ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

Ion-Dependent Conformational Plasticity of Telomeric G-Hairpins and G-Quadruplexes

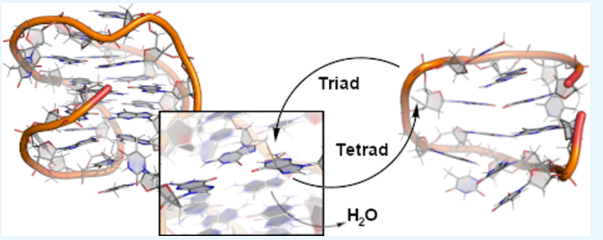
Alexa M. Salsbury, Haley M. Michel, and Justin A. Lemkul*

Cite This: *ACS Omega* 2022, 7, 23368−23379

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Telomeric DNA is guanine-rich and can adopt structures such as G-quadruplexes (GQs) and G-hairpins. Telomeric GQs influence genome stability and telomerase activity, making understanding of enzyme−GQ interactions and dynamics important for potential drug design. GQs have a characteristic tetrad core, which is connected by loop regions. Within this architecture are G-hairpins, fold-back motifs that are thought to represent the first intermediate in GQ folding. To better understand the relationship between G-hairpin motifs and GQs, we performed polarizable simulations of a two-tetrad telomeric GQ and an isolated *SC11* telomeric G-hairpin. The telomeric GQ contains a G-triad, which functions as part of the tetrad core or linker regions, depending on local conformational change. This triad and another motif below the tetrad core frequently bound ions and may represent druggable sites. Further, we observed the unbiased formation of a G-triad and a G-tetrad in simulations of the *SC11* G-hairpin and found that cations can be partially hydrated while facilitating the formation of these motifs. Finally, we demonstrated that K^+ ions form specific interactions with guanine bases, while Na^+ ions interact nonspecifically with bases in the structure. Together, these simulations provide new insights into the influence of ions on GQs, G-hairpins, and G-triad motifs.



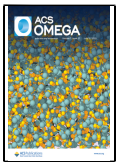
INTRODUCTION

Telomeres are specialized nucleoprotein structures at the terminal ends of eukaryotic chromosomes that help maintain genomic integrity and have been identified as regulators in disease states.[1−4] Human telomeric DNA is composed of tandem repeats of the hexanucleotide $(TTAGGG)_n$, which is well conserved and found at the chromosome ends of most eukaryotes.[5] This primary sequence is guanine-rich, contains guanine triplets, and therefore has the ability to form G-quadruplexes (GQs) and G-hairpins, dynamic and topologically diverse nucleic acid structures that interact with telomerase, an enzyme that lengthens chromosomal DNA and is overactive in disease states.[6−9] GQs form when guanine triplets self-associate, creating stacked tetrads stabilized by cations coordinated to Hoogsteen hydrogen-bonded guanines.[10] GQs comprise these stacked tetrads, called the tetrad core, and linker regions. G-hairpins, fold-back motifs stabilized by G•G base pairing, can connect the guanine sequences of the tetrad core and linker regions in GQs. Further, G-hairpins are believed to be involved in folding and structural transitions and have recently been shown to stably form in native DNA sequences.[7,11−17]

GQ folding is complex and involves the formation of diverse intermediates that help determine the overall topology and dynamics of the folded structure.[18] While several GQ folding intermediates have been proposed and studied, the mechanism of folding is not well characterized and likely differs depending on primary sequences and interactions with other biomolecules.[18,19] Computational and spectroscopic studies have provided descriptions for a simplified, sequential GQ folding pathway with G-hairpin and G-triplex intermediates.[11−17] Broadly, single-stranded DNA is thought to collapse into a G-hairpin before coordinating cations and transitioning into long-lived intermediates like G-triplexes, which undergo slow conformational rearrangements to form the expected G-tetrad geometry and ultimately a folded GQ.[11−17] Though evidence exists that G-hairpins are crucial, on-pathway GQ folding intermediates, understanding the dynamic interplay between these structures requires further investigation, and molecular dynamic (MD) simulations are well suited to doing so.

MD simulations have made valuable contributions to our understanding of guanine-rich nucleic acids and are an important tool for studying their structural dynamics. To better understand GQ dynamics, simulations provided the first description of the rigid tetrad core, described the reliance on coordinated core ions,[20] identified possible intermediates,[21] and proposed folding pathways.[22] Recently, our group has employed the Drude polarizable force field (FF)[23,24] to

Received: March 16, 2022
Accepted: May 19, 2022
Published: June 29, 2022



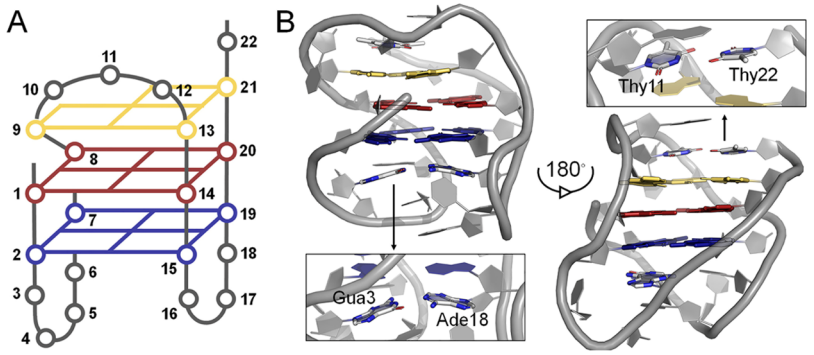


Figure 1. Structure of the two-tetrad telomeric GQ. (A) Schematic of all residues in the GQ. The guanine bases of the core are colored by triad (yellow) and tetrad (1—red and 2—blue), while loop and linker residues are gray. (B) Full structure taken from model 1 of the NMR ensemble deposited in PDB entry 2KF8. Important noncanonical base pairs (Gua3•Ade18 and Thy11•Thy22) are highlighted within the structure.

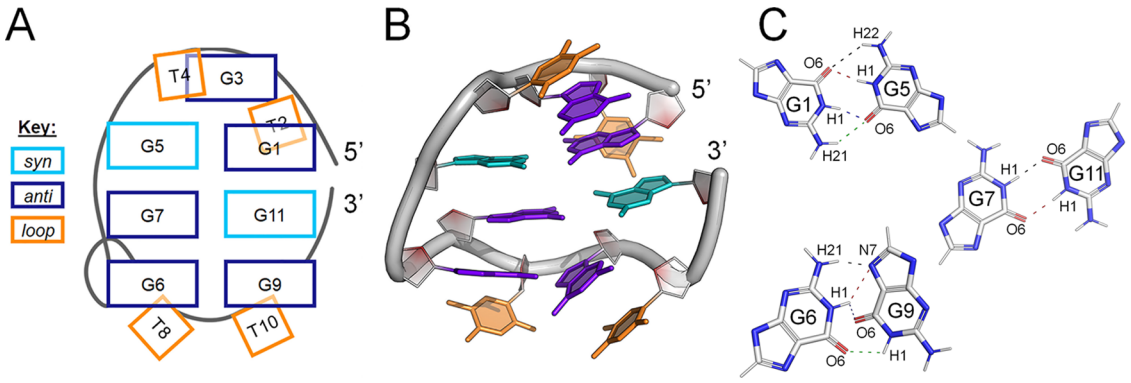


Figure 2. Structure of the telomeric *SC11* G-hairpin. (A) Schematic and (B) structure showing the folded topology and the glycosidic bond classification of the nuclear magnetic resonance (NMR) model 1 deposited in PDB entry 5M1W. (C) Noncanonical base pairs (Gua1•Gua5, Gua6•Gua9, and Gua7•Gua11) stabilizing the G-hairpin. Atoms are labeled exactly as they are assigned in the deposited NMR structure, and we note the assignment of "H21" and "H22" to denote equivalent atoms in different bases. Throughout our discussion, we retain the nomenclature assigned in the PDB structure to avoid confusion despite this discrepancy.

investigate DNA and RNA GQs[25−29] because nonpolarizable FFs are thought to provide insufficient descriptions of ion coordination and ion−ion energetics in the tetrad core.[12,30,31] Our work with the Drude FF to date has shown improved descriptions of GQ−ion interactions and that nucleobase dipole moments vary as a function of ion type and coordination, hydration, and GQ primary and secondary structure. Further, these properties can influence GQ conformational sampling and nonbonded interaction energies in the tetrad core.[25−28] These studies suggest that the inclusion of electronic polarization is critical in MD simulations of nucleic acid structures strongly influenced by ion interactions, such as GQs and G-hairpins. Two such structures, the two-tetrad telomeric GQ[32] and the *SC11* G-hairpin,[7] are the focus of this study.

The two-tetrad telomeric GQ (PDB entry 2KF8, Figure 1) is adopted by four-repeat human telomeric sequences in K^+ solution.[32] It is a structurally unique, basket-type GQ with a guanine triad (G9•G13•G21) above the tetrad core (tetrad 1: G1•G8•G14•G20 and tetrad 2: G2•G7•G15•G19). Two noncanonical base pairs flank the incomplete core, T11•T22 above the triad and G3•A18 below tetrad 2, which are thought to confer extra stability to the structure. In fact, this GQ was found to be more stable than similar three-tetrad GQs, challenging the idea that the Hoogsteen hydrogen bonding is the greatest contribution to GQ stability.[32]

Solving the structure of the thermodynamically stable, isolated *SC11* G-hairpin also challenged pre-existing understanding of guanine-rich DNA (g-DNA).[7] Before this structure was solved, G-hairpins were only characterized using DNA origami[15] and MD simulations.[33] The sequence of the *SC11* G-hairpin, 5′-d(GTGTGGGTGTG)-3′, comprises a protein-binding motif in *Saccharomyces cerevisiae* and forms a mixed parallel/antiparallel fold-back structure (Figure 2).[7] This G-hairpin resembles Hoogsteen stem-loops found in GQs as it is stabilized by three Gua•Gua base pairs (G1•G5, G7•G11, and G6•G9). However, the geometries of the base pairs in this structure (Figure 2C) are distinct from those found in tetrad guanines, reflecting the diversity of g-DNA.

