Molecular Therapy
Nucleic Acids
Original Article

Characterization of G-Quadruplex Motifs in *espB*, *espK*, and *cyp51* Genes of *Mycobacterium tuberculosis* as Potential Drug Targets

Subodh Kumar Mishra,[1,6] Uma Shankar,[1,6] Neha Jain,[1] Kriti Sikri,[2] Jaya Sivaswami Tyagi,[2] Tarun Kumar Sharma,[3] Jean-Louis Mergny,[4,5] and Amit Kumar[1]

[1]Discipline of Biosciences and Biomedical Engineering, Indian Institute of Technology Indore, Indore, Simrol, Indore 453552, India; [2]Department of Biotechnology, All India Institute of Medical Sciences, New Delhi 110029, India; [3]Centre for Bio-design and Diagnostics, Translational Health Science and Technology Institute, Faridabad, Haryana 121001, India; [4]ARNA Laboratory, INSERM U1212, CNRS UMR 5320, Université de Bordeaux, Bordeaux, France; [5]Institute of Biophysics of the Czech Academy of Sciences, Královopolská 135, 612 65 Brno, Czech Republic

G-quadruplex structure forming motifs are among the most studied evolutionarily conserved drug targets that are present throughout the genome of different organisms and susceptible to influencing various biological processes. Here we report highly conserved potential G-quadruplex motifs (PGQs) in three essential genes (*espK*, *espB*, and *cyp51*) among 160 strains of the *Mycobacterium tuberculosis* genome. Products of these genes are involved in pathways that are responsible for virulence determination of bacteria inside the host cell and its survival by maintaining membrane fluidity. The *espK* and *espB* genes are essential players that prevent the formation of mature phagolysosome and antigen presentation by host macrophages. The *cyp51* is another PGQ-possessing gene involved in sterol biosynthesis pathway and membrane formation. In the present study, we revealed the formation of stable intramolecular parallel G-quadruplex structures by *Mycobacterium* PGQs using a combination of techniques (NMR, circular dichroism [CD], and gel electrophoresis). Next, isothermal titration calorimetry (ITC) and CD melting analysis demonstrated that a well-known G-quadruplex ligand, TMPyP4, binds to and stabilizes these PGQ motifs. Finally, polymerase inhibition and qRT-PCR assays highlight the biological relevance of PGQ-possessing genes in this pathogen and demonstrate that G-quadruplexes are potential drug targets for the development of effective anti-tuberculosis therapeutics.

INTRODUCTION

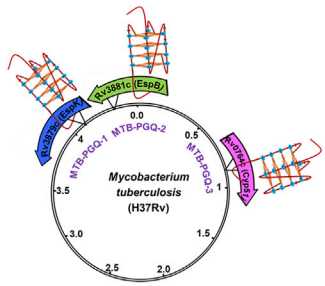
Mycobacterium tuberculosis is a member of the bacterial *Mycobacteriaceae* family and is known to cause tuberculosis (TB) in humans. The World Health Organization (WHO) reported the death of 1.7 million *M. tuberculosis*-infected persons in 2016 and estimated 10.4 million new cases of tuberculosis patients per year worldwide.[1,2] Emergence of multidrug-resistant strains makes the situation more ominous.[1] Therefore, to develop alternative treatments against both drug-susceptible and multidrug-resistant strains, discovery of novel and highly conserved drug targets is warranted. G-quadruplex-forming nucleic acid motifs are one of the most studied evolutionarily conserved drug targets and are copiously found in critical regions of genomes of various organisms ranging from eukaryotes to prokaryotes and viruses.[3,4] In humans, G-quadruplexes have been found to be associated with the regulation of vital cellular processes such as replication, recombination, transcription, and translation.[5] In addition to the telomere, G-quadruplexes have also been identified in the promoter regions of various proto-oncogenes, such as *KRAS*, *Bcl-2*, *c-Kit*, *PDGF-A*, *c-MYC*, *VEGF*, and *c-Myb*, and are considered as potential therapeutic targets for cancer treatment.[6–10] Recently, various viruses such as HIV, Ebola virus, hepatitis B virus (HBV), herpes simplex virus (HSV), human papillomavirus (HPV), hepatitis C virus (HCV), severe acute respiratory syndrome (SARS) coronavirus, Zika virus, and Epstein-Barr virus (EBV) were also reported to harbor conserved G-quadruplex structures in the coding region of various structural and non-structural proteins.[11] In all these cases, G-quadruplexes either play a regulatory role in replication, transcription, and packaging of the viral genome or help the viruses in immune evasion.[11] Among the vast kingdom of prokaryotes, only a few pathogenic bacteria have been explored for the presence of G-quadruplex sequences in their genome such as *Neisseria gonorrhoeae* and *Deinococcus radiodurans*.[12,13] Conserved G-quadruplexes are shown to be responsible for antigenic variations in *Neisseria gonorrhoeae*[14] and *Treponema pallidium*.[15] Previously, several strain-specific G-quadruplexes were also reported in the genome of *Mycobacterium tuberculosis* (H37Rv strain).[16,17] Therefore, here in the present study, we sought to explore novel and highly conserved putative G-quadruplex-forming sequences in all 160 available and completely sequenced strains of *M. tuberculosis*. Bioinformatics analysis revealed the presence of highly conserved potential G-quadruplex motifs

Received 17 September 2018; accepted 22 April 2019;
https://doi.org/10.1016/j.omtn.2019.04.022.

[6]These authors contributed equally to this work.

Correspondence: Amit Kumar, Discipline of Biosciences and Biomedical Engineering, Indian Institute of Technology Indore, Indore, Simrol, Indore 453552, India.
E-mail: amitk@iiti.ac.in

A



B

PGQ-Name	Gene Name	Sequence	G4 Strand	Locus Tag	Gene ID	Associated Functions	Gene Direction
MTB-PGQ1	*espK*	GGGGTTCCCGGGGTGATCGGGGTTCCCGGGG	-	Rv3879c	886212	ESX-1 secretion-associated protein	<<<<
MTB-PGQ4	*espB*	GGGCGGCGGTGGCATGGGAATGCCGATGGGT GCCGCGCATCAGGG	-	Rv3881c	885908	type VII secretion system ESX-1 target EspB	<<<<
MTB-PGQ-9	*cyp51*	GGGGATCGGGGAAGTCTTCGGGGATCCGGTTG GAGATCGCCGGGG	-	Rv0764c	888819	lanosterol 14-alpha demethylase(Cytochrome P450)	>>>>

Figure 1. Conserved G-Quadruplex Motifs in *Mycobacterium tuberculosis*

(A) Schematic representation of MTB-PGQ1 (*espK*), MTB-PGQ2 (*espB*), and MTB-PGQ3 (*cyp51*) sites in the *Mycobacterium tuberculosis* H37Rv genome. (B) List of G-quadruplex motif sequences of MTB-PGQ1 (*espK*), MTB-PGQ2 (*espB*), and MTB-PGQ3 (*cyp51*), the strand in which PGQs present (+, sense; −, antisense strand), gene name, locus ID, gene ID, gene direction (>>>>, sense strand; <<<<, antisense strand), and associated genes function in which PGQs are present.

