ACS Infectious Diseases

Letter

pubs.acs.org/journal/aidcbc

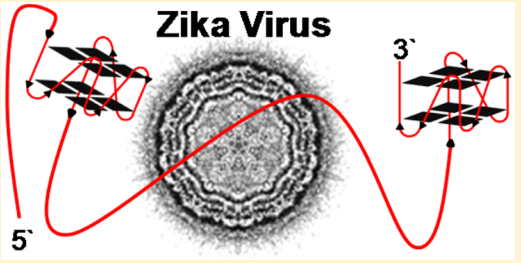
Zika Virus Genomic RNA Possesses Conserved G-Quadruplexes Characteristic of the Flaviviridae Family

Aaron M. Fleming,* Yun Ding, Anton Alenko, and Cynthia J. Burrows*

Department of Chemistry, University of Utah, 315 South 1400 East, Salt Lake City, Utah 84112-0850, United States

Supporting Information

ABSTRACT: Zika virus has emerged as a global concern because neither a vaccine nor antiviral compounds targeting it exist. A structure for the positive-sense RNA genome has not been established, leading us to look for potential G-quadruplex sequences (PQS) in the genome. The analysis identified >60 PQSs in the Zika genome. To minimize the PQS population, conserved sequences in the Flaviviridae family were found by sequence alignment, identifying seven PQSs in the prM, E, NS1, NS3, and NS5 genes. Next, alignment of 78 Zika strain genomes identified a unique PQS near the end of the 3′-UTR. Structural studies on the G-quadruplex sequences found four of the conserved Zika virus sequences to adopt stable, parallel-stranded folds that bind a G-quadruplex-specific compound, and one that was studied caused polymerase stalling when folded to a G-quadruplex. Targeting these PQSs with G-quadruplex binding molecules validated in previous clinical trials may represent a new approach for inhibiting viral replication.



KEYWORDS: *Zika virus, G-quadruplex, RNA genome structure, Flaviviridae*

The recent explosion of Zika virus infections has resulted in focused attention by the World Health Organization (WHO) when they declared in February 2016 that Zika virus is a public health emergency of international concern.[1] This new focus arises due to the strong correlation of Zika virus infections with increased risk of microcephaly and Guillain−Barré syndrome.[2] At present, neither a vaccine nor antiviral compounds have emerged for preventing or treating Zika virus infections, prompting the WHO to strongly advise scientists and public health workers to focus their attentions on these topics.[1] Zika virus is in the genus *Flavivirus* in the Flaviviridae viral family.[2] Characteristics of the Flaviviridae family include mammalian and human hosts with infections spread by arthropod vectors, as in the case of Zika *Aedes* mosquitoes.[2] Flavivirus particles are enveloped, possess icosahedral-like symmetry, and have a diameter of ∼40 nm.[3] More importantly, flaviviruses have positive-sense, single-stranded RNA genomes of ∼11 kb in length.[2,3] Their genomes have a 5′-capped untranslated region (UTR), a code for 10 proteins, three of which are structural (C, prM, and E) and seven nonstructural (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5), as well as a 3′-UTR essential for viral replication.[2,3] The proteins are translated directly from the RNA genome as a polyprotein that is cleaved to the individual proteins by host and viral proteases.[4] A cryo-EM structure for the Zika virus particle has recently been determined,[5] as well as X-ray structures for NS1, the NS3 helicase, and the NS2B−NS3 protease.[6−8] In contrast, the global structure of the RNA genome is not known.[9] This knowledge gap encouraged us to inspect the genome for unique secondary structures found in the Zika virus and the Flaviviridae family.

More specifically, analysis to identify local regions having the ability to adopt potential G-quadruplex sequences (PQSs) in the Zika viral genome was conducted. G-quadruplex folds occur when at least four contiguous runs of two or more guanosine (G) nucleotides exist in a short sequence. The Gs fold around cellular K^+ ions to form G-tetrads composed of four Hoogsteen base-paired Gs.[10] The tetrads stack to adopt G-quadruplex folds in which the intervening nucleotides are the loops connecting the structure (Figure 1). The cellular presence of G-quadruplexes has been the focus of much debate with recent cellular imaging assays supporting their formation in vivo.[11,12] In DNA, G-quadruplexes are conformationally dynamic in a way that is dependent on the sequence and physical context; also, they are generally composed of three tetrads.[13] Cellular

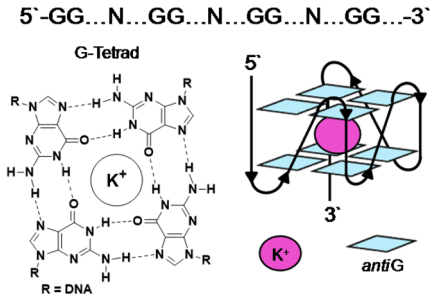
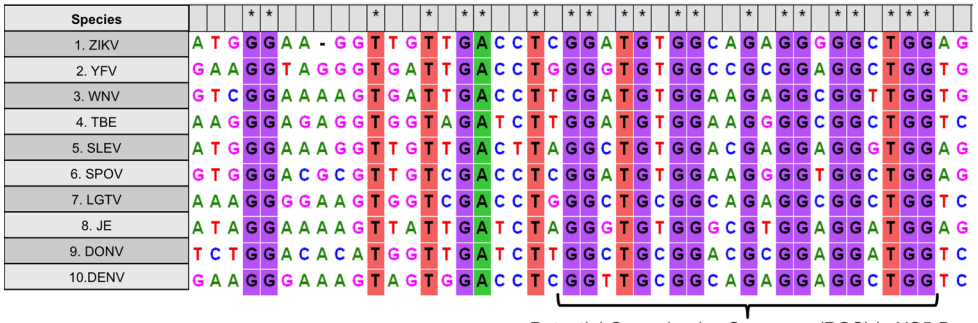


Figure 1. Generic G-quadruplex forming sequence, G-tetrad structure, and a model parallel-stranded, G-quadruplex fold.