Here, we applied the Drude-2017 FF[23,24] to study the dynamics of the telomeric two-tetrad GQ and *SC11* G-hairpins in solutions of KCl and NaCl. These simulations build on our previous GQ studies and, to our knowledge, represent the first time the Drude FF has been employed to study an isolated G-hairpin. Together, these simulations allow us to compare unique g-DNA structures and the effects of varying cation types on their conformational sampling and energetics and provide atomistic detail on base pairing, nucleobase polarization, and ion sampling around different motifs in non-canonical nucleic acids. Such outcomes have the potential to expand our understanding of g-DNA dynamics and GQ folding and can be used to inform novel drug design.

https://doi.org/10.1021/acsomega.2c01600
ACS Omega 2022, 7, 23368−23379

METHODS

System Construction. The starting structure for simulations of the telomeric two-tetrad GQ was taken from the first model of the NMR ensemble from PDB entry 2KF8 (Figure 1).[32] Systems were solvated in ∼150 mM KCl and NaCl solutions, including one core ion added between tetrads 1 and 2 in bipyramidal antiprismatic coordination. This ion was added by calculating the average coordinates of the carbonyl oxygen (O6) atoms of the guanine bases of tetrad 1 (G1•G8•G14•G20) and tetrad 2 (G2•G7•G15•G19) using the CHARMM program[34] and placing the ion there, resulting in canonical bipyramidal antiprismatic coordination.[34]

The first simulations of the *SC11* G-hairpin were prepared using model 1 of the NMR ensemble from PDB entry 5M1W (Figure 2)[7] solvated in a 150 mM concentration of either NaCl or KCl. Throughout the NaCl simulations, large structural rearrangements were observed, revealing possible alternate states in the *SC11* G-hairpin conformational ensemble: G-hairpin state 2 (Supporting Information, Figure S1), stacked triad/tetrad (Supporting Information, Figure S2), and two unfolded conformations (Supporting Information, Figure S3). To better understand the *SC11* conformational ensemble, we extracted the dominant structure of each motif using root-mean-square deviation (RMSD)-based clustering of the pooled trajectories. While these states were only present in NaCl replicates 2 and 3, additional systems were set up in both KCl and NaCl to better understand the influence of the cation type on structure.

Each system described above was constructed, minimized, and equilibrated using the CHARMM36 (C36) FF for nucleic acids[35−37] before being converted to the Drude-2017 nucleic acid FF.[23,24] The nucleic acid structures were centered in cubic boxes with a minimum box−solute distance of 10 Å for the telomeric two-tetrad GQ and 20 Å for the *SC11* G-hairpin. The larger box size used in the *SC11* G-hairpin systems was designed to avoid minimum image violations in anticipation of large conformational changes. These boxes were filled with the SWM4-NDP[38] water and a total salt concentration of ∼150 mM, including neutralizing counterions. Current Drude ion and nonbonded parameters for nucleic acid−ion interactions were applied.[39,40] In total, 18 independent simulations were performed and are listed in Table 1.

MD Simulations. Energy minimization, equilibration, and production runs were performed as described in our previous GQ simulation studies.[25−28] Drude oscillator positions were relaxed with steepest descent minimization and adopted-basis Newton−Raphson energy minimization in CHARMM. Then, *NPT* equilibration was carried out for 1 ns at 298 K and 1 atm by extended Lagrangian integration,[41] implemented in NAMD as Langevin dynamics.[41] Water and bulk ions were unrestrained during equilibration. The real atoms in the system were coupled to a thermostat to maintain temperature at 298 K with a friction coefficient of 5 ps^{-1}, and Drude oscillators were coupled to a low-temperature relative thermostat at 1 K with a friction coefficient of 20 ps^{-1}. Periodic boundary conditions were applied in all spatial dimensions. The short-range van der Waals potential was switched to zero from 10 to 12 Å. All bonds to hydrogen atoms were constrained using the SHAKE algorithm.[42] The water molecules were kept rigid with SETTLE,[43] allowing an integration time step of 1 fs. A "hard wall" constraint[44] was also applied to allow a maximum Drude-atom bond length of 0.2 Å and avoid polarization catastrophe.

Table 1. List of the Contents and Sizes of All Simulation Systems

structure source	solution	replicates	box size (Å)	system notes
2KF8	∼150 mM KCl	3 × 1 μs	50	telomeric two-tetrad GQ (Figure 1)
2KF8	∼150 mM NaCl	3 × 1 μs	50	telomeric two-tetrad GQ (Figure 1)
5M1W	∼150 mM KCl	3 × 2 μs	62	isolated G-hairpin (Figure 2)
5M1W	∼150 mM NaCl	3 × 2 μs	62	isolated G-hairpin (Figure 2)
SC11 snapshot	∼150 mM KCl	1 × 500 ns	72	starting in G-hairpin state 2 (Figure S1)
SC11 snapshot	∼150 mM NaCl	1 × 500 ns	72	starting in G-hairpin state 2 (Figure S1)
SC11 snapshot	∼150 mM KCl	1 × 500 ns	72	starting in the triad/tetrad state (Figure S2)
SC11 snapshot	∼150 mM NaCl	1 × 1 μs	72	starting in the triad/tetrad state (Figure S2)
SC11 snapshot	∼150 mM KCl	2 × 1 μs	62	starting in the unfolded state (Figure S3)

Equilibration was carried out for 1 ns, and independent replicates were generated by applying random velocities at the outset of each equilibration simulation. Unrestrained simulations were then performed on equilibrated systems using OpenMM.[45,46] The *NPT* ensemble was maintained, and the Monte Carlo barostat in OpenMM was used to isotropically regulate pressure with box scaling attempted every 25 integration steps. Simulation analysis was performed using CHARMM[34] and R programs,[47] as described below.

RMSD-Based Clustering. The CHARMM clustering function[48−50] was employed to perform clustering based on RMSD from the starting structures. Clusters were generated by separating frames with a self-defined maximum RMSD radius. Since the two-tetrad telomeric GQ did not show large conformational transitions, replicates of each system (KCl and NaCl) were pooled together and conformations were separated with a maximum radius of 1 Å. Clusters of the ion binding pockets within the two-tetrad telomeric GQ had a maximum radius of 2 Å and considered the deviation of Gua3, Thy5, Thy17, and Ade18. In the *SC11* G-hairpin simulations, we observed large conformational transitions, which varied by the ion type and replicate. To gain an understanding of the conformational ensemble in each, we first pooled together replicates of each system and separated conformations with a maximum radius of 1 Å. To further understand how the conformations of NaCl replicates 2 and 3 changed over time, we determined the dominant cluster for each interval of 250 ns.

Ion Sampling. To characterize ion sampling in our simulations, we tracked ions ≤3.5 Å from nonhydrogen atoms in each base and the sugar-phosphate backbone in the two-tetrad telomeric GQ and *SC11* G-hairpin. The ion counts for each nucleotide were averaged and plotted as heat maps. We also generated ion occupancy maps for the two-tetrad telomeric GQ system. These maps were generated by decomposing the system volume into discrete, 1 $Å^3$ volume elements (voxels). All frames in the trajectories were analyzed, and the location of each K^+ ion was assigned to the nearest voxel. To visualize the K^+ occupancy maps, an isosurface cutoff of ≥1% was chosen, indicating that a voxel must be occupied by an ion for at least 1% of the frames. To perform this

https://doi.org/10.1021/acsomega.2c01600
ACS Omega 2022, 7, 23368−23379

analysis, the three trajectories of each system were pooled, such that the resulting occupancy maps reflect ion sampling in the entire simulation ensemble.

RESULTS AND DISCUSSION

Noncanonical nucleic acid structures are dynamic and complex and, as such, serve as stringent tests of nucleic acid FFs. We have recently employed the Drude polarizable FF to probe the dynamics of different DNA and RNA GQs, demonstrating that the polarizable model outperforms C36, its additive counterpart, in modeling ion−water and ion−nucleic acid interactions as well as tetrad core stability.[25−28] As such, we performed simulations of the two-tetrad telomeric GQ and the *SC11* G-hairpin, noncanonical structures known to be governed by interactions with water and ions, with the Drude polarizable FF. Given our previous findings regarding the low quality of C36 in simulating GQs, we did not perform the nonpolarizable simulations of either of these structures.

Conformational Dynamics of the Two-Tetrad Telomeric GQ. To characterize the dynamics of the two-tetrad telomeric GQ, we analyzed overall and per-nucleotide RMSD, core hydrogen bonding, and per-nucleotide root-mean-square fluctuation (RMSF) and assessed dominant geometries via RMSD-based clustering. RMSD and core hydrogen bonding analyses showed greater deviation from the starting structure in solutions of NaCl than in KCl (Supporting Information, Figure S4). While tracing core hydrogen bonding over time showed some instances of hydrogen bond disruption (particularly in NaCl), in most cases, these interactions were restored over time. Considering both low RMSD and core hydrogen bonding deviation, the GQ structure was stably modeled in both solutions. Per-nucleotide RMSD and RMSF values were similar in KCl and NaCl, demonstrating that core atoms were more rigid than those of linker regions and that the cation type does not substantially alter core stability or flexibility (Supporting Information, Figures S5 and S6). Interestingly, Gua9 and Gua13 in the triad and Gua14 in tetrad 1 showed high deviation. Gua9 had the highest RMSD in both KCl and NaCl systems but low RMSF, suggesting that the nucleotide adopted a different conformation compared to the NMR model and did not vary from that position over time. Gua13 and Gua14 had high RMSD and low RMSF, suggesting that this region can adopt a different conformation compared to the experimental structure. Ion binding may contribute to this conformation and will be discussed below.