(PGQs) in three essential genes of *M. tuberculosis*, namely, *espK*, *espB*, and *cyp51* that were responsible for the virulence of bacteria inside the host cell. Insights into pathogenic mechanisms of *M. tuberculosis* infection indicate that it is a facultative intracellular human pathogen and survives in the host environment by preventing the maturation of phagosomes into phagolysosomes.[18,19] Previously, the *M. tuberculosis* genome has been revealed to possess an *extended region of difference 1* (*extRD1*) locus that harbors genes for early secretory antigenic target 6 (ESAT-6) system 1 (ESX-1) secretion machinery, a type VII secretion system (T7SS).[20] This T7SS utilizes EspK (encoded by *Rv3879*) for secreting ESAT-6 and culture filtrate protein 10 (CFP-10) as a dimer (Figure S1A). After secretion, the ESAT-6 and CFP-10 dimer complex enters into the endoplasmic reticulum (ER) of macrophages and sequesters Beta-2-Microglobulin (β2M) protein, an essential component of MHC class I system.[21] This sequestration reduces MHC class I-mediated antigen presentation, an innate immune response of the host cell.[21] In addition to ESAT-6 and CFP-10, ESX-1 also secretes another ESX-substrate protein, EspB. EspB interacts with the C-terminal region of EspK protein and directly portrays an inhibitory role in the maturation of phagosome.[18,20] Upon EspK-assisted secretion of EspB outside the cytosol by ESX-1 secretion machinery (Figure S1A), EspB undergoes oligomerization, binds to phospholipids of host macrophages, and initiates a membranolytic or pore-forming phenomenon inside the phagosome that serves as a crucial initial step of immune evasion for bacteria.[22] Thus, secretion of EspB helps the bacteria in suppressing the innate immune response of the host cell and promotes its survival within the host[20] (Figure S1A). Above insights into ESX-1 secretion machinery revealed the crucial role of EspB and EspK proteins in the pathogenesis of *M. tuberculosis* within the host and suggested that inhibition of *espK* and *espB* expression could lead to a synergistic reduction in *M. tuberculosis* virulence and pathogenesis, and thus represents a novel therapeutic approach. The *cyp51* gene is another PGQ-possessing gene that has been previously reported to be involved in sterol biosynthesis pathway and membrane formation.[23–26] Because our bioinformatics analysis revealed the PGQs in the *espk*, *espB*, and *cyp51* genes, in the present work, we sought to investigate the regulatory role of these PGQs on the expression of PGQ-possessing genes. To accomplish this, we used various biophysical and biochemical techniques, and confirmed the formation of G-quadruplex structures by predicted PGQs. Isothermal titration calorimetry (ITC) and circular dichroism (CD) melting analysis allowed us to demonstrate the affinity of TMPyP4 to PGQs. Next, polymerase stop assay and qRT-PCR assay of RNA from a *M. tuberculosis* culture treated with TMPyP4 confirmed the downregulation of all PGQs possessing genes *espK*, *espB*, and *cyp51*. Therefore, this study demonstrates a novel approach for developing active therapeutics against *M. tuberculosis* infection by exploiting PGQ targets (Figure S1B).

RESULTS

Putative G-Quadruplex Sequences in the *M. tuberculosis* Genome and Their Evolutionary Conservation

Currently, there are very few effective drugs available for the treatment of the *M. tuberculosis* infection, and the emergence of extremely drug-resistant strains and latent infections ring a global alarm for investigating novel conserved drug targets and drug regimens.[27,28] Depending on nucleotide base composition, nucleic acids can adopt several distinct secondary structures such as duplex, triplex, hairpin, and quadruplex structures.[29] Among these secondary structures, the G-quadruplex structure has been represented as the most studied evolutionarily conserved drug target, and several small molecules have been developed that modulate (stabilize or destabilize) these G-quadruplex structures and have potential to be used as a promising drug target in the battle against human pathogens.

Considering the suitability of G-quadruplexes as an evolutionarily conserved promising drug target, we searched the genomic sequences of all available strains of *M. tuberculosis* in the NCBI database for the presence of PTQs (MTB-PGQs) by employing a G-quadruplex (G4) prediction algorithm previously developed by our group. All predictions were then confirmed by tools developed by other groups (Tables S1, S2, S3, S4, S5, S6, S7, S8, S9, and S10).[30–32] Because G4 structure folding relies on different patterns and length of consecutive G-tracts and loops (Figure S2), we set the constraint of 3 or 4 nt