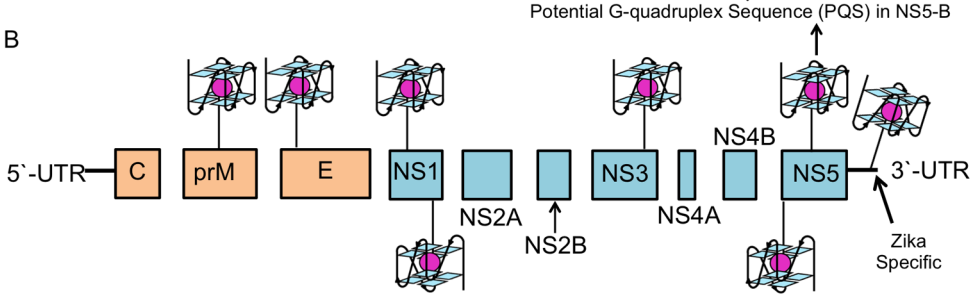
Received: June 14, 2016
Published: August 7, 2016

DOI: 10.1021/acsinfecdis.6b00109
ACS Infect. Dis. 2016, 2, 674−681

A



B



C

PQS Name	Gene or UTR Location	PQS Location	PQS Sequence (red regions were structurally characterized)
prM	473-966	1265-1318	5`-GUGGACAGAGGUUGGGGA AACGGUUGUGGACUUUUUGGCA AAGGGAGCUUGGUG
E	967-2476	2225-2297	5`-AUGGCAGUCCUGGGGGAUACAGCCU GGGACUUCGGAUCAGUCGGGGGUG
NS1-A	2477-3535	2813-2835	5`- AUGGCUGGAAAGCCUGGGGGAA
NS1-B		3455-3487	5`-ACGGCUGCUGGUAUGGAAUGGAGAUAAGGCC
NS3	4601-6451	5027-5059	5`-GUGGAAGAGUGAUAGGACUCUAUGGCAAUGGGGUU
NS5-A	7654-10,364	7654-7692	5`-GUGGAGGUGGGACGGGAGAGACUCUGGGAGAGAAGUGGAA
NS5-B		7875-7917	5`-AUGGGAAGGUUGUUGACCUCGGAUGUGGCAGAGGGGGCUGGAG
3`-UTR*	10,365-10,794	10,762-10,781	5`-GCGGCGGCCGGUGUGGGGAA

Figure 2. Example of sequence alignment, location of conserved PQS sites, and Zika virus sequences for each conserved PQS. (A) Example of alignment of 10 Flaviviridae viruses at the NS5-B site. The complete alignment of all 66 viral genome at each site can be found in Figure S2. (B) Flaviviridae genome model showing pictorially the conserved PQS sites. (C) Zika virus sequences at each of the conserved PQSs, in which the G runs are underlined. * The 3′-UTR sequence is unique to the Zika virus.

experimental studies support DNA PQSs as critical *cis*-acting regulatory elements in many signaling pathways.[14,15] These observations have led to a number of compounds being developed to target these PQSs, whereas some have gone through clinical trials for treatment of disease.[14] In RNA, the single-stranded context favors these structures, in which they are generally less dynamic and less polymorphic than those found in DNA.[16] Additionally, G-quadruplexes in RNA under physiological conditions can fold to stable structures with only two tetrads.[17,18] Roles for RNA PQSs have been ascribed to mRNA splicing, transcriptional termination, and translational control.[15]

Only a few studies have demonstrated important functions for PQSs in viruses. Specifically, the presence of a PQS in the long terminal repeat promoter of the human immunodeficiency virus (HIV) was found to be essential for promoter activity.[19] Stabilization of the PQS with BRACO-19, a trisubstituted acridine derivative specific to G-quadruplexes, was found to inhibit viral replication at the DNA level.[19] The Epstein−Barr virus was shown to utilize an unusual PQS region as a *cis*-acting regulatory region for translation of the Epstein−Barr virus-encoded nuclear antigen 1 (EBNA1) mRNA.[20] Treatment of Epstein−Barr virus-infected cells with the G-quadruplex-specific compound pyridostatin led to decreased EBNA1 protein synthesis. A functional purpose for PQSs has also been observed in the herpes simplex virus,[21] the simian virus 40 T-antigen,[22] and the hepatitis C virus that is also in the Flaviviridae family.[23] These studies highlight a critical role for PQSs in viruses and their ability to be targeted with small-molecule drugs. Therefore, as a first step, inspection of the Zika virus genome for PQS sites was conducted. Depending on the algorithm utilized,[24−26] 64−78 PQSs were found (Figure S1). This population of PQSs was much too large for further study. Thus, we resorted to aligning all known Flaviviridae genome sequences for identification of conserved PQSs throughout the family. Retention of conserved PQSs in the family supports a possibly critical role for these structured sites.

Deposited in the National Center for Biotechnology Information (NCBI) database are 66 Flaviviridae viral genome sequences we used for alignment. Notable viral genomes

