS, Bozdech Z, Bartfai R (2014) Epigenetic memory takes center stage in the survival strategy of malaria parasites. Curr Opin Microbiol 20C: 88–95.

39. Smargiasso N, Gabelica V, Dambion C, Rosu F, De Pauw E, et al. (2009) Putative DNA G-quadruplex formation within the promoters of Plasmodium falciparum var genes. BMC Genomics 10: 362. doi: 10.1038/1471-2164-10-362 PMID: 19660104

40. Merrick CJ, Huttenhower C, Buckee C, Amambua-Ngwa A, Gomez-EScorar N, et al. (2012) Epigenetic dysregulation of virulence gene expression in severe Plasmodium falciparum malaria. J Infect Dis 205: 1593–1600. doi: 10.1093/infdis/jis239 PMID: 22448008

41. Rottmann M, Lavstsen T, Mugasa JP, Kaestli M, Jensen ATR, et al. (2006) Differential expression of var gene groups is associated with morbidity caused by Plasmodium falciparum infection in Tanzanian children. Infect Immun 74: 3904–3911. PMID: 16790763

42. Kyriacou HM, Stone GN, Challis RJ, Raza A, Lyke KE, et al. (2006) Differential var gene transcription in Plasmodium falciparum isolates from patients with cerebral malaria compared to hyperparasitaemia. Mol Biochem Parasitol 150: 211–218. PMID: 16996149

43. Tuteja R (2010) Genome wide identification of Plasmodium falciparum helicases: a comparison with human host. Cell Cycle 9: 104–120. PMID: 20016272

44. Sarkies P, Reams C, Simpson LJ, Sale JE (2010) Epigenetic instability due to defective replication of structured DNA. Mol Cell 40: 703–713. doi: 10.1016/j.molcel.2010.11.009 PMID: 21145480

45. Paeschke K, Bochman ML, Garcia PD, Ciejka P, Friedman KL, et al. (2013) Pf IF1 family helicases suppress genome instability at G-quadruplex motifs. Nature 497: 458–462. doi: 10.1038/nature12149 PMID: 23657261

46. De Cian A, Grellier P, Mouray E, Depoix D, Bertrand H, et al. (2008) Plasmodium telomeric sequences: structure, stability and quadruplex targeting by small compounds. Chembiochem 9: 2730–2739. doi: 10.1002/cbic.200800330 PMID: 18924216

47. Perrone R, Nadai M, Frasson I, Poe JA, Butovskaya E, et al. (2013) A dynamic G-quadruplex region regulates the HIV-1 long terminal repeat promoter. J Med Chem 56: 6521–6530. doi: 10.1021/jm400914r PMID: 23865750
50. Amrane S, Kerkour A, Bedrat A, Vialet B, Andreola M-L, et al. (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. doi: 10.1021/ja501500c PMID: 24649937

51. Piekna-Przybylska D, Sullivan MA, Sharma G, Bambara RA (2014) U3 region in the HIV-1 genome adopts a G-quadruplex structure in its RNA and DNA sequence. Biochemistry 53: 2581–2593. doi: 10.1012/bi4016692 PMID: 24735378

52. Marban C, Suzanne S, Dequiedt F, de Walque S, Redel L, et al. (2007) Recruitment of chromatin-modifying enzymes by CTIP2 promotes HIV-1 transcriptional silencing. EMBO J 26: 412–423. PMID: 17245431

53. Jiang G, Espeseth A, Hazuda DJ, Margolis DM (2007) c-Myc and Sp1 contribute to proviral latency by recruiting histone deacetylase 1 to the human immunodeficiency virus type 1 promoter. J Virol 81: 10914–10923. PMID: 17670825

54. Perrone R, Nadai M, Poe JA, Frasson I, Palumbo M, et al. (2013) Formation of a unique cluster of G-quadruplex structures in the HIV-1 Nef coding region: implications for antiviral activity. PLoS ONE 8: e73121. doi: 10.1371/journal.pone.0073121 PMID: 24015290

55. Basmaciogullari S, Pizzato M (2014) The activity of Nef on HIV-1 infectivity. Front Microbiol 5: 232. doi: 10.3389/fmicb.2014.00232 PMID: 24904546

56. Perrone R, Butovskaya E, Daelemans D, Palù G, Pannecouque C, et al. (2014) Anti-HIV-1 activity of the G-quadruplex ligand BRACO-19. Journal of Antimicrobial Chemotherapy.

57. Murat P, Zhong J, Lekieffre L, Cowieson NP, Clancy JL, et al. (2014) G-quadruplexes regulate Epstein-Barr virus-encoded nuclear antigen 1 mRNA translation. Nat Chem Biol 10: 358–364. doi: 10.1038/nchembio.1479 PMID: 24633353

58. Tellam JT, Zhong J, Lekieffre L, Bhat P, Martinez M, et al. (2014) mRNA Structural Constraints on EBNA1 Synthesis Impact on In Vivo Antigen Presentation and Early Priming of CD8+ T Cells. PLoS Pathog 10: e1004423. doi: 10.1371/journal.ppat.1004423 PMID: 25299404

59. Tlučková K, Marušič M, Tóthová P, Bauer L, Šket P, et al. (2013) Human papillomavirus G-quadruplexes. Biochemistry 52: 7207–7216. doi: 10.1021/bi400897g PMID: 24404463

60. Lopes J, Piazza A, Bermejo R, Kriegsman B, Colosio A, et al. (2011) G-quadruplex-induced instability during leading-strand replication. EMBO J 30: 4033–4046. doi: 10.1038/emboj.2011.316 PMID: 21873979

