THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

An Investigation of G-Quadruplex Structural Polymorphism in the Human Telomere Using a Combined Approach of Hydrodynamic Bead Modeling and Molecular Dynamics Simulation

Huy T. Le,[†] William L. Dean,[†,‡] Robert Buscaglia,[†] Jonathan B. Chaires,[†,‡,§] and John O. Trent*[,†,‡,§]

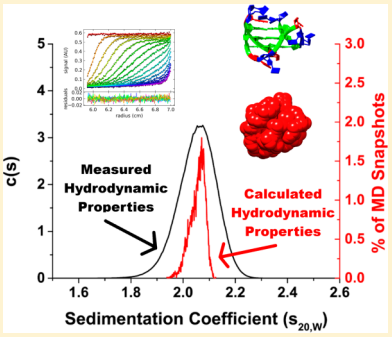
[†]Department of Biochemistry and Molecular Biology, School of Medicine, University of Louisville, HSC-A Building, Room 616, Louisville, Kentucky 40202, United States

[‡]James Graham Brown Cancer Center, University of Louisville, 529 South Jackson Street, Louisville, Kentucky 40202, United States

[§]Department of Medicine, School of Medicine, University of Louisville, 550 South Jackson Street, Louisville, Kentucky 40202, United States

Supporting Information

ABSTRACT: Guanine-rich oligonucleotides can adopt noncanonical tertiary structures known as G-quadruplexes, which can exist in different forms depending on experimental conditions. High-resolution structural methods, such as X-ray crystallography and NMR spectroscopy, have been of limited usefulness in resolving the inherent structural polymorphism associated with G-quadruplex formation. The lack of, or the ambiguous nature of, currently available high-resolution structural data, in turn, has severely hindered investigations into the nature of these structures and their interactions with small-molecule inhibitors. We have used molecular dynamics in conjunction with hydrodynamic bead modeling to study the structures of the human telomeric G-quadruplex-forming sequences at the atomic level. We demonstrated that molecular dynamics can reproduce experimental hydrodynamic measurements and thus can be a powerful tool in the structural study of existing G-quadruplex sequences or in the prediction of new G-quadruplex structures.



INTRODUCTION

In solutions with physiological Na^+ and K^+ concentration, single-stranded guanine-rich oligonucleotide sequences can self-assemble and fold into unimolecular G-quadruplexes, non-canonical DNA tertiary structures composed of a four-stranded helical stem and three interconnecting loops.[1] Within the human genome, over 370 000 putative G-quadruplex-forming sequences have been identified and most of these are observed to localize to genomic regions with important cellular functions, such as the telomere, immunoglobulin switch regions, proto-oncogene promoters, and mRNA untranslated regions.[2,3] Many of these sequences are found to be evolutionarily conserved between humans, mice, and rats[4] suggesting that G-quadruplex structures play important regulatory roles within the cell. The formation of G-quadruplex at the distal 3′ end of the human telomere region,[5] which contains a single-stranded guanine-rich overhang of approximately 100−200 bases, has been investigated as a potential target for novel small-molecule-based anticancer therapy. Small molecules that stabilize telomeric G-quadruplex structures have been shown to decrease the activity of telomerase *in vitro*.[6−10] Since telomerase is activated in more than 90% of all cancers,[11] G-quadruplex-based antitelomerase therapy could be an attractive strategy for the development of anticancer therapeutics.

Despite considerable research being devoted to targeting telomeric G-quadruplexes,[8,12−17] the development of small-molecule G-quadruplex-based inhibitors has progressed slowly with only one candidate drug making it to clinical trials.[18] A challenge to the rational design of small molecules that bind specifically to G-quadruplexes is the lack of, or the ambiguous nature of, high-resolution structural data for many putative G-quadruplex-forming sequences. In fact, the hTel22 sequence, AGGGTTAGGGTTAGGGTTAGGG, which is often used as an *in vitro* model for G-quadruplex formation in the human telomere,[19,20] has been found to exist in many forms depending on experimental conditions and sequence composition (Table 1). In the presence of sodium, it is widely accepted that this sequence folds into an antiparallel "basket" topology[19] which consists of three stacked G-tetrads with one diagonal and two lateral loops. In the presence of potassium, it exists as an ensemble of structures, which includes two mixed "hybrid" topologies (hybrid-1[21−23] and hybrid-2[23,24]), a parallel "propeller" topology,[20] and a new antiparallel "basket" topology.[25,26] Hybrid-1 consists of three stacked G-tetrads with a double chain-reversal loop followed by two lateral loops. Hybrid-2 also consists of three stacked G-tetrads but with reversed loop order, two lateral loops followed by a double chain-reversal loop. The parallel "propeller" topology consists of three stacked G-tetrads