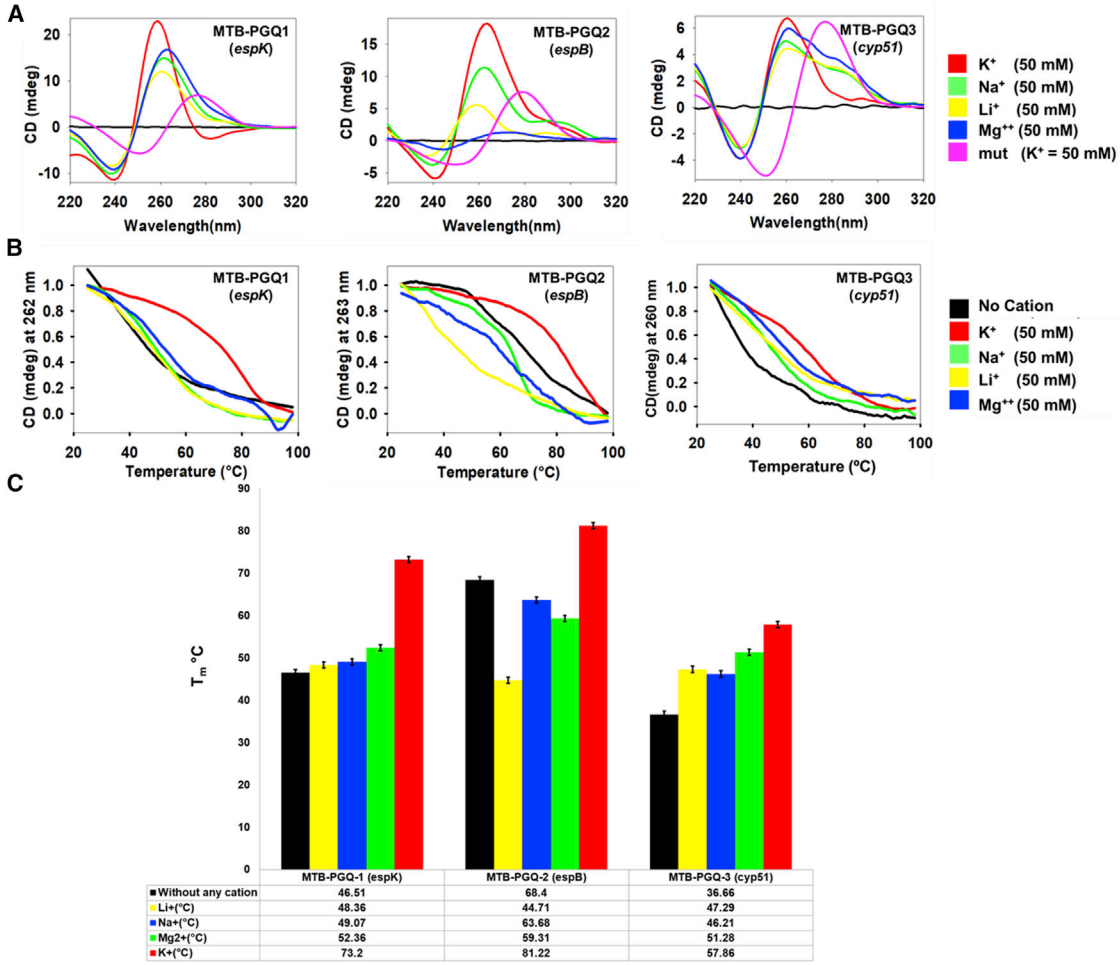


Figure 2. Circular Dichroism Spectra and Melting

(A) Circular dichroism spectra of MTB-PGQ1 (*espK*), MTB-PGQ2 (*espB*), and MTB-PGQ3 (*cyp51*) in Tris-Cl buffer (10 mM) containing either 50 mM KCl (red), 50 mM NaCl (green), 50 mM LiCl (yellow), or 50 mM $MgCl_2$ (blue) and the corresponding mutant in 50 mM KCl (pink). (B) Circular dichroism melting curves of MTB-PGQ1 (*espK*), MTB-PGQ2 (*espB*), and MTB-PGQ3 (*cyp51*) in the absence of any buffer (black) or in the presence of 10 mM Tris-Cl buffer containing 50 mM KCl (red), 50 mM NaCl (green), 50 mM LiCl (yellow), or 50 mM $MgCl_2$ (blue). (C) Bar graph depicting the melting temperature (T_m, °C) for MTB-PGQ1 (*espK*), MTB-PGQ2 (*espB*), and MTB-PGQ3 (*cyp51*).

as the minimum length of G-tract, and 0 and 20 nt as the least and highest length of loops, respectively. The G4 prediction algorithm with the above set of parameters was employed on the 160 completely sequenced *M. tuberculosis* strains (Table S7), allowing us to identify more than 4,000 potential G4-forming sequences (Tables S1 and S2). Considering that the likeness between two sequences may associate with similar functions, we further sought to correlate the sequence similarity between each of the MTB-PGQs and their function. The unweighted pair group method with arithmetic mean (UPGMA) clustering method was utilized to measure the similarity between a large set of MTB-PGQs that defines the likeness hierarchy between the given set of strings.[33] This clustering analysis clusters PGQs that have almost identical sequences, similar function, and are susceptible to form G4s with the same topology (Tables S3 and S4). For combating multidrug-resistant bacterial infection, evolutionary conservation is a requisite that satisfies one of the criteria to work as a promising drug target. Therefore, we employed the BLAST algorithm to calculate the occurrence of each cluster in all 160 available strains of *M. tuberculosis*. To demonstrate the conservation of these clusters, we have employed the following equation: $p = (n \div N) \times 100$, where p is the frequency of occurrence, n is the total number of strains in which a particular PGQ is present, and N is the total number of *M. tuberculosis* strains.

Table S8 demonstrated the details of the MTB-PGQs occurring with >98% frequency and present in the essential genes of the 160 strains of *M. tuberculosis*. For further analysis, we have chosen 100% conserved MTB-PGQs because of their presence in the genome of both drug-susceptible and drug-resistant strains of bacteria. The clustering analysis followed by the conservation mapping revealed three MTB-PGQs as 100% conserved in

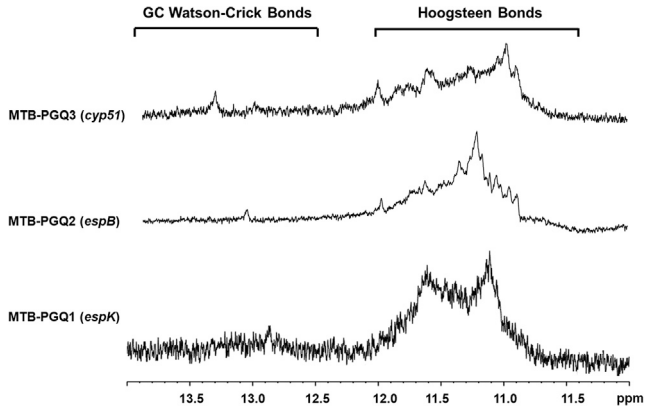


Figure 3. NMR Spectra

1D ^{1}H NMR spectra of MTB-PGQ1 (*espK*), MTB-PGQ2 (*espB*), and MTB-PGQ3 (*cyp51*).

M. tuberculosis (Table S9) and present in three essential genes of the bacteria, namely, *espK* (MTB-PGQ1), *espB* (MTB-PGQ2), and *cyp51* (MTB-PGQ3).

Figure S3 displays the consensus sequence of 100% conserved MTB-PGQs obtained from the R programming-based DECIPHER package. Figures 1A and 1B provide the schematic location, gene ID, locus tag, genome strand of G4, and the direction of the PGQ-possessing genes. A phylogenetic tree obtained from the neighbor-joining methods depicted the conservation of MTB-PGQs among the members of the mycobacteria family that include *Mycobacterium bovis*, *Mycobacterium canettii*, *Mycobacterium microti*, and *Mycobacterium haemophilum* (Figure S4).