DOI: 10.1021/acsinfecdis.6b00109
ACS Infect. Dis. 2016, 2, 674−681

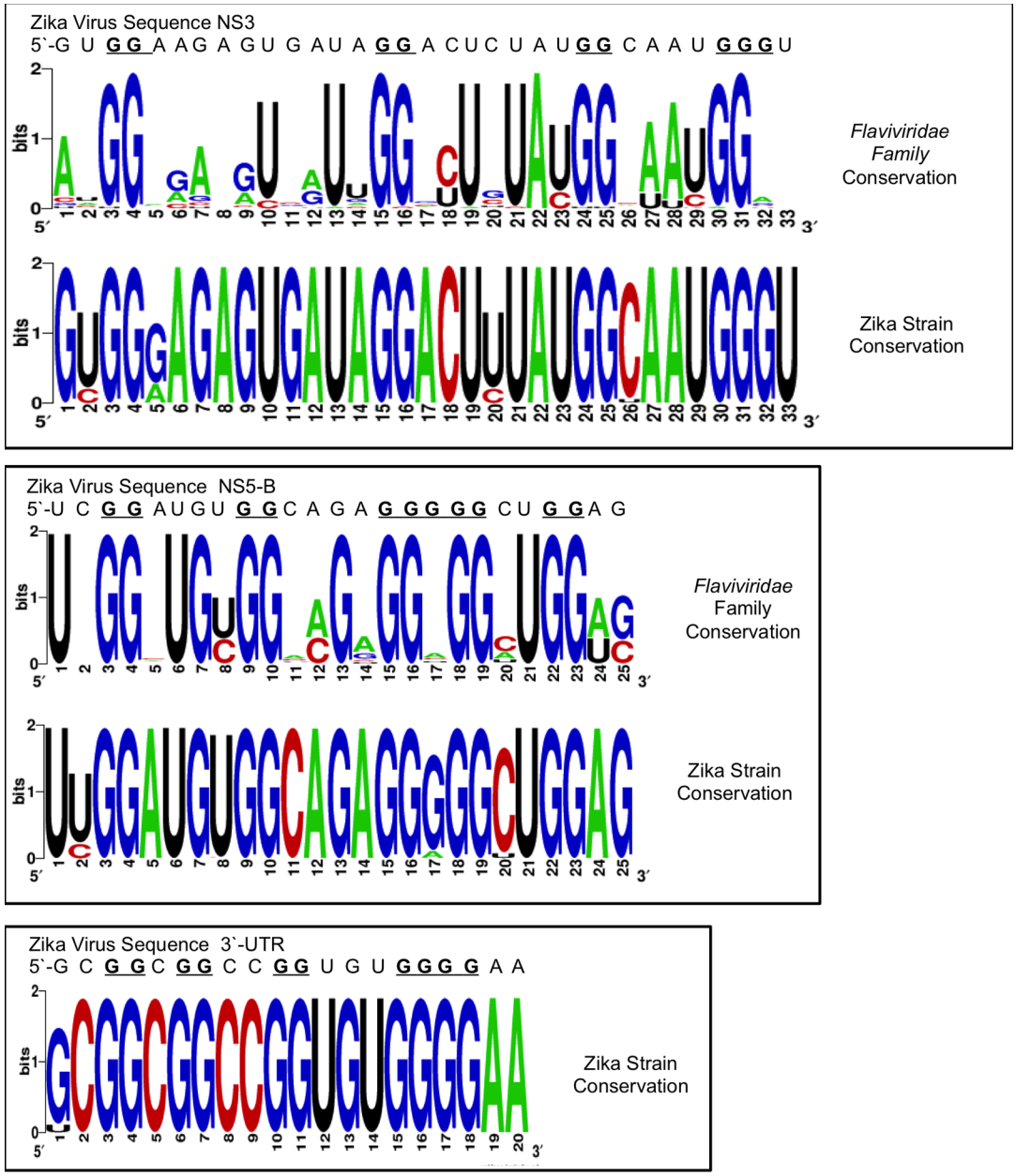


Figure 3. Examples of sequence logos constructed from the NS3 and NS5-B PQS sites from the Flaviviridae family, as well as logos generated for the NS3, NS5-B, and 3′-UTR sequences constructed from the Zika viral strain genomes. The sequence logos were generated using Weblogo,[29] and the genome sequences were found in the NCBI database (Figure S2). The large variability in loop sequence between the viruses is exemplified by the small or nonexistent letters between the highly conserved G runs. Sequence logos for the other PQSs are located in Figure S5.

beyond Zika virus (ZIKV) included in the analysis were the West Nile virus (WNV), dengue virus (DENV), yellow fever virus (YFV), tickborne encephalitis virus (TBE), Japanese encephalitis virus (JE), St. Louis encephalitis virus (SLEV), Donggang virus (DONV), Langat virus (LGTV), and Spondweni virus (SPOV). The hepatitis C viral genome poorly aligned with the other Flaviviridae genome sequences and was not added to the alignment. The genomes were globally aligned using the DECIPHER package in R, and the alignments were visualized using MEGA7 software (see the Supporting Information methods section for details).[27,28] After alignment and examination of the sequences, we found seven conserved PQSs that exist in >56 members of the Flaviviridae virus family studied. In Figure 2A, a representative region of the alignment for one PQS in 10 of the viruses is provided, and the complete alignments for the other PQSs and all viruses can be found in Figures S2 and S3. The seven conserved PQSs were found in the coding regions for the prM, E, NS1, NS3, and NS5 proteins (Figure 2B,C). Because there were two conserved PQSs in the NS1 and NS5 protein coding regions, the first one is denoted with an A following the sequence and the second with a B (e.g., NS1-A and NS1-B). In the next step, alignment of 78 Zika virus strains deposited in the NCBI database (as of June 11, 2016) was conducted to determine the conservation of the PQS sites across strains. The analysis identified the PQS sites to be well maintained through the Zika strains (Figure S4). The alignments also identified a PQS near the 3′ end of the genome specific to the Zika virus (Figure 2).