Received: March 4, 2014
Revised: April 21, 2014
Published: April 29, 2014

   dx.doi.org/10.1021/jp502213y | J. Phys. Chem. B 2014, 118, 5390−5405

Table 1. G-Quadruplex-Forming Sequences for HYDROPRO Calculations

sequence[a]	designation[b]	topology	type	cations
*A***GGGTTAGGGTTAGGGTTAGGG**	143D	basket	solution	Na^+
*A***GGGTTAGGGTTAGGGTTAGGG**	1KF1	propeller	crystal	K^+
*TT***GGGTTAGGGTTAGGGTTAGGG***A*	2GKU	hybrid-1	solution	K^+
*AAA***GGGTTAGGGTTAGGGTTAGGG***AA*	2HY9	hybrid-1	solution	K^+
*TA***GGGTTAGGGTTAGGGTTAGGG**	2JSM	hybrid-1	solution	K^+
*TTA***GGGTTAGGGTTAGGGTTAGGG***TT*	2JPZ	hybrid-2	solution	K^+
*TA***GGGTTAGGGTTAGGGTTAGGG***TT*	2JSL	hybrid-2	solution	K^+
GGGTTAGGGTTAGGGTTAGG*GT*	2KF8	basket	solution	K^+
*A***GGGTTAGGGTTAGGGTTAGG***GT*	2KKA-G	basket	solution	K^+
*A***GGGTTAGGGTTAI**GG**TTAGG***GT*	2KKA-I	basket	solution	K^+

[a]G-quartet stem residues (**bold**) and flanking residues (*italic*). Noncanonical residues (underline). [b]Protein Data Bank code (www.pdb.org).

and three double chain-reversal loops. Lastly, the K^+ antiparallel "basket" topology consists of two stacked G-tetrads with a diagonal and two lateral loops.

The inherent structural polymorphism associated with G-quadruplex formation has severely hindered investigations of G-quadruplex structures and their formation. As a consequence, steps are usually taken to artificially reduce the structural polymorphism with the goal of enrichment of one species for NMR structure elucidation.[27,28] Sequence modification, as seen with human telomere sequence, is one common approach. The reported sequence variants for the human telomere sequence (Table 1) demonstrate how small changes with respect to the flanking bases can result in dramatically different dominant topologies. While each sequence contains the identical G-quadruplex-forming core GGGTTAGGGTTAGGGTTAGGG, the flanking bases differ from the hTel22 sequence (5′-A-core-3′), which contains a mixture of G-quadruplex structures, compared to the hybrid-1 dominant sequences (2GKU, 5′-TT-core-A-3′; 2HY9, 5′-AAA-core-AA-3′; 2JSM, 5′-TA-core-3′),[21−23] the hybrid-2 dominant sequences (2JPZ, 5′-TTA-core-TT-3′; 2JSL, 5′-TA-core-TT-3′),[23,24] and the antiparallel dominant sequences (2KF8, 5′-core-T-3′; 2KKA, 5′-A-core-T-3′).[25,26] In addition to changes in the flanking bases, G-quadruplex-forming sequences can also be truncated or elongated. Often, sequence modifications also involve the incorporation of noncanonical bases, as is the case with the 2KKA sequence which contains an inosine substitution for guanine. A list of G-quadruplex modifying constituents and their effects on G-quadruplex formation can be found in a recent review.[29]

Aside from sequence modification, another common approach to reduce the structural polymorphism is by changing the solution conditions. The presence of biological molecules[30] (e.g., sugar, proteins), cosolvents (e.g., acetonitrile, PEG),[31−33] the use of divalent versus monovalent cations,[34,35] and cation concentration[36−38] all play a major role in determining G-quadruplex stability. The parallel topology of hTel22 clearly illustrates the effect of experimental conditions on G-quadruplex formation. This topology was first reported in potassium conditions as a crystal structure.[20] It was later determined that this is not the predominant topology in solution[33,39,40] and accounts for only about 14% of the total G-quadruplex structures.[33] However, under the effect of dehydration[32] or in the presence of PEG[33] (both factors present in the crystallization conditions), the parallel topology is enriched to become the predominant form. In fact, as proof of this principle, a recently reported NMR solution structure[41] for hTel22 in 40% poly(ethylene glycol) (PEG 200) was similar to the previously reported crystal structure.