Evaluation of the Formation of G4 Structures and Their Stability by Employing Circular Dichroism Spectra and Melting Analysis

Circular dichroism is among the most dependable methods to portray the formation of G4 structures by DNA oligonucleotides. Previous studies have observed that: (1) a positive peak at ~265 nm and a negative peak at ~240 nm correspond to a parallel G4, whereas (2) a positive peak at 290 nm and a negative peak at 260 nm indicate anti-parallel G4 formation, and (3) two positive peaks at 260 and 290 nm coupled with a negative peak at 240 nm represent a mixed or hybrid G4 structure.[34] Different cations exert variable effects on the stability and folding pattern of G4 structures. The ranking of stabilizing ability among the well-studied cations is as follows: $K^+ > Na^+ > Mg^{2+} > Li^+$.[35] Therefore, we assessed the topology of conserved MTB-PGQs using CD spectra and CD melting curve analysis in four different buffers containing related cations K^+, Na^+, Mg^+, and Li^+. CD spectra analysis revealed a predominantly parallel G4 topology exhibited by all conserved MTB-PGQs. As expected, the molar ellipticity and melting temperature of each MTB-PGQ were found to be higher in the presence of K^+ ions (Figure 2). CD spectra analysis performed in the presence of an increasing concentration of K^+ ions revealed an increase in molar ellipticity as a function of K^+ ion concentration (Figure S5). We then mutated the central guanine nucleotide of all of the G stretches into thymine and evaluated the G4-forming ability of mutated MTB-PGQs (mut-MTB-PGQs) in K^+ buffer (Table S10). All of the mutants (mut-MTB-PGQ1, mut-MTB-PGQ2, and mut-MTB-PGQ3) did not show CD signals that correspond to G4 topology and suggested the disruption of the G4 structure formation by the guanine mutation (Figure 2).

Electrophoretic Mobility Shift Assay Indicated the Formation of the Intramolecular G4 by MTB-PGQs

Further, we performed an electrophoretic gel mobility shift assay (EMSA) to check the presence of multimeric structures and determined the molecularity of these assemblies in solution. In the case of G4 folding, a fast-moving oligonucleotide band is considered to correspond to an intramolecular topology, whereas a slow-moving oligonucleotide band is likely to correspond to an intermolecular structure.[36–38] We examined the rate of migration for all MTB-PGQs oligos, as well as mutant sequences on native PAGE. All of the MTB-PGQ oligonucleotides migrate at a faster rate as compared with their mutant sequences, suggesting the formation of intramolecular complexes by all PGQs (Figure S6).

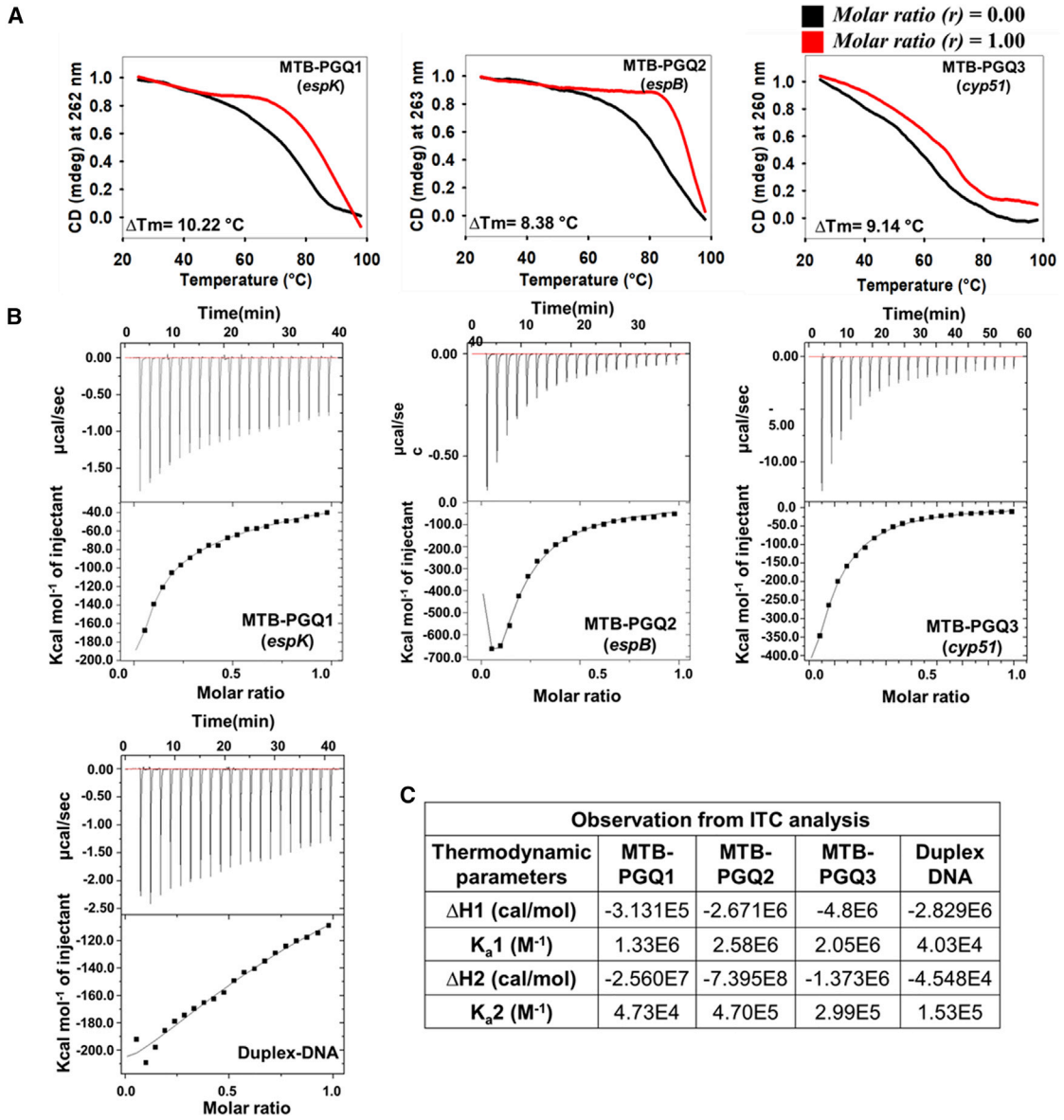
Exploration of G4 Structure Formation by NMR

NMR spectroscopy is a highly reliable technique to confirm the formation of stable G4 structures by nucleic acids.[39,40] Therefore, we examined the one-dimensional proton NMR (1D ^{1}H NMR) spectra of MTB-PGQs *in vitro*. In 1D ^{1}H NMR, the chemical shift signals appear between 10.5 and 12 ppm pertaining to Hoogsteen base pairing involved in G4 formation, whereas signals between 12 and 14 ppm pertain to Watson-Crick base pairing. Encouragingly, 1D ^{1}H NMR spectra of MTB-PGQ1, MTB-PGQ2, and MTB-PGQ3 exhibited imino proton resonances between 10.5 and 12 ppm, and confirmed the formation of G4s by these MTB-PGQs (Figure 3).