61. Shen W, Gao L, Balakrishnan M, Bambara RA (2009) A recombination hotspot in HIV-1 contains guanosine runs that can form a G-quartet structure and promote strand transfer in vitro. J Biol Chem 284: 33883–33893. doi: 10.1074/jbc.M109.055368 PMID: 19822521

62. Piekna-Przybylska D, Sharma G, 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. doi: 10.1074/jbc.M113.477265 PMID: 23839990

63. Balasubramanian S, Neidle S (2009) G-quadruplex nucleic acids as therapeutic targets. Curr Opin Chem Biol 13: 345–353. doi: 10.1016/j.cbpa.2009.04.637 PMID: 19156062

64. Norseen J, Johnson FB, 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. doi: 10.1128/JVI.00747-09 PMID: 19656898

65. Iverson-Cabral SL, Astete SG, Cohen CR, Totten PA (2007) mgpB and mgpC sequence diversity in Mycoplasma genitalium is generated by segmental reciprocal recombination with repetitive chromosomal sequences. Mol Microbiol 66: 55–73. PMID: 17880423

66. McCulloch R, Horn D (2009) What has DNA sequencing revealed about the VSG expression sites of African trypanosomes? Trends Parasitol 25: 359–363. doi: 10.1016/j.pt.2009.05.007 PMID: 19632154

67. Brayton KA, Lau AOT, Herndon DR, Hannick L, Kappmeyer LS, et al. (2007) Genome sequence of Babesia bovis and comparative analysis of apicomplexan hemoproteozoan. PLoS Pathog 3: 1401–1413. PMID: 17953480

68. Biffi G, Tannahil D, McCafferty J, Balasubramanian S (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. Nat Chem 5: 182–186. doi: 10.1038/nchem.1548 PMID: 23422569

69. Biffi G, Di Antonio M, Tannahil D, Balasubramanian S (2014) Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. Nat Chem 6: 75–80. doi: 10.1038/nchem.1805 PMID: 24345950

70. Wong HM, Stegle O, Rodgers S, Huppert JL (2010) A toolbox for predicting g-quadruplex formation and stability. J Nucleic Acids 2010. doi: 10.4061/2010/182894 PMID: 21234397

71. Kikin O, D’Antonio L, Bagga PS (2006) GQRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. Nucleic Acids Research 34: W676–W682. PMID: 16845096
72. Yadav VK, Abraham JK, Mani P, Kulshrestha R, Chowdhury S (2008) QuadBase: genome-wide database of G4 DNA—occurrence and conservation in human, chimpanzee, mouse and rat promoters and 146 microbes. Nucleic Acids Research 36: D381–D385. PMID: 17962308

73. Gowan SM, Harrison JR, Patterson L, Valenti M, Read MA, et al. (2002) A G-quadruplex-interactive potent small-molecule inhibitor of telomerase exhibiting in vitro and in vivo antitumor activity. Mol Pharmacol 61: 1154–1162. PMID: 11961134

74. Kim M-Y, Vankayalapati H, Shin-ya K, Wierzba K, Hurley LH (2002) Telomestatin, a potent telomerase inhibitor that interacts quite specifically with the human telomeric intramolecular g-quadruplex. J Am Chem Soc 124: 2098–2099. PMID: 11878947

75. Riou JF, Guittat L, Mailliet P, Laoui A, Renou E, et al. (2002) Cell senescence and telomere shortening induced by a new series of specific G-quadruplex DNA ligands. Proc Natl Acad Sci USA. 99: 2672–2677. PMID: 11854467

76. Paeschke K, Simonsson T, Postberg J, Rhodes D, Lipps HJ (2005) Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. Nat Struct Mol Biol 12: 847–854. PMID: 16142245

77. Schaffitzel C, Berger I, Postberg J, Hanes J, Lipps HJ, et al. (2001) In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei. Proc Natl Acad Sci USA 98: 8572–8577. PMID: 11438689

78. Fang G, Cech TR (1993) The beta subunit of Oxytricha telomere-binding protein promotes G-quartet formation by telomeric DNA. Cell 74: 875–885. PMID: 8374954

79. Fang G, Cech TR (1993) Characterization of a G-quartet formation reaction promoted by the beta-subunit of the Oxytricha telomere-binding protein. Biochemistry 32: 11646–11657. PMID: 8218232

80. Etzioni S, Yafe A, Khateb S, Weisman-Shomer P, Bengali E, et al. (2005) Homodimeric MyoD preferentially binds tetraplex structures of regulatory sequences of muscle-specific genes. J Biol Chem 280: 26805–26812. PMID: 15923190

81. Weisman-Shomer P, Fry M (1994) Stabilization of tetrahelical DNA by the quadruplex DNA binding protein QUAD. Biochem Biophys Res Commun 205: 305–311. PMID: 7999041

82. Zaug AJ, Podell ER, Cech TR (2005) Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro. Proceedings of the National Academy of Sciences 102: 10864–10869. PMID: 16043710

83. Fukuda H, Katahira M, Tsuchiya N, Enokizono Y, Sugimura T, et al. (2002) Unfolding of quadruplex structure in the G-rich strand of the minisatellite repeat by the binding protein UP1. Proc Natl Acad Sci USA 99: 12685–12690. PMID: 12235355

84. Enokizono Y, Konishi Y, Nagata K, Ouhashi K, Uesugi S, et al. (2005) Structure of hnRNP D Complexed with Single-stranded Telomere DNA and Unfolding of the Quadruplex by Heterogeneous Nuclear Ribonucleoprotein D. J Biol Chem 280: 18862–18870. PMID: 15734733

85. Khateb S (2004) Destabilization of tetraplex structures of the fragile X repeat sequence (CGG)n is mediated by homolog-conserved domains in three members of the hnRNP family. Nucleic Acids Research 32: 4145–4154 PMID: 15302914