The unintended consequence of sequence modification or alteration of experimental conditions is that such an approach can result in an unpredictable perturbation of the system and the selection of a topology which may or may not be representative of the original ensemble of topologies.[42] However, these approaches for the artificial reduction of the structural polymorphism are used because of the limitations of traditional biophysical methods with regard to elucidating G-quadruplex structure. Low-resolution spectroscopy methods, such as circular dichroism or UV−vis spectroscopy, usually cannot distinguish between the different topologies within the ensemble,[38,43] while high-resolution structural methods, such as NMR spectroscopy and X-ray crystallography, are often of limited utility when it comes to resolving the structural polymorphism of G-quadruplexes. As observed with hTel22, a definitive structure cannot be obtained by NMR spectroscopy because this sequence exists as a mixture of multiple G-quadruplex species in solution.[1] The alternative method, X-ray crystallography, may not be appropriate either, as, under dehydrating crystallization conditions, this sequence adopts a topology that may not be representative of the ensemble in solution[33] or *in vivo*.[40] This has a significant effect on what structure or structures can be claimed as "biologically relevant". Thus, there is a need for new experimental approaches that can explore the conformational space surrounding the G-quadruplex topologies without significantly disrupting or perturbing the system.

We propose an alternative approach for the unperturbed investigation of G-quadruplex structures, hydrodynamic bead modeling (HBM). HBM has emerged as a useful tool for studying biological macromolecules and their complexes for which high-resolution structural data are either unavailable or ambiguous.[44] HBM has been used, in a limited scope, to study G-quadruplex structures. In 1999, Niermann et al. used HBM to calculate the rotational and translational diffusion coefficients for the Watson−Crick double helical B-DNA, the single-stranded duplex "hairpin", and the tetramolecular G-quadruplex structures.[45] In 2005, Li et al. employed HBM to demonstrate that the crystal structure of the hTel22 telomere sequence in potassium is not the predominant topology in solution.[39] More recently, Petraccone et al. used HBM to study higher-order G-quadruplex formation by the human telomeric sequence $(T_2AG_3)_nT_2$ (n = 4, 8, 12).[46−48] The purpose of the current work is to use hydrodynamic bead modeling in tandem with molecular dynamics (MD) simulations to explore the structural polymorphism of the human telomere G-quadruplex sequence. In particular, we exploited recent advances in computing hardware,

dx.doi.org/10.1021/jp502213y | J. Phys. Chem. B 2014, 118, 5390−5405

which makes it feasible for routine microsecond-time scale simulations through traditional MD methods. Compared to shorter nanosecond simulations, longer simulations are better at sampling conformations while avoiding the bias of the starting structures.[49] Using MD, we explored the conformational space surrounding the five different folding topologies associated with the human telomere sequence. HBM was used to calculate sedimentation coefficients ($s_{20,W}$) and other hydrodynamic properties for "snapshot" structures obtained by MD simulations. Overall, the calculated hydrodynamic values agreed with experimental values obtained via analytical ultracentrifugation (AUC) studies. Clustering of the MD trajectories using $s_{20,W}$ values revealed the existence of different hydrodynamic substates within the ensemble of structures. Using principal component analysis (PCA) of the MD trajectory, the key motions of the G-quadruplex structures that are responsible for variations observed by hydrodynamic measurements were identified. Grid mapping of water and cations around the DNA also provides valuable insights into the role of hydration and ion binding in hydrodynamic measurements. Lastly, a novel use for HBM is proposed to estimate the number of counterions bound to a particular G-quadruplex structure when accurate information about the size and shape of the DNA is known. This work demonstrates that hydrodynamic bead modeling in conjunction with MD simulations is a powerful technique to study G-quadruplex structures.