Stabilizing and Energetically Favorable Interaction of the PGQs with the Representative G4 Ligand

Various G4 binding ligands have been investigated for their therapeutic potential by affecting the stability of the G4 structure, such as TMPyP4, BRACO-19, etc.[11,41,42] TMPyP4 acts as an anti-cancer agent by stabilizing G4 structures present at human telomeres.[43] It also inhibits viral RNA processing inside host cells by stabilizing the G4 structure present in the L gene of the Ebola virus and acts as an anti-Ebola virus (EBOV) agent.[11,41] Recently, BRACO-19 was shown to inhibit *M. tuberculosis* growth, but its cellular target remains to be identified.[17]

In our study, we assessed the stabilizing activity of TMPyP4 for MTB-PGQs by CD melting curve analysis. Melting curves of MTB-PGQs in the absence and presence of TMPyP4 yielded an increase in melting temperature (T_m) of ~8°C, indicative of a stabilizing effect of TMPyP4 on the thermostability of G4 structures formed by MTB-PGQs (Figure 4A). Further, we assessed the binding affinity of TMPyP4 for the G4 forming MTB-PGQ oligos by ITC. Changes in enthalpy (ΔH_1) associated with binding to MTB-PGQ1, MTB-PGQ-2, and MTB-PGQ3 were very negative, indicative of an



Observation from ITC analysis				
Thermodynamic parameters	MTB-PGQ1	MTB-PGQ2	MTB-PGQ3	Duplex DNA
ΔH1 (cal/mol)	-3.131E5	-2.671E6	-4.8E6	-2.829E6
K_a1 (M^{-1})	1.33E6	2.58E6	2.05E6	4.03E4
ΔH2 (cal/mol)	-2.560E7	-7.395E8	-1.373E6	-4.548E4
K_a2 (M^{-1})	4.73E4	4.70E5	2.99E5	1.53E5

Figure 4. Effect of TMPyP4 on MTB-PGQs

(A) A circular dichroism melting curve depicting the change in T_m of MTB-PGQ1 (*espK*), MTB-PGQ2 (*espB*), and MTB-PGQ3 (*cyp51*) dissolved in 50 mM K^+ buffer in the absence and presence of TMPyP4. (B) Isothermograms of MTB-PGQ1 (*espK*), MTB-PGQ2 (*espB*), MTB-PGQ3, (*cyp51*) and duplex DNA (negative control) obtained from isothermal calorimetry upon addition of TMPyP4. (C) Thermodynamic parameters obtained by two-mode fitting, i.e., change in enthalpy (ΔH1 and ΔH2) and association constants (K_A1 and K_A2) for MTB-PGQ1 (*espK*), MTB-PGQ2 (*espB*), MTB-PGQ3, (*cyp51*) and duplex DNA to TMPyP4.

energetically favorable binding of TMPyP4 to MTB-PGQs.[44] The association constants (K_A) for the high-affinity site of MTB-PGQ1, MTB-PGQ2, and MTB-PGQ3 were 1.33×10^6 M^{-1}, 2.58×10^6 M^{-1}, and 2.05×10^6 M^{-1}, respectively. These results confirm a thermodynamically stable interaction between TMPyP4 and MTB-PGQs with affinities in the micromolar range (Figure 4B). We also took a non-G4-forming sequence as a negative control to observe selectivity of TMPyP4 for the G4 structure. ITC results indicate 33-, 64-, and 50-fold tighter binding of TMPyP4 to MTB-PGQ1, MTB-PGQ2, and MTB-PGQ3, respectively, than to the non-quadruplex-forming sequence.

Effect of TMPyP4 on M. tuberculosis *Growth*

NMR, CD, and EMSA experiments indicated the formation of G4 structures by conserved MTB-PGQ. Moreover, ITC and CD melting analysis revealed energetically favorable high affinity of TMPyP4 for the conserved MTB-PGQs. Therefore, we further addressed the effect of TMPyP4 on *M. tuberculosis* survival using an inhibition assay

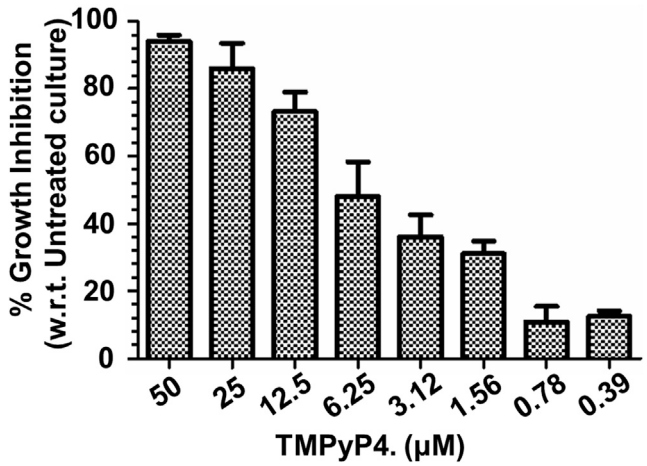


Figure 5. Growth Inhibition of *M. tuberculosis*
Growth inhibition of *M. tuberculosis* H37Rv in the presence of various concentrations of TMPyP4.

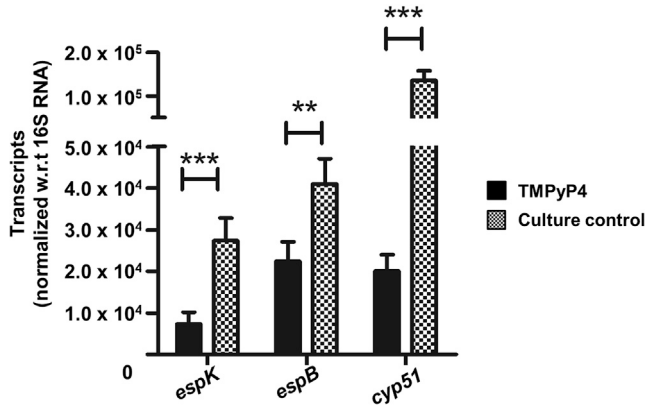


Figure 6. qRT-PCR Analysis of *espK* and *espB* Transcripts
The normalized level (with respect to 16S rRNA transcript) of *espK*, *espB*, and *cyp51* transcripts in *M. tuberculosis* determined by qRT-PCR in the absence (culture control) and presence of TMPyP4. ***$p < 0.001$, **$p < 0.01$.