Critical for maintaining G-quadruplex folds is strong conservation of the G nucleotides, whereas variations in the

DOI: 10.1021/acsinfecdis.6b00109
ACS Infect. Dis. 2016, 2, 674−681

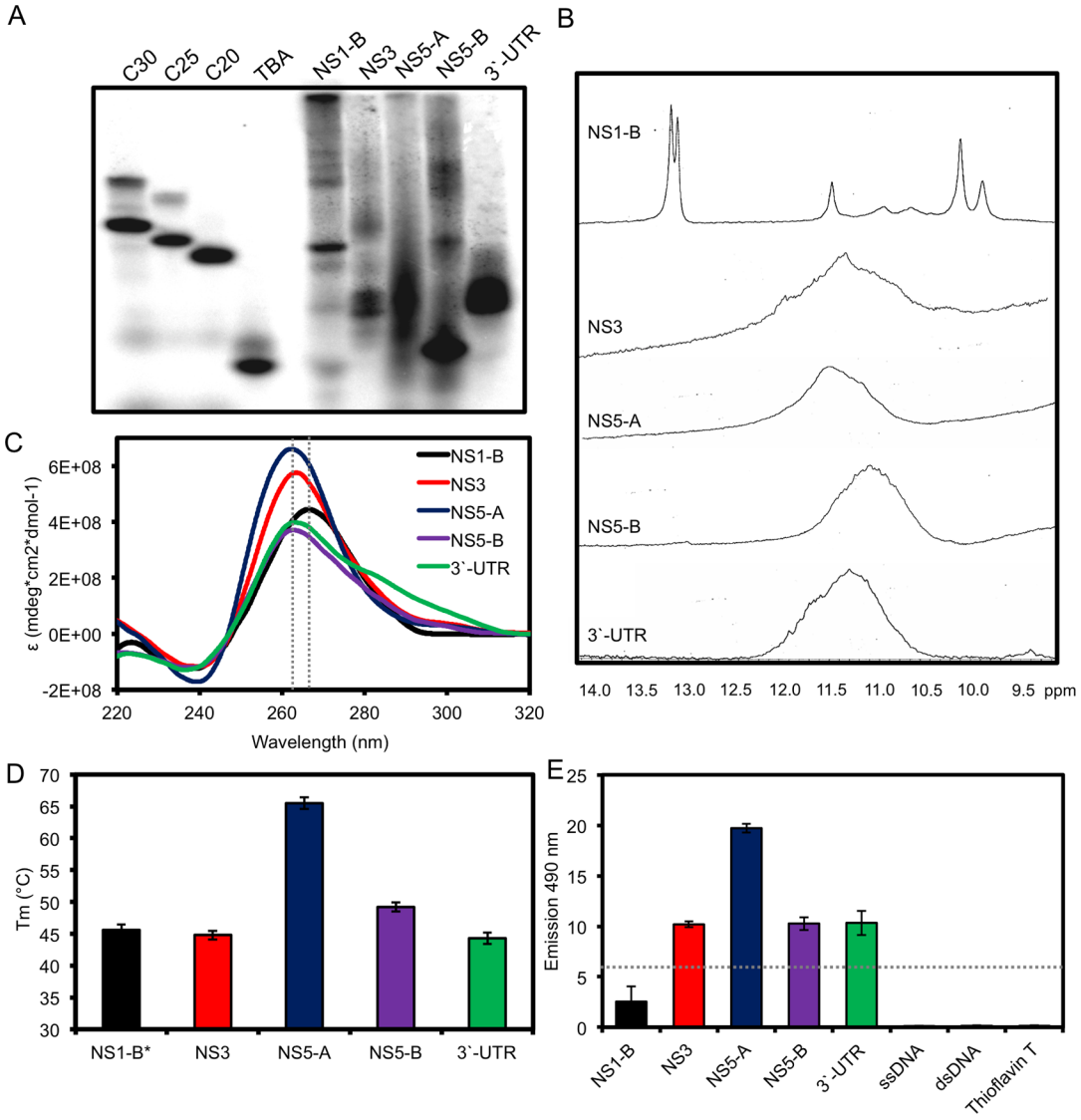


Figure 4. Characterization of selected conserved Zika viral PQSs. (A) Native gel electrophoresis of the PQSs. The sequences were compared to the single-stranded controls composed of 20, 25, and 30 nucleotide length homopolymers of 2′-deoxycytidine (C20, C25, and C30), as well as TBA, a well-established two-tetrad G-quadruplex. (B) ^{1}H NMR of each sequence studied. (C) CD spectra for the sequences. The dotted lines highlight the subtle difference in λ_{max} for the NS1-B sequence.[18] (D) T_m values measured for the sequences. * This value was obtained by monitoring the melting profile at 260 nm. (E) Thioflavin T fluorescence assay for each PQS compared to ssDNA, dsDNA, and thioflavin T controls. The dotted line represents the threshold for acceptable thioflavin T fluorescence enhancement to support G-quadruplex formation.[33]

loop sequences (i.e., those between G runs) can occur and still allow G-quadruplex formation. Determination of conservation of these necessary G runs was achieved by constructing sequence logos[29] of the PQS sites from each of the aligned sequences from the Flaviviridae family (Figure 3A for the NS3 and NS5-B sequences). The sequence logos are plotted alongside the Zika virus sequence, in which the critical Gs are underlined (Figure 3). From these logos, nearly complete retention of Gs required for G-quadruplex formation was observed in the NS3 and NS5-B sequences, whereas the variable sites depicted by smaller letters in the logo reside in the loop regions of the PQS. Furthermore, a nearly complete retention of necessary G-quadruplex-forming Gs was also observed in the PQS sites identified in the prM, E, and NS5-A sequences, whereas the PQS sites in NS1-A and NS1-B were not as strongly conserved across the Flaviviridae family, although they still showed the ability to form a G-quadruplex (Figure S5). Sequence logos generated from the Zika virus strain sequences found complete retention of the core Gs, and variations occurred only in loop regions (Figure 3). This final observation again supports the hypothesis that these PQS sites are being favorably selected.