To better understand the observed structural deviations, we plotted backbone dihedral angle distributions for the α, ε, and ζ torsions and compared these against the values in the NMR ensemble. For most of these dihedrals, the NMR ensemble exhibits wide ranges of values (Supporting Information, Figure S7), suggesting either difficulty in unambiguously assigning dihedral values or reflecting a flexible structure with a heterogeneous ensemble. All distributions obtained from the MD simulations indicate that the simulations sampled dihedral conformers consistent with the experimental ensemble, with the largest deviations in Gua9 (α sampling) and Gua14 (α and ζ sampling). Further, these distributions show that dihedral sampling can differ as a function of cation type (KCl vs NaCl), though these differences were modest.

Per-nucleotide RMSF analysis showed that linker nucleotides were flexible, while guanine nucleotides in the triad and tetrads were rigid, except for Gua13 in the triad, which was more flexible than other core guanines in both KCl and NaCl solutions. The most noticeable difference between the GQ in the presence of KCl and that in NaCl was the sampling of the 3′-terminal Thy22, which was more flexible in solutions of KCl because the Thy11•Thy22 base pair broke in these simulations. Frequent interactions with K^+ ions were observed in these instances. In the NMR ensemble, five models included paired Thy11•Thy22; however, these base-pairing geometries differed and Thy11•Thy22 are unpaired in the remaining five models. Thus, the flexibility observed in our simulations agrees with the experimental ensemble. In contrast, when simulating this structure in NaCl solution, the Thy11•Thy22 base pair rarely broke (it was disrupted for only ∼25 ns of a total 3 μs sampling time), suggesting that Na^+ does not perturb this interaction in the way that K^+ does.

To better characterize the relevant substates of the simulation ensembles, we performed RMSD-based clustering of all heavy atoms in the system (Supporting Information, Figure S8). These clusters demonstrate strong overall agreement with the NMR structure and show that in KCl, the central structures of each cluster deviated less from the experimental structure than those produced in simulations with NaCl. Further, the clusters show the flexibility of linker regions and the Thy11•Thy22 base pair in the simulations. Linker nucleotides Gua3, Thy5, Ade6, Thy17, and Ade18 deviated from their starting positions, which we further characterized by RMSD-based clustering specific to these nucleotides (Supporting Information, Figure S9). This change from the starting structure created an ion binding pocket in the vicinity of the Gua3•Ade18 base pair. NMR data show that Gua3, Ade6, and Ade18 adopt a planar arrangement below tetrad 2 that is reminiscent of a GCA triad[32,51] and that Thy5 and Thy17 base stack below this motif to help preserve the structure (Supporting Information, Figure S9A,B). The planar configuration of Gua3•Ade6•Ade18 was maintained throughout the simulations, but the overall structure of this motif deviated over time. Gua3 and Ade18 flexed downward toward Thy5 and Thy17, which did not remain stacked, and this small rearrangement created a pocket that became amenable to ion binding. The pocket generally consisted of Gua3 O6, Thy5 O2 and O4, Thy17 O4, and Ade18 N6 (Supporting Information, Figure S9C,D). In NaCl simulations, Thy17 was more flexible (Supporting Information, Figure S6), enabling it to recruit ions from bulk solution, similar to thymine-assisted bulk ion binding in previous simulations of the *c-kit* GQs.[25,26] We discuss these ion interactions in detail in the next section.

Ion Sampling around the Two-Tetrad Telomeric GQ. To understand and compare ion sampling in our systems, we performed ion interaction analysis on the two-tetrad telomeric GQ. While only one ion was initially placed in the core of this GQ, occupying core binding site 1, three core ion binding sites persisted throughout the simulations (Figure 4A,B). Ions quickly bound and coordinated between the triad and tetrad 1, a location we call core binding site 2, occupying this location 99% of the time in both KCl and NaCl solutions. These ions desolvated to bind to the core and did not exchange with bulk ions after the initial coordination event (Supporting Information, Video S1). Bulk ions also quickly and consistently occupied the area above the triad and below the Thy11•Thy22 base pair, which we call core binding site 3. Ions occupying core binding site 3 and the ion binding pocket below tetrad 2 were more dynamic, exchanging multiple times throughout the simulations (Supporting Information, Video S2). In core binding site 3, bulk K^+ ions coordinated linearly with the other

https://doi.org/10.1021/acsomega.2c01600
ACS Omega 2022, 7, 23368−23379

two core ions and exchanged with other bulk ions between 10 and 13 times within 1 μs. In the NaCl simulations, bulk Na^+ bound and exchanged 1−7 times in 1 μs. In the ion binding pocket, both K^+ and Na^+ ions interacted with Gua3 O6, Thy5 O2 and O4, Thy17 O4, and Ade18 N6 and exchanged with bulk ions frequently (as many as to 40 times in 1 μs). The high rate of exchange in these ion binding sites might be partially due to their coordination by linker nucleotides that are more flexible than core guanines. However, it is likely that ion−water interactions contribute to the rate of exchange (discussed below).

We further investigated ion sampling in our simulations by tracing ion interactions with nucleobase and sugar-phosphate backbone atoms throughout the simulations (Figure 3C,

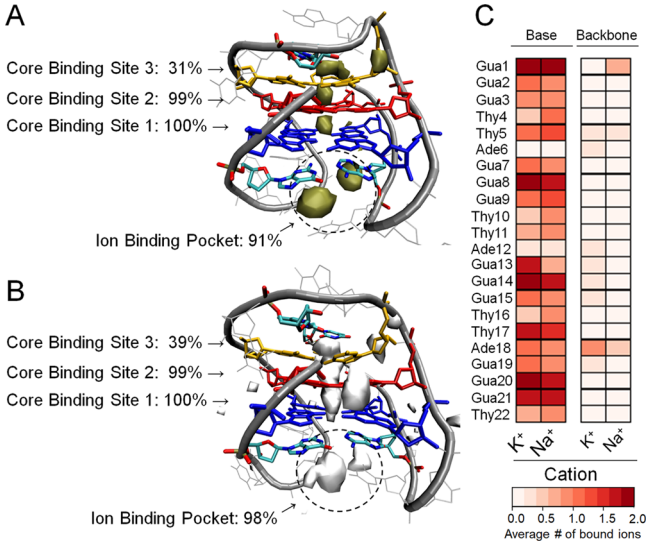


Figure 3. Ion−DNA interactions in the two-tetrad GQ. (A) K^+ and (B) Na^+ occupancy maps show ion sampling around the GQ at an occupancy threshold of ≥1%. The percentages shown indicate the persistence of each ion at that location throughout the three replicate simulations. (C) Heat map that shows the average number of ions within 3.5 Å of nonhydrogen atoms in GQ bases (left) and the sugar-phosphate backbone (right) in KCl and NaCl.

Supporting Information, Table S2). In doing so, we observed that ion−base interactions were more common than ion−sugar-phosphate backbone interactions, which were consistently low regardless of the cation type. We also observed that guanine bases had the highest number of ion interactions, regardless of the ion type. Interestingly, we found increased thymine−Na^+ interactions compared to K^+.

Water and Ion Interactions in the Gua9•Gua13•Gua21 Triad. The two-tetrad telomeric GQ topology is atypical among solved structures, with an incomplete core and a pronounced Thy11•Thy22 base pair, and was the site of interesting water and ion dynamics throughout the simulations. The frequency of bulk ion exchange in core ion binding site 3 was greater than those in our previous simulations of DNA or RNA GQs,[25−28] which we attribute to the water interactions around the triad. A previous GQ simulation study described the formation of a G-triad−water complex in which coordinated waters occupied the "vacant site" of the triad, mimicking the position of the missing guanine carbonyl oxygen ligand.[52] As in that study, we observed high water occupancies around the triad, specifically between Gua13 and Gua21 in the triad. These waters occasionally interacted with the ion occupying core binding site 2; however, water primarily interacted with the ions occupying core binding site 3 and the guanine bases. In fact, the exchange of ions in site 3 appeared to be induced by interactions of water in the vacant site. By tracing these interactions in KCl simulations, we found 2−3 water molecules within 3.5 Å of core binding site 3 (defined as the center of mass between O6 atoms of triad guanines) and 4−5 water molecules within 7 Å allowed for bipyramidal coordination of the ion (Figure 4A). However, when more water molecules occupied the volume enclosed within this 7 Å cutoff, the ions adopted nonlinear coordination, such that the bulk ion interacted with the O6 of triad guanine but was not aligned along the core axis (Figure 4B). Such coordination was often followed by ion exchange shortly after. In NaCl simulations, we found 1−2 water molecules within 3.5 Å of core binding site 3 and 3−4 water molecules within 7 Å, which resulted in coplanar coordination of the ion (Figure 4C), the expected behavior for Na^+ ions.[26,53] When a greater number of water molecules occupied the 7 Å cutoff, the Na^+ ions adopted nonlinear coordination and subsequent bulk ion exchange (Figure 4D). Increased water occupancies near the vacant site corresponded to ion exchange between coordinated and bulk ions. Further, simulations in KCl had higher water occupancies near the vacant site than NaCl, which is interesting because K^+ has a lower dehydration penalty than Na^+.[54] It is likely that K^+ engaged in more water interactions because it was more solvent accessible in bipyramidal coordination compared to coplanar Na^+ coordination.