(Figure 5). The observed IC_{50} value was 6.25 μM, a value lower than the IC_{50} values of the TMPyP4 for both malignant and normal eukaryotic cells.[45]

TMPyP4 Stalls the Replication Machinery at G4 Sites

Next, we sought to investigate whether TMPyP4 binding to the PGQ motif inhibits DNA replication or not. To accomplish this, we employed two variants of stop assay: (1) PCR stop assay and (2) Taq DNA polymerase stop assay.[46] In PCR stop assay, a G-region overlapping primer was utilized for PCR amplification. In the absence of TMPyP4, a double-stranded product formed because of the primer annealing and extension.[46] The presence of TMPyP4 leads to the stabilization of G4 in the template that results in the unavailability of the template for primer annealing, and thus PCR got inhibited (fully inhibited at higher concentration and partially inhibited at a lower concentration). As shown in Figure S7, increasing concentrations of TMPyP4 have a strong impact on PCR amplification. On the contrary, when mut-MTB-PGQs were used as a DNA template, TMPyP4 did not inhibit DNA synthesis leading to full product generation.

However, the Taq polymerase stop assay assesses the ability of TMPyP4 to stabilize the G4 structure and stop the movement of polymerase enzyme during PCR extension reaction at MTB-PGQ sites.[46] Figure S8 depicts the results of Taq polymerase stop assay for the three templates and negative control. In the absence of TMPyP4, full-length product is formed but with increasing concentrations of TMPyP4, the shorter band was observed with the decreasing band intensity, whereas neither smaller product nor diminishing bands intensities were observed in the negative control.[46]

TMPyP4 Reduced the Transcription of espK, espB, *and* cyp51 *Genes*

Next, we sought to examine the effect of TMPyP4 on transcription of the G4 motif-possessing genes *espK*, *espB*, and *cyp51* by qRT-PCR. The transcripts of genes were quantified relative to the transcript of 16S rRNA gene (a housekeeping gene), and quantification cycle (Cq) of non-template controls was >35 cycles. As shown in Figure 6, treatment of the *M. tuberculosis* cells with 6.25 μM TMPyP4 reduced the transcription of the *espK*, *espB*, and *cyp51* genes by 3.70-, 1.91-, and 6.25-fold, respectively, relative to the untreated culture control. These results strengthen the likelihood of a role for a G4-mediated mechanism in the transcription of the G4 motif-containing genes *espK*, *espB*, and *cyp51*, and open a novel therapeutic avenue for combating this deadly human pathogen.

Recently, G4s were recognized as highly conserved and promising drug targets for combating the infection of various polymorphic and lethal human pathogens.[11,47] For example, the presence of the G4 motif in the pre-S2/S gene promoter upregulates the activity of HBV inside human cells.[48] Recently, G4 motifs in the genes of the various structural proteins of the Zika virus were found to be promising drug targets.[49] Similarly, in the case of the Kaposi's sarcoma-associated herpesvirus (KSHV), a G4 motif was found to lie in between the kaposin (DR6) and oriLyt-R, oriLyt-L and K5 9, and v-IRF2 and v-IRF3 genes, and was observed to regulate the replication and transcription of the viral genome inside the host cell.[50] In HIV, a G4 found at unique long terminal repeat (LTR) promoter regions has been proposed as a potent target for anti-HIV therapy.[51] The *pac1* signal required for successful cleavage and packaging of the Human Herpesvirus, and the presence of the G4 motif in this region suggested their novel function in the viral genome.[52]

Altogether, bioinformatics analysis, 1D ^{1}H NMR, CD spectroscopy, and EMSA revealed the presence of 100% conserved intramolecular parallel G4-forming motifs in three essential bacterial genes, namely, *espK* (Rv3879c), *espB* (Rv3881c), and *cyp51* (Rv0764c). Similar to the G4 target in the above-mentioned human pathogen, a G4 motif present in these genes may also serve as a potential target for developing

an anti-bacterial therapy. These G-rich targets can also overcome the problem of the drug resistance because of their 100% conservation in both drug-susceptible and drug-resistant bacterial strains (including extremely drug-resistant strains).

Because these G4s are present in the coding regions of the *espK*, *espB*, and *cyp51* genes, we probed the presence of GQs and the consequence of stabilizing these motifs using TMPyP4, a well-known G4 stabilizing ligand. First, the observed increase in T_m by >8°C in the presence of TMPyP4 established the presence of PGQ sequences in these genes. Second, the presence of PGQ sequences was confirmed by the observed stalling of DNA replication in the presence of TMPyP4 in a PCR stop assay. The binding interaction between TMPyP4 and MTB-PGQs was determined by ITC analysis to be thermodynamically stable, energetically favorable, and selective. Finally, we evaluated the inhibitory effect of TMPyP4 on *M. tuberculosis* growth; IC_{50} value was calculated to be 6.25 μM. Highly stable G4s in the open reading frame of genes were previously shown to downregulate the expression of genes.[13,41,53] The treatment of *M. tuberculosis* cultures with TMPyP4 led to a significant decrease in the expression of *espK* and *espB* relative to 16S rRNA, a housekeeping gene, suggesting that a G4-mediated inhibition mechanism is involved in this process. As a schematic model elaborated in Figure S1B, G4-mediated inhibition of *espK* and *espB* genes is expected to increase the innate immune response of host cell. The inhibited expression of the *espB* and *espK* proteins would restore the phagosome maturation process and reduce the secretion of CFP-10 and ESAT-6, in turn leading to a rescued antigen presentation to the host immune cells.

In conclusion, the current study has identified and confirmed the presence of highly conserved G4 motifs in virulence determining genes of *M. tuberculosis*. Moreover, the presence of these PGQs in other *Mycobacterium* species suggests their conserved role in the survival and virulence of bacteria, and allows us to propose them as a suitable therapeutic target for inhibiting the bacterial pathogenicity. The presence of GQs was confirmed using various biophysical approaches in the presence of a G4 stabilizing agent, TMPyP4. Lastly, the inhibitory role of GQs in the expression of *espK* and *espB* virulence genes and in the survival of *M.* tuberculosis was demonstrated. TMPyP4 can inhibit intracellular transcription of the G4-containing genes. Taking together all the results, this study demonstrates that the G4s in the *espK* and *espB* genes can be considered as novel targets for the development of anti-tuberculosis compounds.

MATERIALS AND METHODS

Please see the Supplemental Information for the detailed explanation of all experimental procedures and materials used in this study.

Availability

Putative G4 prediction tool is an open web-based tool to predict putative G4-forming sequences in different nucleotide stretches (http://bsbe.iiti.ac.in/bsbe/ipdb/pattern2.php).

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.omtn.2019.04.022.

AUTHOR CONTRIBUTIONS

A.K. conceptualized the idea, and A.K. and S.K.M. designed the methodology. Data mining and curation and biophysical experiments were performed by S.K.M., U.S., and N.J. Cell-based studies and RT-PCR were performed by S.K.M. and K.S. under the supervision of A.K., T.K.S., and J.S.T. Original draft was written by S.K.M. and was reviewed by A.K., T.K.S., J.S.T., and J.-L.M.

CONFLICTS OF INTEREST

The authors declare no competing interests.