The strong retention of the PQSs was even more surprising when genome variability across the family was examined. Specifically, in the Flaviviridae sequences analyzed, the percent G content ranged from 25.3 to 33.9%, and the percent similarity in sequence with respect to the Zika viral genome as the reference ranged from 42.3 to 70.3% (Figure S3). The large variation in total sequence similarity and distribution of G nucleotides between family members suggests that these PQSs are not appearing by random chance and adds additional support for their conservation. To further support that the PQSs in the Zika virus are selected by evolution and not appearing by random chance, the Zika viral genome was computationally randomized, and then the number of PQSs was counted (Figure S6). The randomization was conducted

DOI: 10.1021/acsinfecdis.6b00109
ACS Infect. Dis. 2016, 2, 674−681

two ways and 10 times each. In the first study, the single nucleotide content was held constant during the randomization, and significantly fewer PQSs were observed ($P = 7.0 \times 10^{-5}$) than in the native genome sequence. When the genome was randomized while maintaining the same dinucleotide content to maintain the same number of 5′-GG-3′ dinucleotides, there were again significantly fewer PQSs observed ($P = 5.3 \times 10^{-4}$) than in the native genome. These observations support the observation that the Zika virus has evolved to maintain these PQSs, and they are not appearing by random chance.

Flaviviruses replicate by first synthesizing the negative-sense strand from the positive-sense strand, which is then used as a template to synthesize more positive-sense strands for packaging into new viral capsids.[4] Another location for PQSs could be in the negative-sense strand; thus, we looked through the aligned positive-sense strands for four or more contiguous runs of two or more cytidine (C) nucleotides, because these would be complementary to PQSs in the negative-sense strand. This inspection failed to identify any PQS in the negative-sense strand. This observation identifies large asymmetry with respect to the PQS content between the two strands. The strand asymmetry for PQS sites likely results from the high G content relative to C content in the positive-sense strands (%G = 25.3−33.9%, %C = 19.2−24.8%; Figure S3). In conclusion, if G-quadruplex-binding drugs were administered to Zika virus-infected cells, they would target the positive-sense strand.

The PQS positions were inspected with respect to their location in the Zika viral genome (NCBI reference sequence NC_012532.1, Figure 2C).[30] The PQSs in the prM, E, NS1-A, NS3, and NS5-B sites were located in the interior of the coding region for each protein.[30] The NS5-A PQS was located on the very 5′ end of the coding sequence. Lastly, the PQS specific to Zika in the 3′-UTR was located 13 nucleotides from the very end of the genome. The 3′-terminal PQS is in a location critical for initial viral replication of the negative-sense strand.[4,9] Targeting the 3′-UTR PQS with drugs may represent an alternative approach to diminishing viral replication within the cell; moreover, this approach would be a new avenue for targeting viruses through the use of small-molecule drugs aimed at the RNA viral genome.

Previous studies have demonstrated that G-quadruplexes with short loops tend to be more stable (<7 nucleotides);[31] therefore, we studied the folding potential of the NS1-B, NS3, NS5-A, NS5-B, and 3′-UTR PQSs. If more than four runs of G exist in the conserved region, the sequence predicted to be the best to fold using QGRS mapper[25] was studied (Figure 2C, red sequences). The reason for this simplification is because additional G runs cause the structures to be more polymorphic and more challenging to characterize, but these runs may have important functions during oxidative stress, as we previously demonstrated in DNA.[32] Additionally, all strands studied had two-nucleotide overhangs on the 5′ and 3′ ends to ensure they remained in a more natural sequence context. The sequences were constructed by solid-phase synthesis and purified by HPLC prior to characterization. Initially, native gel electrophoresis was conducted on samples annealed under the analysis conditions and concentrations (20 mM lithium cacodylate, pH 7.4, 140 mM KCl, and 12 mM NaCl at 10 μM RNA). The gel analysis verified that the PQSs NS3, NS5-A, NS5-B, and 3′-UTR adopted >85% unimolecular G-quadruplexe folds (Figure 4A). The sequence NS1-B, as described below, failed to adopt a G-quadruplex. These claims are based on comparisons to poly-2′-deoxycytidine single-stranded controls and the thrombin binding aptamer (TBA) that is a well established two-tetrad DNA G-quadruplex.[18]

Next, to obtain suitable ^{1}H NMR spectra on each PQS, the concentrations were increased 30-fold to 300 μM. The NMR spectra were recorded for each sequence in slightly lower ionic strength (50 mM KCl buffered with 20 mM KP_i at pH 7.0, Figure 4B). In the NMR studies, we looked for diagnostic imino protons from 9.5 to 12.0 ppm indicating G-tetrad formation; in contrast, peaks between 13.0 and 14.0 ppm support Watson−Crick base pair formation and possible hairpin structures.[10] The first sequence inspected, NS1-B, produced three weak imino peaks for G-tetrads and strong peaks supporting Watson−Crick base pairs. This initial observation supports NS1-B as unlikely to form a G-quadruplex. In contrast to the first sequence studied, NS3, NS5-A, NS5-B, and the 3′-UTR PQS all produced imino profiles around 10.5−12.0 ppm, supporting formation of G-tetrads leading to G-quadruplex folds.

Circular dichroism (CD) spectroscopy is a routine method to gain an initial handle on the folded structure of a G-quadruplex.[34] In general, RNA G-quadruplexes only adopt parallel-stranded structures with all G nucleotides in the *anti* conformation.[16] The strong preference for *anti* G results from the 2′-OH of the ribose ring favoring the 2′-endo configuration causing the guanine heterocycle to favor the *anti* conformation in the nucleotides.[16] The CD spectra for NS3, NS5-A, NS5-B, and the 3′-UTR PQS all yielded a $\lambda_{max} \sim 260-262$ nm and a $\lambda_{min} = 240$ nm, supporting parallel-stranded G-quadruplex folds (Figure 4C). On the other hand, the NS1-B sequence gave a CD profile with a $\lambda_{max} = 268$ nm and a $\lambda_{min} = 240$ nm supporting a hairpin structure.[34] The native-gel analysis, NMR, and CD spectroscopy results combined support NS3, NS5-A, NS5-B, and the 3′-UTR Zika virus PQS as capable of folding to parallel-stranded G-quadruplexes.