The simulations performed by Heddi et al. did not result in ion exchange because the ions were restricted using distance constraints to the core O6 atoms.[52] It is possible that the fixed position of the ion in the triad contributed to restricted water dynamics in the vacant site and influenced ion−water and water−triad interactions within the simulation. The differences in water sampling between these simulations may also be due to FF selection (AMBER[55] vs Drude-2017[23,24]) or differing GQ topology. In fact, conformational sampling influenced core ion binding across replicates in our NaCl simulations. While the ion in core binding site 3 was exchanged multiple times throughout most of the simulations, in NaCl replicate 2, the Na^+ ion in core binding site 3 did not exchange due to a network of interactions (<3.5 Å) with Thy11 and Thy22 O4 atoms (Supporting Information, Figure S10). Further, in replicate 3 of NaCl simulations, an ion bound to core binding site 2, while ions already occupied core binding sites 1 and 3. This binding mode was unexpected considering our previous observations of external ion binding, in which ions bound in GQ tetrad cores preceded binding on outer faces of the tetrad core. In this case, the observed binding was possible because backbone sampling in residues 10−12 caused the gap between Gua13 and Gua21 to widen (Supporting Information, Figure S11). In most cases of ion binding, the triad mimicked the behavior of tetrad 1 in other systems, meaning that it helped attract ions that eventually bound to the core. During the binding event in replicate 3, the triad played a role similar to flanking nucleotides in other systems. While this binding mechanism was an isolated event, it shows how the triad motif can function like the tetrad core or linker regions. More work is needed to show if this event occurred in NaCl simulations due to increased DNA−ion interactions in the backbone region of residues 10−12 or because Na^+ ions are small enough to fit through the gap between Gua13 and Gua21.

https://doi.org/10.1021/acsomega.2c01600
ACS Omega 2022, 7, 23368−23379

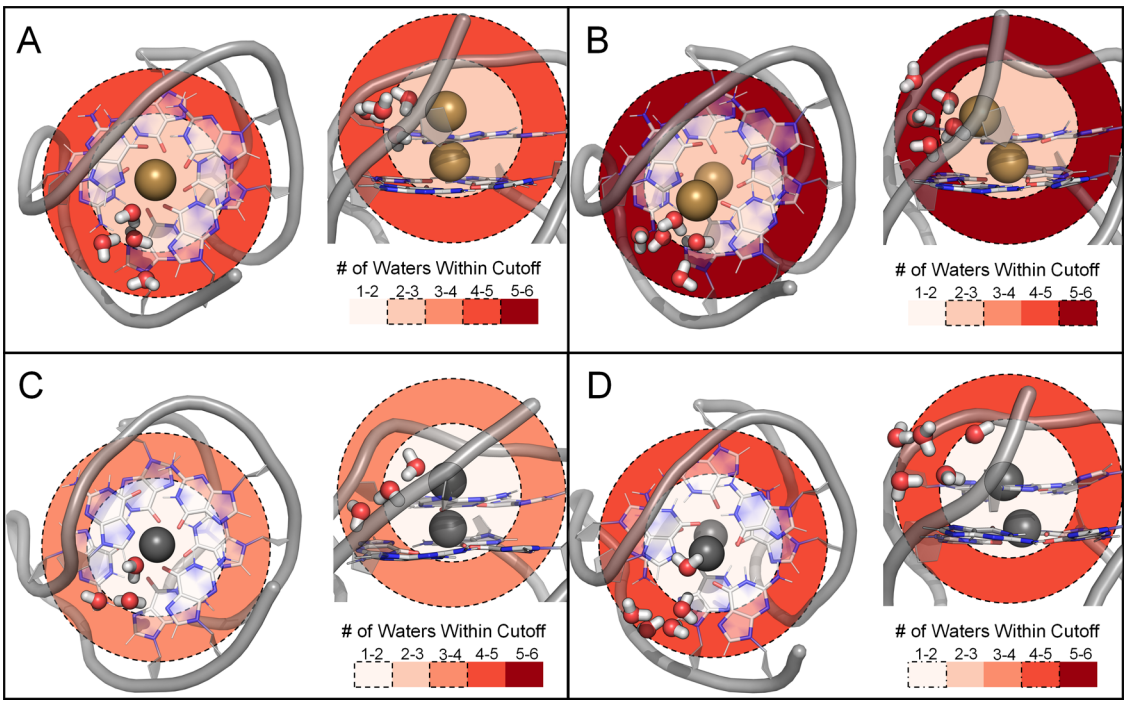


Figure 4. Images of water−triad and ion−triad interactions overlaid on a heatmap of water within 3.5 Å (inner circle) and 7 Å (outer circle) from the ion binding site. Structures are representative examples of water−ion−triad interactions during (A) bipyramidal K^+ coordination, (B) nonlinear K^+ coordination, (C) coplanar Na^+ coordination, and (D) nonlinear Na^+ coordination.

Conformational Dynamics of the SC11 G-Hairpin. To characterize the dynamics of the isolated *SC11* G-hairpin, we performed RMSD, per-nucleotide RMSD, and per-nucleotide RMSF analysis, traced important G•G base pairing, and generated RMSD-based clusters. While the NMR ensemble showed little variation with a pairwise RMSD of 0.63 Å,[7] the ensemble obtained from our Drude FF simulations indicated considerable flexibility (Supporting Information, Figures S12, S13 and Table S3). In solutions of KCl, G•G base pairs were maintained in all replicates, but their hydrogen bonding interactions rearranged over time, deviating from their initial NMR assignment. In the Gua1•Gua5 base pair, Gua1 O6•Gua5 H22 and O6•Gua5 H1 hydrogen bonds were rarely observed in KCl or NaCl (Supporting Information, Figure S14). However, Gua1 H1•Gua5 O6 and Gua1 H21•Gua5 O6 hydrogen bonds were preserved in most replicates. The Gua6•Gua9 base pair was maintained across replicates, except for the Gua6 O6•Gua9 H1 hydrogen bond (Supporting Information, Figure S15). In the Gua7•Gua11 base pair, the Gua7 H1•Gua11 N7 hydrogen bond was maintained, while the Gua7 O6•Gua11 H1 hydrogen bond broke and reformed in KCl simulations but broke irreversibly in NaCl simulations (Supporting Information, Figure S16). Despite these differences, the overall G-hairpin motif was maintained in KCl as well as NaCl replicate 1, agreeing with the overall experimental ensemble (Supporting Information, Figures S17 and 18). RMSD-based clustering showed that the linker nucleotides (those not participating in G•G base pairing) and backbone were most flexible. Gua3, which is stacked tightly to Gua1 in the NMR model, could flip to interact with the solvent, much like the thymine bases in the system (at positions 2, 4, 8, and 10).

We also calculated NOE violations for the G-hairpin simulations performed in KCl. We calculated H−H distances for all internucleotide NOEs except those associated with thymine "H7" atoms, which are not uniquely assigned to a specific H atom. We also omitted intranucleotide NOEs, as they reflect subtle conformational changes and not the agreement with the overall fold of the structure. Therefore, we included 80 NOEs in this analysis. We considered a violation to be more than 1 Å outside of the boundaries defined in the experimental restraint file. The results of this analysis are shown in the Supporting Information, Table S4. The greatest number of violations occurred in the terminal nucleotides, Gua1 and Gua11, which were somewhat flexible despite preserving some key interactions noted above. The variation in their positions led to some NOE violations associated with Gua7, which is hydrogen-bonded to Gua1 and Gua11 in the experimental ensemble. Most of the violations in Gua7 involved H4′, suggesting some deviation likely in the γ torsion, which is known to be slightly up-shifted in the Drude-2017 FF. Other infrequent NOE violations in other nucleotides typically involved H4′, H5′, and H5″, again all suggestive of small γ deviations.

In NaCl simulations, some large conformational changes occurred, which coincided with coordinated DNA−ion interactions. In NaCl replicate 2, notable conformational changes occurred around 40, 366, 780 ns, and 1.14 μs (Supporting Information, Figures S14−S16 and S19). At ∼40 ns, ions were coordinated by the carbonyl oxygens of guanines 1, 7, and 11, causing the stacked Gua1 and Gua3 bases to slide down toward the Gua7•Gua11 base pair. This rearrangement resembled strand slippage and resulted in six carbonyl oxygens facing the same direction (guanines 1, 3, 6, 7, 9, and 11), creating electronegative sites for ion binding. As ions occupied these sites, guanine conformations continued to fluctuate. By 366 ns, bound ions helped separate the Gua1•Gua5 and Gua7•Gua11 base pairs and Gua1•Gua11 base stacking, resulting in an unfolded G-hairpin by 780 ns. The unfolded state persisted between 400 ns and 1.2 μs until guanines 3, 6,

https://doi.org/10.1021/acsomega.2c01600
ACS Omega 2022, 7, 23368−23379

and 9 rearranged around an ion coordinated to their carbonyl oxygens and formed a triad. The triad motif was preserved for the remainder of the simulation with ions roughly equidistant to these O6 atoms. This motif was rigid, demonstrated by low RMSF values for these triad residues (Supporting Information, Figure S13), while the other eight nucleotides in the system sampled many states. These important ion−DNA interactions are highlighted in the Supporting Information, Video S3, and are further discussed in the next section ("Ion Sampling and Energetics in *SC11* G-Hairpin States").

In NaCl replicate 3, notable conformational changes occur around 20, 860, 980 ns, and 1.1 μs (Supporting Information, Figure S20), again coinciding with coordinated ion interactions. Around 20 ns, multiple bulk ions were coordinated to carbonyl oxygens of guanines 1, 3, 7, 9, and 11 and important G•G base pairs shifted, altering the conformation of the hairpin. The stacked Gua1 and Gua3 bases slid down toward the Gua7•Gua11 base pair, which adopted a Hoogsteen hydrogen bonding arrangement. As in replicate 2, these changes created electronegative sites, which led to high ion occupancies and unfolding by 40 ns. The unfolded state persisted until ∼860 ns when a G-hairpin temporarily reformed. At this time, the G•G base pairing was remarkably similar to the starting structure; however, the glycosidic torsion angles were different. This conformation is referred to as G-hairpin state 2 (Supporting Information, Figure S1) and is used as the starting structure of additional simulations (discussed later). Again, bulk Na^+ ions were coordinated to the center of the hairpin, which induced G•G base pair rearrangement. This time, Gua1 and Gua11 interacted, eventually adopting a Hoogsteen hydrogen bonding arrangement and disrupting the G-hairpin (within ∼10 ns). Guanines 1, 7, and 11 formed a triad around a coplanar Na^+ ion, and shortly after, Gua5 became coplanar with the triad, forming a tetrad. At this time, the Gua6•Gua9 base pair was also stacked above the tetrad in Hoogsteen hydrogen bonding arrangement. This arrangement helped attract a bulk ion, which aligned above the ion bound to the tetrad in a configuration reminiscent of a typical GQ tetrad core. As this ion coordination persisted, Gua3 rearranged around the ion bound to Gua6•Gua9, forming a stacked triad/tetrad structure (∼980 ns). At first, the Gua3•Gua6•Gua9 triad and Gua1•Gua5•Gua7•Gua11 tetrad were both flexible. Then, the ion coordinated to the triad left and the guanine backbone and base conformations fluctuated, until ∼1.1 μs when a bulk ion bound to the triad and stabilized the structure. After this ion binding event, both the triad and tetrad structures were rigid, and the coordinated ions were retained for the remainder of the simulation. These important ion−DNA interactions are highlighted in the Supporting Information, Video S4.