ACKNOWLEDGMENTS

The authors are thankful to the SIC Facility at IIT Indore for NMR and CD experiments. S.K.M., N.J., and U.S. are thankful to UGC, New Delhi, CSIR, New Delhi, and MHRD, New Delhi for their individual fellowships, respectively. K.S. is thankful to the Department of Science and Technology, Government of India (PDF/2016/000745) for her post-doctoral funding. A.K. and J.-L.M. acknowledges funding from the SERB, DST, New Delhi for the EMR grant (EMR/2016/003897) and SYMBIT project (reg. no. CZ.02.1.01/0.0/0.0/15_003/0000477) financed by the ERDF, respectively.

REFERENCES

1. Silva, D.R., Mello, F.C.Q., Kritski, A., Dalcolmo, M., Zumla, A., and Migliori, G.B. (2018). Tuberculosis series. J. Bras. Pneumol. *44*, 71–72.
2. Ogundipe, T., Otolorin, A., Ogundipe, F., Sran, G., Abbas, G., and Filani, O. (2018). Multidrug-resistant Tuberculosis Lymphadenitis as the Initial Presentation of Secondary Multidrug-resistant Tuberculosis: A Case Report. Cureus *10*, e2363.
3. Rawal, P., Kummarasetti, V.B.R., Ravindran, J., Kumar, N., Halder, K., Sharma, R., Mukerji, M., Das, S.K., and Chowdhury, S. (2006). Genome-wide prediction of G4 DNA as regulatory motifs: role in Escherichia coli global regulation. Genome Res. *16*, 644–655.
4. Murat, P., and Balasubramanian, S. (2014). Existence and consequences of G-quadruplex structures in DNA. Curr. Opin. Genet. Dev. *25*, 22–29.
5. Hänsel-Hertsch, R., Di Antonio, M., and Balasubramanian, S. (2017). DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. Nat. Rev. Mol. Cell Biol. *18*, 279–284.
6. Jafri, M.A., Ansari, S.A., Alqahtani, M.H., and Shay, J.W. (2016). Roles of telomeres and telomerase in cancer, and advances in telomerase-targeted therapies. Genome Med. *8*, 69.
7. Tawani, A., and Kumar, A. (2015). Structural insight into the interaction of flavonoids with human telomeric sequence. Sci. Rep. *5*, 17574.
8. Tawani, A., Amanullah, A., Mishra, A., and Kumar, A. (2016). Evidences for Piperine inhibiting cancer by targeting human G-quadruplex DNA sequences. Sci. Rep. *6*, 39239.
9. Tawani, A., Mishra, S.K., and Kumar, A. (2017). Structural insight for the recognition of G-quadruplex structure at human c-myc promoter sequence by flavonoid Quercetin. Sci. Rep. *7*, 3600.
10. Bonsignore, R., Trippodo, E., and Barone, G. (2017). Targeting G-quadruplex DNA as potential anti-cancer therapy: from hybrid DNA nanostructures to functional devices. In DNA Nanotechnology for Bioanalysis, G. Arrabito and L. Wang, eds. (World Scientific), pp. 129–162.

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

12. Kota, S., Dhamodharan, V., Pradeepkumar, P.I., and Misra, H.S. (2015). G-quadruplex forming structural motifs in the genome of Deinococcus radiodurans and their regulatory roles in promoter functions. Appl. Microbiol. Biotechnol. *99*, 9761–9769.

13. Endoh, T., Kawasaki, Y., and Sugimoto, N. (2013). Suppression of gene expression by G-quadruplexes in open reading frames depends on G-quadruplex stability. Angew. Chem. Int. Ed. Engl. *52*, 5522–5526.

14. Cahoon, L.A., and Seifert, H.S. (2009). An alternative DNA structure is necessary for pilin antigenic variation in Neisseria gonorrhoeae. Science *325*, 764–767.

15. Giacani, L., Brandt, S.L., Puray-Chavez, M., Reid, T.B., Godornes, C., Molini, B.J., Benzler, M., Hartig, J.S., Lukehart, S.A., and Centurion-Lara, A. (2012). Comparative investigation of the genomic regions involved in antigenic variation of the TprK antigen among treponemal species, subspecies, and strains. J. Bacteriol. *194*, 4208–4225.

16. Thakur, R.S., Desingu, A., Basavaraju, S., Subramanya, S., Rao, D.N., and Nagaraju, G. (2014). Mycobacterium tuberculosis DinG is a structure-specific helicase that unwinds G4 DNA: implications for targeting G4 DNA as a novel therapeutic approach. J. Biol. Chem. *289*, 25112–25136.

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

18. Solomonson, M., Setiaputra, D., Makepeace, K.A.T., Lameignere, E., Petrotchenko, E.V., Conrady, D.G., Bergeron, J.R., Vuckovic, M., DiMaio, J., Borchers, C.H., et al. (2015). Structure of EspB from the ESX-1 type VII secretion system and insights into its export mechanism. Structure *23*, 571–583.

19. Huang, D., and Bao, L. (2016). Mycobacterium tuberculosis EspB protein suppresses interferon-γ-induced autophagy in murine macrophages. J. Microbiol. Immunol. Infect. *49*, 859–865.

20. McLaughlin, B., Chon, J.S., MacGurn, J.A., Carlsson, F., Cheng, T.L., Cox, J.S., and Brown, E.J. (2007). A mycobacterium ESX-1-secreted virulence factor with unique requirements for export. PLoS Pathog. *3*, e105.

21. Sreejit, G., Ahmed, A., Parveen, N., Jha, V., Valluri, V.L., Ghosh, S., and Mukhopadhyay, S. (2014). The ESAT-6 protein of Mycobacterium tuberculosis interacts with beta-2-microglobulin (β2M) affecting antigen presentation function of macrophage. PLoS Pathog. *10*, e1004446.

22. Chen, J.M., Zhang, M., Rybniker, J., Boy-Röttger, S., Dhar, N., Pojer, F., and Cole, S.T. (2013). Mycobacterium tuberculosis EspB binds phospholipids and mediates EsxA-independent virulence. Mol. Microbiol. *89*, 1154–1166.

23. Jackson, C.J., Lamb, D.C., Marczylo, T.H., Parker, J.E., Manning, N.L., Kelly, D.E., and Kelly, S.L. (2003). Conservation and cloning of CYP51: a sterol 14 alpha-demethylase from Mycobacterium smegmatis. Biochem. Biophys. Res. Commun. *301*, 558–563.

24. Sagatova, A.A., Keniya, M.V., Wilson, R.K., Monk, B.C., and Tyndall, J.D. (2015). Structural Insights into Binding of the Antifungal Drug Fluconazole to Saccharomyces cerevisiae Lanosterol 14α-Demethylase. Antimicrob. Agents Chemother. *59*, 4982–4989.