The thermal stability of G-quadruplex folds is dependent on many parameters including primary sequence, cation identity, and cation concentration, to name a few.[13] Under the conditions we studied to closely match the cation ionic strength of a cell, the thermal stabilities (T_m) for the Zika virus PQS were measured by monitoring the temperature-dependent change in UV absorption at 295 nm.[13] The sequences that adopted G-quadruplex folds had T_m values ranging from 44.3 to 65.5 °C, supporting their ability to remain folded at physiological temperatures (Figures 4D and S7). Individually, the least stable were the NS3 and 3′-UTR sequences, with T_m values of 44.8 and 44.3 °C, respectively; the NS5-B sequence was slightly more stable with a T_m of 49.2 °C, and the NS5-A was most stable with a T_m of 65.5 °C. The T_m value for the NS1-B hairpin was monitored at 260 nm and found to have a stability of 45.6 °C that would remain folded at physiological temperatures. These secondary structures are even stable enough to remain intact at the extreme temperatures associated with viral-induced fevers (~40−41 °C).

In the last study, demonstration that the Zika virus PQSs adopting G-quadruplex folds can be targeted with a G-quadruplex-specific molecule was undertaken. A G-quadruplex-specific binding compound is the fluorophore thioflavin T.[33] Binding of the fluorophore with each PQS was determined from a fluorescence assay in which the fluorescence emission at 490 nm increased >60-fold upon binding to a G-quadruplex motif (Figure 4E, gray line). The fluorescence study found the Zika virus sequences to increase the thioflavin T fluorescence (Figures 4E and S8). The best folded G-quadruplexes on the

DOI: 10.1021/acsinfecdis.6b00109
ACS Infect. Dis. 2016, 2, 674−681

basis of the ^{1}H NMR, CD, and T_m experiments (NS3, NS5-A, NS5-B, and the 3′-UTR) all yielded a fluorescence enhancement of >60-fold characteristic of G-quadruplexes; in contrast, the NS1-A sequence that adopts a possible hairpin structure yielded only a 25-fold enhancement (Figure 4E). This final study identifies the ability to bind G-quadruplex folds in the Zika viral genome with G-quadruplex specific molecules.

The recent global spread of Zika virus is alarming because Zika infections in pregnant individuals can apparently cause microcephaly, whereas in others these infections induce Guillain−Barré syndrome that can be fatal.[2] Due to the rapid increase in Zika viral infections, there remain many mysteries surrounding this virus, one of which is detailed information concerning the viral genome structure. The present study inspected the Zika viral genome for PQS and found >60 possible sites in the genome capable of G-quadruplex formation (Figure S1). To address whether any of the PQSs have important roles, we aligned all Flaviviridae genomes from the NCBI database to identify seven conserved PQSs in their genomes. The strong conservation at sites such as NS3, NS5-A, and NS5-B supports evolutionary conservation of these sequence motifs, even if these sequences exist for coding a necessary region of the protein. Furthermore, analysis of the Zika viral strains found these sites to not change, and the strain sequence alignment identified a unique PQS near the very 3′-end of the Zika virus genome (Figure 3). In the 3′-UTR PQS specific to Zika, this sequence may be essential for other purposes such as viral replication.[9] The structural studies found four of the five sequences studied to adopt G-quadruplex folds. Interestingly, one PQS failed to adopt a G-quadruplex fold, highlighting how hard it is to predict whether sequences fold in a specific way. Also, this underscores the reason that many complementary structural methods are required to establish G-quadruplex folding. Moreover, all four PQSs that did fold provided binding sites for the G-quadruplex specific compound thioflavin T on the basis of a fluorescence study (Figure 4E).

With regard to Zika virus, vectors, reservoirs, amplifying hosts, and their potential to spread worldwide, what do we know and what should we investigate urgently? There are two sequences in the present sample worth urgent further inspection. First, the NS5-A PQS adopted the most stable parallel-stranded G-quadruplex with a T_m = 65.5 °C, and it is located at the very 5′ end of the coding region for the NS5 protein. The NS5 sequence codes for the critical RNA-dependent RNA polymerase essential for viral replication. Whether this sequence is essential or not for polyprotein cleavage is an interesting future question. If this fold is important, stabilizing it with a G-quadruplex-specific compound could interfere with viral protein synthesis. Second is the Zika-specific PQS near the very 3′ end of the genome. For the virus to replicate, it will need access to the 3′ end to allow proper replication of the genome. Many studies have demonstrated the power of binding G-quadruplexes with molecules specific for these structures to inhibit the advancement of polymerases on a template strand.[35] Therefore, targeting these PQS sites, particularly the 3′-UTR or the most stable NS5-A sequence, with G-quadruplex-binding molecules could have the potential for preventing Zika virus replication. Therefore, preliminary polymerase stop assays were conducted to determine if the NS5-A or 3′-UTR PQSs could stall polymerase bypass when folded to a G-quadruplex. The Zika viral RNA-dependent RNA polymerase is not available; thus, we used a commercial reverse transcriptase for these experiments. These studies identified NS5-A could stall a polymerase when folded, whereas the 3′-UTR sequence did not lead to significant stalling (Figure S9). Future studies to determine if adding G-quadruplex binding compounds could enhance polymerase stalling are warranted, especially while using a Zika viral RNA-dependent RNA polymerase.