The conformational changes observed in these replicates, to our knowledge, represent the first time that unbiased MD simulations have captured the formation of a triad, tetrad, or stacked triad/tetrad motif. These structural rearrangements appear to be driven by interactions with ions (discussed later) and revealed rare states in the ensemble that may be relevant to the G-hairpin folding pathway. To further characterize their conformational ensembles, we performed simulations of the telomeric *SC11* G-hairpin state 2 (Supporting Information, Figure S1, Table 1) and stacked triad/tetrad state (Supporting Information, Figure S2, Table 1) in solutions of KCl and NaCl. We also performed simulations of unfolded states (Supporting Information, Figure S3, Table 1) in solutions of KCl to see if triad, tetrad, or stacked triad/tetrad motifs would form.

In solutions of KCl and NaCl, *SC11* G-hairpin state 2 was not preserved. The initial structure was disrupted quickly, remaining stable for only ∼2% of the simulation in KCl and ∼1% in NaCl (Supporting Information, Figures S21 and S22). No specific motifs, such as a triad, tetrad, or stacked triad/tetrad, were adopted. This outcome suggests that the structure is not an important intermediate in the conformational landscape of the G-hairpin.

In solutions of KCl, the stacked triad/tetrad state was not preserved in a 500 ns simulation (Supporting Information, Figure S23). The tetrad began to distort within 5 ns, and the motif was disrupted within 50 ns, representing <10% of the simulation time. The only persistent interaction in the simulation was Gua6•Gua9 base pairing, while the rest of the structure fluctuated. However, in solutions of NaCl, the stacked triad/tetrad state was preserved for the entire 1 μs simulation time, resulting in low RMSD values (Supporting Information, Figure S24). Only thymine 2, 4, and 10 were variable throughout the simulation, while the tetrad and triad motifs were rigid. The unbiased formation of the structure and preservation in an independent 1 μs simulation suggests that it is a stable intermediate or motif in this g-DNA sequence, one that may be specifically stabilized by NaCl. Finally, simulations of unfolded states did not adopt hairpin, triad, or tetrad motifs (Supporting Information, Figures S25 and S26). The structure was variable throughout the 1 μs simulations, with the only consistent interaction being the Gua6•Gua9 base pair, which engaged in stable Hoogsteen hydrogen bonding.

G-hairpins have been identified as potential GQ folding intermediates,[11−19,56−59] so the formation of triad and tetrad motifs in NaCl systems is plausible; however, the formation and preservation of these motifs in NaCl but not KCl is interesting. Since KCl preferentially folds GQs, one might expect to see these motifs in KCl simulations, but studies have shown that misfolded or intermediate structures can persist for seconds to minutes and that a greater number of kinetic intermediates are present in solutions of NaCl, influencing the folding kinetics of potential GQ (pGQ) forming sequences.[60−62] So in this case, the formation of the stacked triad/tetrad in NaCl may represent long-lasting intermediates that frustrate the folding pathway of pGQ sequences. Of course, extending the sequence to include regions upstream and downstream may influence its ability to fold into the described motifs. Nevertheless, these simulations demonstrate the conformational complexity of g-DNA and further suggest that DNA−cation interactions are integral to the folding landscape of noncanonical DNA structures.

Ion Sampling and Energetics in SC11 G-Hairpin States. Ion sampling in the *SC11* G-hairpin simulations was variable, like the conformational ensemble of the G-hairpin itself. To understand and compare ion interactions within our systems, we generated ion maps for and traced ion−base and ion−backbone interactions throughout all simulations. As mentioned above, there could be up to six carbonyl oxygens facing one direction, depending on base fluctuation, creating electronegative sites for ion binding. In KCl, ions primarily occupied these sites, near guanine O6 atoms (Supporting Information, Figure S27A). In contrast, NaCl replicate 1 produced more diffuse ion sampling, interacting with guanine O6 as well as thymine O2 and O4 atoms (Supporting Information, Figure S27B). The ion−DNA heat maps also

https://doi.org/10.1021/acsomega.2c01600
ACS Omega 2022, 7, 23368−23379

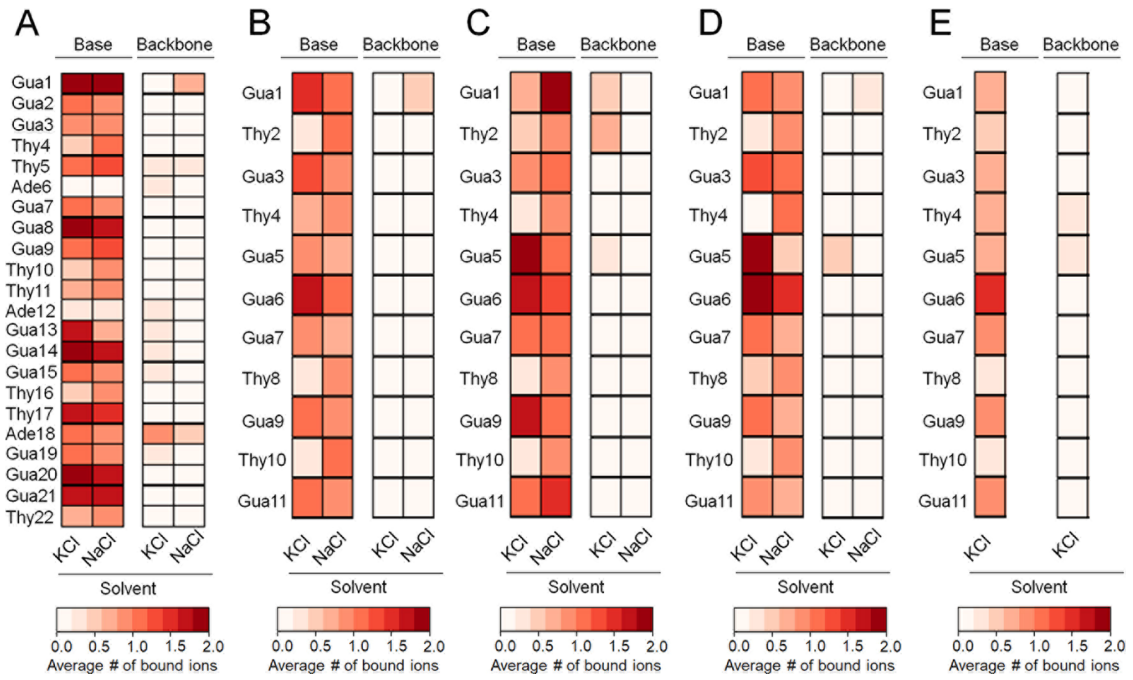


Figure 5. Ion−DNA interactions in g-DNA simulations. The heat maps show the average number of ions within 3.5 Å of nonhydrogen atoms of the bases and sugar-phosphate backbone in (A) telomeric two-tetrad GQ, (B) isolated G-hairpin simulations, (C) simulations starting in the triad/tetrad state, (D) simulations starting in G-hairpin state 2, and (E) simulations starting in unfolded states.

show distinct guanine base−K^+ interactions, whereas Na^+ ions bound nonspecifically to bases in the structure (Figure 5). It is possible that such nonspecific Na^+ sampling either fails to stabilize important interactions in the G-hairpin structure or contributes to structural transition. There were no specific ion−backbone interactions throughout these simulations. Since the conformational dynamics of the backbone varied across simulations, this outcome suggests that the affinity of ions for a given atom type takes precedence over location or the constituent base of the nucleotide at a given backbone location.

To better understand how ion−guanine interactions influenced the conformational sampling and energetics of the G-hairpins, we compared structural deviation, a number of ion−guanine interactions, and guanine−ion interaction energies over time. In all replicates without large conformational change, all KCl replicates and NaCl replicate 1, the RMSD, the number of interacting ions, and guanine−ion interaction energies were consistent throughout simulations (Supporting Information, Figures S28 and S29). However, in NaCl replicates 2 and 3, there were large changes in the number of interacting ions and guanine−ion interaction energies that corresponded to conformational changes (Supporting Information, Figure S29). In general, increased ion interactions coincided with increased RMSD and strengthened interaction energies. Interestingly, in replicate 3, in which the stacked triad/tetrad structure formed, there were decreased ion interactions and strengthened guanine−ion interaction energies when the triad/tetrad was forming (between 860 and 980 ns). When the triad ion was expelled and the stacked triad/tetrad structure was flexible (between 980 ns and 1.1 μs), interaction energies were weakened. Once the stacked triad/tetrad structure stabilized, so did the number of ion interactions and the ion−guanine interaction energies. In short, an increased number of ions interacting with guanine bases did not always result in strengthened interaction energies. This outcome suggests that specific ion interactions, like those coordinating triad/tetrad formation between 860 and 980 ns, have greater influence on interaction energy than the number of ions interacting. Further, there may be a saturation effect such that proximal ions do not additively impact interaction energy, with only the closest few ions contributing strongly.