25. Trösken, E.R., Adamska, M., Arand, M., Zarn, J.A., Patten, C., Völkel, W., and Lutz, W.K. (2006). Comparison of lanosterol-14 alpha-demethylase (CYP51) of human and Candida albicans for inhibition by different antifungal azoles. Toxicology *228*, 24–32.

26. Warrilow, A.G., Martel, C.M., Parker, J.E., Melo, N., Lamb, D.C., Nes, W.D., Kelly, D.E., and Kelly, S.L. (2010). Azole binding properties of Candida albicans sterol 14-alpha demethylase (CaCYP51). Antimicrob. Agents Chemother. *54*, 4235–4245.

27. Pontali, E., Sotgiu, G., D'Ambrosio, L., Centis, R., and Migliori, G.B. (2016). Bedaquiline and multidrug-resistant tuberculosis: a systematic and critical analysis of the evidence. Eur. Respir. J. *47*, 394–402.

28. Getahun, H., Matteelli, A., Abubakar, I., Aziz, M.A., Baddeley, A., Barreira, D., Den Boon, S., Borroto Gutierrez, S.M., Bruchfeld, J., Burhan, E., et al. (2015). Management of latent Mycobacterium tuberculosis infection: WHO guidelines for low tuberculosis burden countries. Eur. Respir. J. *46*, 1563–1576.

29. Burge, S., Parkinson, G.N., Hazel, P., Todd, A.K., and Neidle, S. (2006). Quadruplex DNA: sequence, topology and structure. Nucleic Acids Res. *34*, 5402–5415.

30. Mishra, S.K., Tawani, A., Mishra, A., and Kumar, A. (2016). G4IPDB: A database for G-quadruplex structure forming nucleic acid interacting proteins. Sci. Rep. *6*, 38144.

31. Kikin, O., D'Antonio, L., and Bagga, P.S. (2006). QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. Nucleic Acids Res. *34*, W676–W682.

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

33. Matsen, F.A., 4th, and Evans, S.N. (2013). Edge principal components and squash clustering: using the special structure of phylogenetic placement data for sample comparison. PLoS ONE *8*, e56859.

34. Kypr, J., Kejnovská, I., Renčiuk, D., and Vorlícková, M. (2009). Circular dichroism and conformational polymorphism of DNA. Nucleic Acids Res. *37*, 1713–1725.

35. Largy, E., Marchand, A., Amrane, S., Gabelica, V., and Mergny, J.-L. (2016). Quadruplex Turncoats: Cation-Dependent Folding and Stability of Quadruplex-DNA Double Switches. J. Am. Chem. Soc. *138*, 2780–2792.

36. Víglaský, V., Bauer, L., and Tlucková, K. (2010). Structural features of intra- and intermolecular G-quadruplexes derived from telomeric repeats. Biochemistry *49*, 2110–2120.

37. Wang, Y., Zhao, M., Zhang, Q., Zhu, G.F., Li, F.F., and Du, L.F. (2015). Genomic distribution and possible functional roles of putative G-quadruplex motifs in two subspecies of Oryza sativa. Comput. Biol. Chem. *56*, 122–130.

38. Garg, R., Aggarwal, J., and Thakkar, B. (2016). Genome-wide discovery of G-quadruplex forming sequences and their functional relevance in plants. Sci. Rep. *6*, 28211.

39. Weldon, C., Eperon, I.C., and Dominguez, C. (2016). Do we know whether potential G-quadruplexes actually form in long functional RNA molecules? Biochem. Soc. Trans. *44*, 1761–1768.

40. Adrian, M., Heddi, B., and Phan, A.T. (2012). NMR spectroscopy of G-quadruplexes. Methods *57*, 11–24.

41. 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. (2016). Chemical Targeting of a G-Quadruplex RNA in the Ebola Virus L Gene. Cell Chem. Biol. *23*, 1113–1122.

42. Kleideiter, E., Piotrowska, K., and Klotz, U. (2007). Screening of telomerase inhibitors. Methods Mol. Biol. *405*, 167–180.

43. Zheng, X.H., Nie, X., Liu, H.Y., Fang, Y.M., Zhao, Y., and Xia, L.X. (2016). TMPyP4 promotes cancer cell migration at low doses, but induces cell death at high doses. Sci. Rep. *6*, 26592.

44. Du, X., Li, Y., Xia, Y.-L., Ai, S.-M., Liang, J., Sang, P., Ji, X.L., and Liu, S.Q. (2016). Insights into Protein-Ligand Interactions: Mechanisms, Models, and Methods. Int. J. Mol. Sci. *17*, 144.

45. Brito, H., Martins, A.C., Lavrado, J., Mendes, E., Francisco, A.P., Santos, S.A., Ohnmacht, S.A., Kim, N.S., Rodrigues, C.M., Moreira, R., et al. (2015). Targeting KRAS Oncogene in Colon Cancer Cells with 7-Carboxylate Indolo[3,2-b]quinoline Tri-Alkylamine Derivatives. PLoS ONE *10*, e0126891.

46. Guittat, L., Lacroix, L., Gomez, D., Arimondo, P.B., De Cian, A., Pennarun, G., Amrane, S., Alberti, P., Lemarteleur, T., Aouali, N., et al. (2004). Quadruplex structures and quadruplex ligands. In Dynamical Genetics, V. Parisi, V.D. Fonzo, and F. Aluffi-Pentini, eds. (Research Signpost), pp. 199–236.

47. Harris, L.M., and Merrick, C.J. (2015). G-quadruplexes in pathogens: a common route to virulence control? PLoS Pathog. *11*, e1004562.

48. 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.

49. Fleming, A.M., Ding, Y., Alenko, A., and Burrows, C.J. (2016). Zika Virus Genomic RNA Possesses Conserved G-Quadruplexes Characteristic of the Flaviviridae Family. ACS Infect. Dis. *2*, 674–681.

50. Madireddy, A., Purushothaman, P., Loosbroock, C.P., Robertson, E.S., Schildkraut, C.L., and Verma, S.C. (2016). G-quadruplex-interacting compounds alter latent DNA replication and episomal persistence of KSHV. Nucleic Acids Res. *44*, 3675–3694.

51. Perrone, R., Butovskaya, E., Daelemans, D., Palù, G., Pannecouque, C., and Richter, S.N. (2014). Anti-HIV-1 activity of the G-quadruplex ligand BRACO-19. J. Antimicrob. Chemother. *69*, 3248–3258.

52. 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.

53. Endoh, T., Kawasaki, Y., and Sugimoto, N. (2013). Stability of RNA quadruplex in open reading frame determines proteolysis of human estrogen receptor α. Nucleic Acids Res. *41*, 6222–6231.