Recently, a call to the scientific community was made to identify compounds that have progressed through clinical trials and could be repurposed to fight Zika virus.[36] There exist many qualified compounds that target G-quadruplexes that fit this requirement. *We strongly encourage any laboratory that has the necessary facilities and ability to conduct such studies to consider G-quadruplex binding compounds to counter the Zika virus.* If these types of compounds can inhibit Zika virus replication by targeting essential RNA G-quadruplexes, this will represent a new avenue for combating viral infections. Previous studies aimed at G-quadruplexes in viruses targeted either the DNA integrated genome of HIV or the 5′-UTR of essential viral mRNAs, as demonstrated with the Epstein−Barr virus.[19,20]

As a final note, the sequence alignment of the Flaviviridae genomes identified seven conserved PQS sites. Our observation of the strong conservation of the G nucleotides required for G-quadruplex formation through the whole family supports a hypothesis that these sequences are very important for the Flaviviridae family of viruses. A second interesting observation is with regard to the NS3 protein, very unique to this virus family, which possesses both protease and helicase domains. The most interesting observation is that the helicase domain has homology to DEAH-box helicases. These types of helicases are essential for unwinding G-quadruplex folds in RNA.[37] The utility of this type of helicase by flaviviruses adds additional support for the importance of these genomic G-quadruplex folds. Whether these G-quadruplexes are important for maintaining global genome structure or serve another purpose remains an open question. A last interesting observation is that the 3′-UTRs of LINE-1 retrotransposons also harbor a conserved PQS similar to the one we have identified in the Zika virus,[38] and whether there exists an evolutionary reason for these PQSs provides exciting future prospects. Further studies are anticipated to gain a better understanding of these PQSs in flaviviruses and more specifically the Zika virus.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acsinfecdis.6b00109.

Predicted PQSs, sequence alignments, alignment statistics, experimental procedures, CD spectra, T_m curves, NMR spectra, and thioflavin T fluorescence spectra (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*(C.J.B.) Phone: (801) 585-7290. E-mail: burrows@chem.utah.edu.

*(A.M.F.) E-mail: afleming@chem.utah.edu.

Notes

The authors declare no competing financial interest.

DOI: 10.1021/acsinfecdis.6b00109
ACS Infect. Dis. 2016, 2, 674−681

■ ACKNOWLEDGMENTS

We are deeply appreciative of the National Cancer Institute (R01 CA090689) for partial financial support of this work. We thank Dr. Peter Flynn (University of Utah) for help with NMR experiments. The oligonucleotides were provided by the DNA/Peptide core facility at the University of Utah, which is supported in part by NCI Cancer Support Grant (P30 CA042014).

■ ABBREVIATIONS

C, capsid protein; CD, circular dichroism; DENV, dengue virus; DONV, Donggang virus; E, envelope protein; EBNA1, Epstein−Barr virus-encoded nuclear antigen 1; G, guanosine; HIV, human immunodeficiency virus; JE, Japanese encephalitis; LGTV, Langat virus; NCBI, National Center for Biotechnology Information; NS1, nonstructural protein 1; NS2A, nonstructural protein 2A; NS2B, nonstructural protein 2B; NS3, nonstructural protein 3; NS4A, nonstructural protein 4A; NS4B, nonstructural protein 4B; NS5, nonstructural protein 5; PQS, potential G-quadruplex sequence; prM, precursor membrane protein; SPOV, Spondweni virus; SLEV, St. Louis encephalitis virus; TBA, thrombin binding aptamer; TBE, tickborne encephalitis virus; T_m, thermal melting temperature; UTR, untranslated region; WNV, West Nile virus; WHO, World Health Organization; ZIKV, Zika virus