To better understand the relationship between structure and ion−guanine interaction energy, we plotted their distributions for each *SC11* G-hairpin state (Supporting Information, Figure S30). These distributions show that the guanine−Na^+ interaction energies were stronger than guanine−K^+ interaction energies, regardless of structure, which agrees with our previous simulation outcomes.[26,29] The interaction energy distribution of the stacked triad/tetrad structure in NaCl was very narrow as the structure was rigid and the core ion interactions were consistent throughout the simulation (Supporting Information, Figure S30B). However, the isolated G-hairpin simulated in KCl, which was a stable throughout all replicates, had a broad distribution, suggesting that small changes in conformational sampling or ion binding could greatly influence the strength of guanine−ion interactions (Supporting Information, Figure S30C). There were also wide guanine−ion interaction energy distributions for systems that underwent a greater degree of conformational change (isolated G-hairpin in NaCl, stacked triad/tetrad in KCl, G-hairpin state 2, and unfolded states). Together, these outcomes show that the relationship between ion−guanine interactions and g-DNA conformational change is complex and is a function of the nature of the ion itself as well as the architecture adopted by the g-DNA at the time of coordination.

Core Ion and Water Interactions in the Stacked Triad/Tetrad. The unbiased formation of the stacked triad/tetrad structure provided an opportunity to track the triad−water and

https://doi.org/10.1021/acsomega.2c01600
ACS Omega 2022, 7, 23368−23379

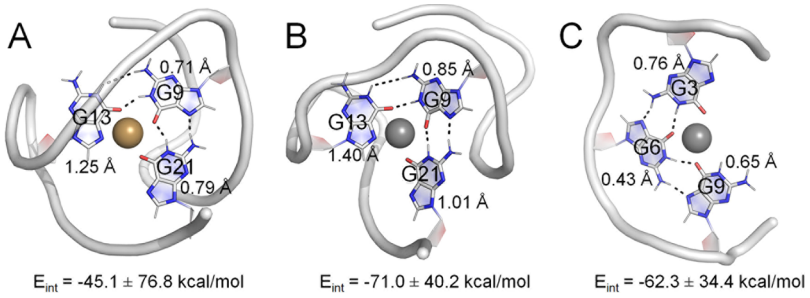


Figure 6. Comparison of the triad motif within (A and B) two-tetrad telomeric GQ and (C) *SC11* G-hairpin. Structure, Hoogsteen hydrogen bonding interactions, and average RMSF values for each guanine base are highlighted. The average guanine−ion interaction energy in each system is listed below the representative structure.

triad−ion interactions during and after formation. Before conformational change to the stacked triad/tetrad, an ion partially dehydrated (3−4 coordinated water molecules) and interacted with the carbonyl oxygens of the Hoogsteen hydrogen-bonded Gua7•Gua11 pair. This ion interaction was long-lasting and helped coordinate additional guanine bases comprising the tetrad. The carbonyl oxygens of Gua1 and Gua5 coordinated directly with the bound ion (<3 Å) for ∼120 ns before forming Hoogsteen hydrogen bonds with Gua7 and Gua11 to yield the tetrad. During these 120 ns, the bound ion further dehydrated, ultimately retaining only 1−2 coordinated water molecules. The formation of the triad was similar, with an ion partially dehydrating (3−4 coordinated water molecules) and binding to the Hoogsteen hydrogen bonded Gua6•Gua9 base pair before recruiting Gua3. As Gua3 repositioned into the triad, the ion bound to the Gua6•Gua9 base pair was expelled. Shortly after, another ion bound to the triad coordination site, stabilizing the final structure of the triad. As in the two-tetrad telomeric GQ, there was a vacant site in the triad. In this system, 1−2 water molecules occupied this vacant site. This observation also extends to the simulation starting in the stacked triad/tetrad state, in which the water−ion interactions in the core were very consistent. Here, there was generally one water molecule interacting above the Na^+ bound to the triad and 1−2 water molecules interacting in the vacant site. Further, there were 1−2 water molecules interacting below the Na^+ bound to the tetrad. These consistent interactions help further explain the narrow interaction energy distribution in the system (Supporting Information, Figure S30).

We previously showed that smaller ions like Na^+ and Li^+ can partition into the folded *c-kit1* GQ without completely dehydrating,[29] challenging the longstanding notion that cations have to desolvate completely to access the GQ core.[63−66] These outcomes also suggest that Na^+ ions do not have to be completely dehydrated to aid folding or otherwise stabilize folded G-tetrads or G-triads. These observations could be topology-dependent, as the *SC11* G-hairpin is not a full GQ-forming sequence and there are no flanking nucleotides above or below the core. It is also possible that the propensity for Na^+ to retain several water molecules when coordinating triad or tetrad folding contributes to a greater number of folding intermediates or differing folded topologies commonly observed in solutions of NaCl.[32] More work is needed to answer these questions; still, the formation of this stacked triad/tetrad structure provides interesting insight on the conformational diversity of g-DNA.

Comparison of the Two-Tetrad Telomeric GQ and SC11 G-Hairpin Systems. Together, these simulations allowed us to compare two distinct g-DNA structures with similar motifs. Overall, the two-tetrad telomeric GQ was much more rigid than the *SC11* G-hairpin, demonstrated by lower RMSD (Supporting Information, Tables S1 and S3) and RMSF (Supporting Information, Figures S6 and S13). This outcome is sensible as GQs exhibit greater stability than most nucleic acid structures.[67] Though the *SC11* G-hairpin sampled many states, the average number of ions interacting with each base type was very similar across the simulations (Supporting Information, Tables S2 and S5−S8), as were the dipole moment distributions (Supporting Information, Figure S31). As mentioned above, we observed increased ion−thymine interactions in solutions of NaCl, resulting in an increase in thymine base polarization in the GQ and G-hairpin systems. Guanine and adenine base polarizations were consistent, independent of the cation type. These outcomes suggest that the relationship between thymine and Na^+ might play a role in GQ preference for K^+, as K^+ engaged in more specific contact with guanine bases that may better stabilize folded g-DNA.

The formation of a triad motif in the *SC11* G-hairpin systems allows us to compare this secondary structure motif in the context of two different folded topologies. To better understand the triad motif, we compared the dominant triad structures with RMSD and compared RMSF, water coordination, and interaction energies. Since the stacked triad/tetrad structure was not preserved in solutions of KCl, it was not included in this comparison. The hydrogen bonding networks within the triad motifs were similar across systems; however, the triad in the two-tetrad telomeric GQ was more flexible, demonstrated by higher RMSF values (Figure 6). This flexibility may be due to interactions of proximal linker nucleotides or because of frequent ion exchange at the site. Further, the ion−guanine interaction energies within the triad were stronger in NaCl and, interestingly, the highest in the GQ system. Since the ion interactions within the triad were more consistent in the *SC11* G-hairpin, one might expect a stronger interaction energy, but we have previously shown that there is electrostatic cooperativity within GQ tetrad cores,[25] which may result in stronger interaction energies in a system with two tetrads instead of one. However, the high statistical variance in the ion−guanine interaction energies calls for additional studies. Further, the high rate of ion exchange in the GQ triad compared to the absence of ion exchange in the *SC11* triad also raises questions. As noted above, the triad in the two-tetrad telomeric GQ can mimic the behavior of a tetrad or a linker region, depending on small conformational changes.

https://doi.org/10.1021/acsomega.2c01600
ACS Omega 2022, 7, 23368−23379

Since this behavior was not observed in the stacked triad/tetrad structure, the overall folded topology of the structure may contribute to these differences. It is also possible that the ability of the triad to act as a linker or core region may function as an on/off switch for small molecule or ligand binding. These outcomes once again demonstrate the complexity of secondary structure within g-DNA and that more work is needed to characterize guanine-containing triads.

CONCLUSIONS

Using polarizable MD simulations, we have studied two interesting g-DNA structures, the two-tetrad telomeric GQ and the telomeric *SC11* G-hairpin, in atomistic detail. In doing so, we found that the two-tetrad telomeric GQ structure is rigid and shows dense ion accumulation in four locations, three within the tetrad core and one below tetrad 2 in the ion binding pocket. While the number and location of these ion binding sites resemble that of a three-tetrad GQ, triad−water and triad−ion interactions were distinct from those of a tetrad. Throughout simulations of the two-tetrad GQ, we also observed that the triad could function as a part of the tetrad core, attracting and binding bulk ions to the core, or function as a linker region, binding ions outside the core, on the GQ exterior. The functional diversity of the triad motif might play a role in modulating biological activity in living systems. More work is needed to better understand the dynamics and function of G-triad motifs within GQs.

Simulations of the *SC11* G-hairpin showed that the structure is flexible and we found that in solutions of NaCl, the G-hairpin underwent conformational change to a stacked triad/tetrad, a structure that was stable in additional simulations. These outcomes suggest that the conformational ensemble of the G-hairpin may be more diverse than originally thought and provided insight on the formation of triad and tetrad motifs. We previously showed that Na^+ could bind to the *c-kit1* tetrad core without completely dehydrating,[29] and we observed a similar phenomenon in the formation of the stacked triad/tetrad structure. Ions coordinated the formation of a tetrad and then a triad while retaining 1−2 coordinated water molecules, challenging the idea that complete dehydration is needed for core−ion coordination. In comparing ion sampling in all simulations, we showed that K^+ ions preferentially bind to guanine bases, regardless of the secondary or tertiary structure, while Na^+ ions commonly interact with thymine bases. We previously observed nonspecific Na^+ ion sampling in the *c-kit1* GQ, and since this observation extends to the two-tetrad telomeric GQ, the *SC11* G-hairpin, and intermediate states of the G-hairpin, nonspecific ion sampling may play an important role in preference for K^+ over Na^+ in GQ folding. Altogether, these simulations highlight the complexity of g-DNA structures and their interactions with water and ions.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.2c01600.

Tables of RMSD, ion counts, NOE violations, structural schematics, snapshots of interesting configurations, structural analyses (RMSD, RMSF, and hydrogen bonding), dihedral distributions, central structures from RMSD-based clustering, snapshots of ion coordination, per-nucleotide RMSD and RMSF of the *SC11* hairpin, hydrogen bond analysis of the *SC11* hairpin, snapshots of *SC11* conformational change, ion occupancy maps, ion interaction energies, and base dipole moment distributions (PDF)

(Video S1) Ion desolvation at core binding site 2 in the two-tetrad GQ (MOV)

(Video S2) Ion exchanges at core binding site 3 in the two-tetrad GQ (MOV)

(Video S3) Ion-triad interactions in G-hairpin replicate 2 in NaCl (MOV)

(Video S4) Ion-triad/tetrad interactions in G-hairpin replicate 3 in NaCl (MOV)

AUTHOR INFORMATION

Corresponding Author

Justin A. Lemkul − *Department of Biochemistry and Center for Drug Discovery, Virginia Tech, Blacksburg, Virginia 24061, United States;* orcid.org/0000-0001-6661-8653; Phone: +1 (540) 231-3129; Email: jalemkul@vt.edu

Authors

Alexa M. Salsbury − *Department of Biochemistry, Virginia Tech, Blacksburg, Virginia 24061, United States;* orcid.org/0000-0001-5736-1294

Haley M. Michel − *Department of Biochemistry, Virginia Tech, Blacksburg, Virginia 24061, United States;* orcid.org/0000-0002-4787-3410

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.2c01600

Funding

This work was supported by the National Institutes of Health (grant R35GM133754 to J.A.L.), the Thomas F. and Kate Miller Jeffress Memorial Trust (Bank of America, Trustee), USDA-NIFA (project number VA-160092 to J.A.L.), and The American Association of University Women (American Dissertation Fellowship to A.M.S.).

Notes

The authors declare no competing financial interest.

Simulation trajectories, corresponding topologies in the PSF format, and example input files are supplied via the Open Science Framework at https://osf.io/2pfj8/.

ACKNOWLEDGMENTS

The authors thank Darcy S. Davidson, Tanner J. Dean, and Melanie Hempel for contributions in the initial stages of the project, Diane F. Eilerts for helpful discussions regarding analysis, Marcelo D. Polêto for help in data processing, and Virginia Tech Advanced Research Computing for providing computing time and resources.

REFERENCES

(1) Sfeir, A. J.; Chai, W.; Shay, J. W.; Wright, W. E. Telomere-End Processing: The Terminal Nucleotides of Human Chromosomes. *Mol. Cell* **2005**, *18*, 131−138.

(2) Yeh, J.-K.; Wang, C.-Y. Telomeres and Telomerase in Cardiovascular Diseases. *Genes* **2016**, *7*, 58.

(3) Eitan, E.; Hutchison, E. R.; Mattson, M. P. Telomere Shortening in Neurological Disorders: An Abundance of Unanswered Questions. *Trends Neurosci.* **2014**, *37*, 256−263.

(4) Maciejowski, J.; De Lange, T. Telomeres in Cancer: Tumour Suppression and Genome Instability. *Nat. Rev. Mol.* **2017**, *18*, 175−186.

(5) Aksenova, A. Y.; Mirkin, S. M. At the Beginning of the End and in the Middle of the Beginning: Structure and Maintenance of Telomeric Dna Repeats and Interstitial Telomeric Sequences. *Genes* **2019**, *10*, 118.
(6) Phan, A. T.; Mergny, J. L. Human Telomeric DNA: G-Quadruplex, i-Motif and Watson-Crick Double Helix. *Nucleic Acids Res.* **2002**, *30*, 4618−4625.
(7) Gajarský, M.; Živković, M. L.; Stadlbauer, P.; Pagano, B.; Fiala, R.; Amato, J.; Tomáška, L.; Šponer, J.; Plavec, J.; Trantírek, L. Structure of a Stable G-Hairpin. *J. Am. Chem. Soc.* **2017**, *139*, 3591−3594.
(8) Jansson, L. I.; Hentschel, J.; Parks, J. W.; Chang, T. R.; Lu, C.; Baral, R.; Bagshaw, C. R.; Stone, M. D. Telomere DNA G-Quadruplex Folding within Actively Extending Human Telomerase. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 9350−9359.
(9) Moye, A. L.; Porter, K. C.; Cohen, S. B.; Phan, T.; Zyner, K. G.; Sasaki, N.; Lovrecz, G. O.; Beck, J. L.; Bryan, T. M. Telomeric G-Quadruplexes Are a Substrate and Site of Localization for Human Telomerase. *Nat. Commun.* **2015**, *6*, 1−12.
(10) Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. Quadruplex DNA: Sequence, Topology and Structure. *Nucleic Acids Res.* **2006**, *34*, 5402−5415.
(11) Stadlbauer, P.; Krepl, M.; Cheatham, T. E.; Koca, J.; Šponer, J. Structural Dynamics of Possible Late-Stage Intermediates in Folding of Quadruplex DNA Studied by Molecular Simulations. *Nucleic Acids Res.* **2013**, *41*, 7128−7143.
(12) Šponer, J.; Bussi, G.; Stadlbauer, P.; Kührová, P.; Banáš, P.; Islam, B.; Haider, S.; Neidle, S.; Otyepka, M. Folding of Guanine Quadruplex Molecules−Funnel-like Mechanism or Kinetic Partitioning? An Overview from MD Simulation Studies. *Biochim. Biophys. Acta, Gen. Subj.* **2017**, *1861*, 1246−1263.
(13) Mashimo, T.; Yagi, H.; Sannohe, Y.; Rajendran, A.; Sugiyama, H. Folding Pathways of Human Telomeric Type-1 and Type-2 G-Quadruplex Structures. *J. Am. Chem. Soc.* **2010**, *132*, 14910−14918.
(14) Li, W.; Hou, X.-M.; Wang, P.-Y.; Xi, X.-G.; Li, M. Direct Measurement of Sequential Folding Pathway and Energy Landscape of Human Telomeric G-Quadruplex Structures. *J. Am. Chem. Soc.* **2013**, *135*, 6423−6426.
(15) Rajendran, A.; Endo, M.; Hidaka, K.; Sugiyama, H. Direct and Single-Molecule Visualization of the Solution-State Structures of G-Hairpin and G-Triplex Intermediates. *Angew. Chem., Int. Ed.* **2014**, *53*, 4107−4112.
(16) Zhang, A. Y. Q.; Balasubramanian, S. The Kinetics and Folding Pathways of Intramolecular G-Quadruplex Nucleic Acids. *J. Am. Chem. Soc.* **2012**, *134*, 19297−19308.
(17) Kuo, M. H. J.; Wang, Z. F.; Tseng, T. Y.; Li, M. H.; Hsu, S. T. D.; Lin, J. J.; Chang, T. C. Conformational Transition of a Hairpin Structure to G-Quadruplex within the WNT1 Gene Promoter. *J. Am. Chem. Soc.* **2015**, *137*, 210−218.
(18) Hou, X.-M.; Fu, Y.-B.; Wu, W.-Q.; Wang, L.; Teng, F.-Y.; Xie, P.; Wang, P.-Y.; Xi, X.-G. Involvement of G-Triplex and G-Hairpin in the Multi-Pathway Folding of Human Telomeric G-Quadruplex. *Nucleic Acids Res.* **2017**, *45*, 11401.
(19) Luu, K. N.; Phan, A. T.; Kuryavyi, V.; Lacroix, L.; Patel, D. J. Structure of the Human Telomere in K^+ Solution: An Intramolecular (3 + 1) G-Quadruplex Scaffold. *J. Am. Chem. Soc.* **2006**, *128*, 9963−9970.
(20) Šponer, J.; Mládek, A.; Špačková, N.; Cang, X.; Cheatham, T. E.; Grimme, S. Relative Stability of Different DNA Guanine Quadruplex Stem Topologies Derived Using Large-Scale Quantum-Chemical Computations. *J. Am. Chem. Soc.* **2013**, *135*, 9785−9796.
(21) Štefl, R.; Cheatham, T. E.; Špačková, N.; Fadrná, E.; Berger, I.; Koča, J.; Šponer, J. Formation Pathways of a Guanine-Quadruplex DNA Revealed by Molecular Dynamics and Thermodynamic Analysis of the Substates. *Biophys. J.* **2003**, *85*, 1787−1804.
(22) Ming, G. L.; Song, H. DISC1 Partners with GSK3beta in Neurogenesis. *Cell* **2009**, *136*, 990−992.
(23) Lemkul, J. A.; MacKerell, A. D. Polarizable Force Field for DNA Based on the Classical Drude Oscillator: II. Microsecond Molecular Dynamics Simulations of Duplex DNA. *J. Chem. Theory Comput.* **2017**, *13*, 2072−2085.
(24) Lemkul, J. A.; MacKerell, A. D. Polarizable Force Field for DNA Based on the Classical Drude Oscillator: I. Refinement Using Quantum Mechanical Base Stacking and Conformational Energetics. *J. Chem. Theory Comput.* **2017**, *13*, 2053−2071.
(25) Salsbury, A. M.; Lemkul, J. A. Molecular Dynamics Simulations of the *c-kit1* Promoter G-Quadruplex: Importance of Electronic Polarization on Stability and Cooperative Ion Binding. *J. Phys. Chem. B* **2019**, *123*, 148−159.
(26) Salsbury, A. M.; Dean, T. J.; Lemkul, J. A. Polarizable Molecular Dynamics Simulations of Two *c-kit* Oncogene Promoter G-Quadruplexes: Effect of Primary and Secondary Structure on Loop and Ion Sampling. *J. Chem. Theory Comput.* **2020**, *16*, 3430−3444.
(27) Lemkul, J. A. Same Fold, Different Properties: Polarizable Molecular Dynamics Simulations of Telomeric and TERRA G-Quadruplexes. *Nucleic Acids Res.* **2019**, *48*, 561−575.
(28) Ratnasinghe, B. D.; Salsbury, A. M.; Lemkul, J. A. Ion Binding Properties and Dynamics of the *bcl*-2 G-Quadruplex Using a Polarizable Force Field. *J. Chem. Inf. Model.* **2020**, *60*, 6476−6488.
(29) Salsbury, A. M.; Lemkul, J. A. Cation Competition and Recruitment around the c-kit1 G-Quadruplex Using Polarizable Simulations. *Biophys. J.* **2021**, *120*, 2249−2261.
(30) Gkionis, K.; Kruse, H.; Platts, J. A.; Mládek, A.; Koča, J.; Šponer, J. Ion Binding to Quadruplex DNA Stems. Comparison of MM and QM Descriptions Reveals Sizable Polarization Effects Not Included in Contemporary Simulations. *J. Chem. Theory Comput.* **2014**, *10*, 1326−1340.
(31) Fadrná, E.; Špačková, N.; Štefl, R.; Koča, J.; Cheatham, T. E.; Šponer, J. Molecular Dynamics Simulations of Guanine Quadruplex Loops: Advances and Force Field Limitations. *Biophys. J.* **2004**, *87*, 227−242.
(32) Lim, K. W.; Amrane, S.; Bouaziz, S.; Xu, W.; Mu, Y.; Patel, D. J.; Luu, K. N.; Phan, A. T. Structure of the Human Telomere in K^+ Solution: A Stable Basket-Type G-Quadruplex with Only Two G-Tetrad Layers. *J. Am. Chem. Soc.* **2009**, *131*, 4301−4309.
(33) Stadlbauer, P.; Kührová, P.; Banáš, P.; Koča, J.; Bussi, G.; Trantírek, L.; Otyepka, M.; Šponer, J. Hairpins Participating in Folding of Human Telomeric Sequence Quadruplexes Studied by Standard and T-REMD Simulations. *Nucleic Acids Res.* **2015**, *43*, 9626−9644.
(34) Brooks, B. R.; Brooks, C. L.; MacKerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; et al. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30*, 1545−1614.
(35) MacKerell, A. D.; Banavali, N. K. All-Atom Empirical Force Field for Nucleic Acids: II. Application to Molecular Dynamics Simulations of DNA and RNA in Solution. *J. Comput. Chem.* **2000**, *21*, 105−120.
(36) Foloppe, N.; MacKerell, A. D. All-Atom Empirical Force Field for Nucleic Acids: I. Parameter Optimization Based on Small Molecule and Condensed Phase Macromolecular Target Data. *J. Comput. Chem.* **2000**, *21*, 86−104.
(37) Hart, K.; Foloppe, N.; Baker, C. M.; Denning, E. J.; Nilsson, L.; MacKerell, A. D. Optimization of the CHARMM Additive Force Field for DNA: Improved Treatment of the BI/BII Conformational Equilibrium. *J. Chem. Theory Comput.* **2012**, *8*, 348−362.
(38) Lamoureux, G.; Harder, E.; Vorobyov, I. V.; Roux, B.; MacKerell, A. D. A Polarizable Model of Water for Molecular Dynamics Simulations of Biomolecules. *Chem. Phys. Lett.* **2006**, *418*, 245−249.
(39) Yu, H.; Whitfield, T. W.; Harder, E.; Lamoureux, G.; Vorobyov, I.; Anisimov, V. M.; MacKerell, A. D.; Roux, B. Simulating Monovalent and Divalent Ions in Aqueous Solution Using a Drude Polarizable Force Field. *J. Chem. Theory Comput.* **2010**, *6*, 774−786.
(40) Savelyev, A.; MacKerell, A. D. Balancing the Interactions of Ions, Water, and DNA in the Drude Polarizable Force Field. *J. Phys. Chem. B* **2014**, *118*, 6742−6757.