■ REFERENCES

(1) World Health Organization (WHO). (2016) *Zika Strategic Response Framework and Joint Operations Plan (January−June 2016)*.
(2) Vorou, R. (2016) Zika virus, vectors, reservoirs, amplifying hosts, and their potential to spread worldwide: what we know and what we should investigate urgently. *Int. J. Infect. Dis. 48*, 85−90.
(3) Chambers, T. J., Hahn, C. S., Galler, R., and Rice, C. M. (1990) *Flavivirus* genome organization, expression, and replication. *Annu. Rev. Microbiol. 44*, 649−688.
(4) Suthar, M. S., Diamond, M. S., and Gale, M., Jr. (2013) West Nile virus infection and immunity. *Nat. Rev. Microbiol. 11*, 115−128.
(5) Sirohi, D., Chen, Z., Sun, L., Klose, T., Pierson, T. C., Rossmann, M. G., and Kuhn, R. J. (2016) The 3.8 Å resolution cryo-EM structure of Zika virus. *Science 352*, 467−470.
(6) Song, H., Qi, J., Haywood, J., Shi, Y., and Gao, G. F. (2016) Zika virus NS1 structure reveals diversity of electrostatic surfaces among *flaviviruses*. *Nat. Struct. Mol. Biol. 23*, 456−458.
(7) Jain, R., Coloma, J., Garcia-Sastre, A., and Aggarwal, A. K. (2016) Structure of the NS3 helicase from Zika virus. *Nat. Struct. Mol. Biol. 23*, 752 10.1038/nsmb.3258.
(8) Lei, J., Hansen, G., Nitsche, C., Klein, C. D., Zhang, L., and Hilgenfeld, R. (2016) Crystal structure of Zika virus NS2B-NS3 protease in complex with a boronate inhibitor. *Science 353*, 503−505.
(9) Villordo, S. M., Carballeda, J. M., Filomatori, C. V., and Gamarnik, A. V. (2016) RNA structure duplications and *Flavivirus* host adaptation. *Trends Microbiol. 24*, 270−283.
(10) Patel, D. J., Phan, A. T., and Kuryavyi, V. (2007) Human telomere, oncogenic promoter and 5′-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics. *Nucleic Acids Res. 35*, 7429−7455.
(11) Biffi, G., Tannahill, D., McCafferty, J., and Balasubramanian, S. (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem. 5*, 182−186.
(12) Henderson, A., Wu, Y., Huang, Y. C., Chavez, E. A., Platt, J., Johnson, F. B., Brosh, R. M., Jr., Sen, D., and Lansdorp, P. M. (2014) Detection of G-quadruplex DNA in mammalian cells. *Nucleic Acids Res. 42*, 860−869.
(13) Gray, R. D., and Chaires, J. B. (2008) Kinetics and mechanism of K^+- and Na^+-induced folding of models of human telomeric DNA into G-quadruplex structures. *Nucleic Acids Res. 36*, 4191−4203.
(14) Balasubramanian, S., Hurley, L. H., and Neidle, S. (2011) Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nat. Rev. Drug Discovery 10*, 261−275.
(15) Rhodes, D., and Lipps, H. J. (2015) G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res. 43*, 8627−8637.
(16) Zhang, D. H., Fujimoto, T., Saxena, S., Yu, H. Q., Miyoshi, D., and Sugimoto, N. (2010) Monomorphic RNA G-quadruplex and polymorphic DNA G-quadruplex structures responding to cellular environmental factors. *Biochemistry 49*, 4554−4563.
(17) Joachimi, A., Benz, A., and Hartig, J. S. (2009) A comparison of DNA and RNA quadruplex structures and stabilities. *Bioorg. Med. Chem. 17*, 6811−6815.
(18) Sacca, B., Lacroix, L., and Mergny, J. L. (2005) The effect of chemical modifications on the thermal stability of different G-quadruplex-forming oligonucleotides. *Nucleic Acids Res. 33*, 1182−1192.
(19) Perrone, R., Butovskaya, E., Daelemans, D., Palu, 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.
(20) Murat, P., Zhong, J., Lekieffre, L., Cowieson, N. P., Clancy, J. L., Preiss, T., Balasubramanian, S., Khanna, R., and Tellam, J. (2014) G-quadruplexes regulate Epstein-Barr virus-encoded nuclear antigen 1 mRNA translation. *Nat. Chem. Biol. 10*, 358−364.
(21) Horsburgh, B. C., Kollmus, H., Hauser, H., and Coen, D. M. (1996) Translational recoding induced by G-rich mRNA sequences that form unusual structures. *Cell 86*, 949−959.
(22) Tuesuwan, B., Kern, J. T., Thomas, P. W., Rodriguez, M., Li, J., David, W. M., and Kerwin, S. M. (2008) Simian virus 40 large T-antigen G-quadruplex DNA helicase inhibition by G-quadruplex DNA-interactive agents. *Biochemistry 47*, 1896−1909.
(23) Wang, S. R., Min, Y. Q., Wang, J. Q., Liu, C. X., Fu, B. S., Wu, F., Wu, L. Y., Qiao, Z. X., Song, Y. Y., Xu, G. H., Wu, Z. G., Huang, G., Peng, N. F., Huang, R., Mao, W. X., Peng, S., Chen, Y. Q., Zhu, Y., Tian, T., Zhang, X. L., and Zhou, X. (2016) A highly conserved G-rich consensus sequence in hepatitis C virus core gene represents a new anti-hepatitis C target. *Sci. Adv. 2*, e1501535.
(24) Huppert, J. L., and Balasubramanian, S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res. 33*, 2908−2916.
(25) 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.
(26) Bedrat, A., Lacroix, L., and Mergny, J. L. (2016) Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res. 44*, 1746−1759.
(27) Kumar, S., Stecher, G., and Tamura, K. (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol. 33*, 1870−1874.
(28) Wright, E. S. (2015) DECIPHER: harnessing local sequence context to improve protein multiple sequence alignment. *BMC Bioinf. 16*, 1−14.
(29) Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) WebLogo: a sequence logo generator. *Genome Res. 14*, 1188−1190.
(30) Kuno, G., and Chang, G. J. (2007) Full-length sequencing and genomic characterization of Bagaza, Kedougou, and Zika viruses. *Arch. Virol. 152*, 687−696.
(31) Guedin, A., Gros, J., Alberti, P., and Mergny, J. L. (2010) How long is too long? Effects of loop size on G-quadruplex stability. *Nucleic Acids Res. 38*, 7858−7868.
(32) Fleming, A. M., Zhou, J., Wallace, S. S., and Burrows, C. J. (2015) A role for the fifth G-track in G-quadruplex forming oncogene promoter sequences during oxidative stress: do these "spare tires" have an evolved function? *ACS Cent. Sci. 1*, 226−233.
(33) Renaud de la Faverie, A., Guedin, A., Bedrat, A., Yatsunyk, L. A., and Mergny, J. L. (2014) Thioflavin T as a fluorescence light-up probe for G4 formation. *Nucleic Acids Res. 42*, e65.
(34) Karsisiotis, A. I., Hessari, N. M., Novellino, E., Spada, G. P., Randazzo, A., and Webba da Silva, M. (2011) Topological character-

DOI: 10.1021/acsinfecdis.6b00109
ACS Infect. Dis. 2016, 2, 674−681

ization of nucleic acid G-quadruplexes by UV absorption and circular dichroism. *Angew. Chem., Int. Ed. 50*, 10645−10648.
(35) Baumann, P., Sun, D., and Hurley, L. (2010) Biochemical techniques for the characterization of G-quadruplex structures: EMSA, DMS footprinting, and DNA polymerase stop assay. *Methods Mol. Biol. 608*, 65−79.
(36) Halford, B. (2016) Scientists scramble to develop tools, treatments for Zika virus. *Chem. Eng. News 94* (8), 33−36.
(37) Chen, M. C., Murat, P., Abecassis, K., Ferre-D'Amare, A. R., and Balasubramanian, S. (2015) Insights into the mechanism of a G-quadruplex-unwinding DEAH-box helicase. *Nucleic Acids Res. 43*, 2223−2231.
(38) Honda, T. (2016) Links between human LINE-1 retrotransposons and hepatitis virus-related hepatocellular carcinoma. *Front. Chem. 4*, DOI: 10.3389/fchem.2016.00021.

DOI: 10.1021/acsinfecdis.6b00109
ACS Infect. Dis. 2016, 2, 674−681