https://doi.org/10.1021/acsomega.2c01600
ACS Omega 2022, 7, 23368−23379

(41) Lamoureux, G.; Roux, B. Modeling Induced Polarization with Classical Drude Oscillators: Theory and Molecular Dynamics Simulation Algorithm. *J. Chem. Phys.* **2003**, *119*, 3025−3039.
(42) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.
(43) Miyamoto, S.; Kollman, P. A. SETTLE: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992**, *13*, 952−962.
(44) Chowdhary, J.; Harder, E.; Lopes, P. E. M.; Huang, L.; MacKerell, A. D.; Roux, B. A Polarizable Force Field of Dipalmitoylphosphatidylcholine Based on the Classical Drude Model for Molecular Dynamics Simulations of Lipids. *J. Phys. Chem. B* **2013**, *117*, 9142−9160.
(45) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; et al. OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *PLoS Comput. Biol.* **2017**, *13*, No. e1005659.
(46) Huang, J.; Lemkul, J. A.; Eastman, P. K.; MacKerell, A. D. Molecular Dynamics Simulations Using the Drude Polarizable Force Field on GPUs with OpenMM: Implementation, Validation, and Benchmarks. *J. Comput. Chem.* **2018**, *39*, 1682−1689.
(47) R Development Core Team, R. *R: A Language and Environment for Statistical Computing*, 2011.
(48) Carpenter, G. A.; Grossberg, S. ART 2: Self-Organization of Stable Category Recognition Codes for Analog Input Patterns. *Appl. Opt.* **1987**, *26*, 4919.
(49) Carpenter, G. A.; Grossberg, S. The ART of Adaptive Pattern Recognition by a Self-Organizing Neural Network. *Computer* **1988**, *21*, 77−88.
(50) Karpen, M. E.; Tobias, D. J.; Brooks, C. L. Statistical Clustering Techniques for the Analysis of Long Molecular Dynamics Trajectories: Analysis of 2.2-ns Trajectories of YPGDV. *Biochemistry* **1993**, *32*, 412−420.
(51) Kettani, A.; Basu, G.; Gorin, A.; Majumdar, A.; Skripkin, E.; Patel, D. J. A Two-Stranded Template-Based Approach to G·(C-A) Triad Formation: Designing Novel Structural Elements into an Existing DNA Framework. *J. Mol. Biol.* **2000**, *301*, 129−146.
(52) Heddi, B.; Martín-Pintado, N.; Serimbetov, Z.; Kari, T. M. A.; Phan, A. T. G-Quadruplexes with (4n - 1) Guanines in the G-Tetrad Core: Formation of a G-Triad·water Complex and Implication for Small-Molecule Binding. *Nucleic Acids Res.* **2016**, *44*, 910−916.
(53) Meyer, M.; Steinke, T.; Brandl, M.; Suhnel, J. Density Functional Study of Guanine and Uracil Quartets and of Guanine Quartet/Metal Ion Complexes. *J. Comput. Chem.* **2001**, *22*, 109−124.
(54) Eisenman, G. Cation Selective Glass Electrodes and Their Mode of Operation. *Biophys. J.* **1962**, *2*, 259−323.
(55) Pérez, A.; Marchán, I.; Svozil, D.; Šponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of α/γ Conformers. *Biophys. J.* **2007**, *92*, 3817−3829.
(56) Wang, Y.; Patel, D. J. Solution Structure of the Human Telomeric Repeat d[$AG_3(T_2AG_3)_3$] G-Tetraplex. *Structure* **1993**, *1*, 263−282.
(57) Dai, J.; Carver, M.; Punchihewa, C.; Jones, R. A.; Yang, D. Structure of the Hybrid-2 Type Intramolecular Human Telomeric G-Quadruplex in K^+ Solution: Insights into Structure Polymorphism of the Human Telomeric Sequence. *Nucleic Acids Res.* **2007**, *35*, 4927−4940.
(58) Parkinson, G. N.; Lee, M. P. H.; Neidle, S. Crystal Structure of Parallel Quadruplexes from Human Telomeric DNA. *Nature* **2002**, *417*, 876−880.
(59) Dai, J.; Chen, D.; Jones, R. A.; Hurley, L. H.; Yang, D. NMR Solution Structure of the Major G-Quadruplex Structure Formed in the Human BCL2 Promoter Region. *Nucleic Acids Res.* **2006**, *34*, 5133−5144.
(60) Marchand, A.; Gabelica, V. Folding and Misfolding Pathways of G-Quadruplex DNA. *Nucleic Acids Res.* **2016**, *44*, 10999−11012.
(61) Gray, R. D.; Trent, J. O.; Chaires, J. B. Folding and Unfolding Pathways of the Human Telomeric G-Quadruplex. *J. Mol. Biol.* **2014**, *426*, 1629−1650.
(62) Gray, R. D.; Chaires, J. B. Kinetics and Mechanism of K^+- and Na^+-Induced Folding of Models of Human Telomeric DNA into G-Quadruplex Structures. *Nucleic Acids Res.* **2008**, *36*, 4191−4203.
(63) Hud, N. V.; Smith, F. W.; Anet, F. A. L.; Feigon, J. The Selectivity for K^+ versus Na^+ in DNA Quadruplexes Is Dominated by Relative Free Energies of Hydration: A Thermodynamic Analysis by 1 H NMR. *Biochemistry* **1996**, *35*, 15383−15390.
(64) Gu, J.; Leszczynski, J. Origin of Na^+/K^+ Selectivity of the Guanine Tetraplexes in Water: The Theoretical Rationale. *J. Phys. Chem. A* **2002**, *106*, 529−532.
(65) Nieuwland, C.; Zaccaria, F.; Fonseca Guerra, C. Understanding Alkali Metal Cation Affinities of Multi-Layer Guanine Quadruplex DNA. *Phys. Chem. Chem. Phys.* **2020**, *22*, 21108.
(66) Bhattacharyya, D.; Mirihana Arachchilage, G.; Basu, S. Metal Cations in G-Quadruplex Folding and Stability. *Front. Chem.* **2016**, *4*, 38.
(67) Lane, A. N.; Chaires, J. B.; Gray, R. D.; Trent, J. O. Stability and Kinetics of G-Quadruplex Structures. *Nucleic Acids Res.* **2008**, *36*, 5482−5515.

https://doi.org/10.1021/acsomega.2c01600
ACS Omega 2022, 7, 23368−23379