EXPERIMENTAL METHODS

Molecular Dynamics (MD) Simulation. Molecular models of G-quadruplex structures were obtained from the Protein Data Bank using the PDB IDs in Table 1. For structures containing multiple models, the first model in the file was selected for AMBER MD simulations. Appropriate coordinating ions were added to the stacked G-tetrads of each model, and additional ions were added to neutralize the G-quadruplex structures. The system was solvated in a truncated octahedral box of TIP3P water molecules with 10 Å buffer. The system was heated and equilibrated using the following protocol: (i) minimize water and ions (1000 steps − 500 steepest descents) holding the DNA fixed (50 kcal/mol/Å), (ii) 50 ps MD (heating to 300 K) holding the DNA fixed, (iii) repeat step i, (iv) minimize all atoms (2500 steps − 1000 steepest descents), (v) repeat step ii, (vi) 50 ns MD (T = 300 K) equilibration holding the DNA fixed (50 kcal/mol/Å), and (vii) 50 ns MD to finish the equilibrium period. Production runs of 1 μs after the final equilibration step were carried out to obtain snapshots at 100 ps interval for a total of 10 000 snapshots. Simulations were performed in the isothermal isobaric ensemble (P = 1 atm, T = 300 K) using sander and GPU version of pmemd. Periodic boundary conditions and particle-mesh-Ewald algorithms were used. A 2.0 fs time step was used with bonds involving hydrogen atoms frozen using SHAKE. Analysis of the trajectory was performed using the *cpptraj* module of the AmberTools 13 Package. Calculations of hydrodynamic properties were done using the program HYDROPRO. All AMBER and HYDROPRO calculations were conducted in part using the resources of the University of Louisville's research computing group and the Cardinal Research Cluster.

Oligonucleotide Preparation and Annealing. The human telomere G-quadruplex-forming oligonucleotide, $dAG_3(T_2AG_3)_3$, and its derivatives (Table 1) were purchased from Integrated DNA Technologies (Coralsville, IA). A stock solution (1 mM) of each oligonucleotide was prepared by dissolving the lyophilized DNA in TBAP buffer (10 mM tetrabutylammoniumphosphate monobasic, 1 mM EDTA, pH 7.0). The DNA was quantified using a Nanodrop 2000 instrument (Thermo Scientific, Wilmington, DE). The molar extinction coefficient (ε) for each oligonucleotide was calculated via the nearest-neighbor method. Prior to sedimentation velocity experiments, the DNA samples were diluted in TBAP buffer to an A_{260} value of 0.5, and salt was added to the solution to bring the final concentration of NaCl or KCl to 400 mM. The oligonucleotide samples were annealed in a water bath by heating to 100 °C, holding the sample at temperature for 10 min, and gradually cooling to room temperature overnight.

Analytical Ultracentrifugation (AUC). AUC was carried out in a Beckman Coulter ProteomeLab XL-A analytical ultracentrifuge (Beckman Coulter Inc., Brea, CA) at 20.0 °C overnight at 50 000 rpm in standard 2 sector cells. Data were analyzed using the program Sedfit (www.sefit.com). The discrete noninteracting species model in Sedfit was used to determine experimental values for $s_{20,W}$ and $D_{t20,W}$ for comparison with HYDROPRO. The concentration-dependent distributions of sedimenting species were calculated using the $c(s)$ continuous distribution model using measured values for buffer density and viscosity. Buffer density was determined on a Mettler/Paar Calculating Density Meter DMA 55A at 20.0 °C, and viscosity was measured using an Anton Parr AMVn Automated Microviscometer. For the calculation of frictional ratio, 0.55 mL/g was used for partial specific volume and 0.3 g/g was assumed for the amount of water bound.

Calibration of HYDROPRO Parameters. The atomic element radius (AER), or bead size, of the primary HYDROPRO hydrodynamic model was calibrated to the Stokes radius of the DNA obtained by AUC. In HYDROPRO, the Stokes radius is reported as the equivalent translational radius, defined as the radius of the sphere with equivalent translational diffusion coefficient value. The calculated values were taken as an average over the number of poses available in the Protein Data Bank. For 143D, 6 poses were deposited in the PDB record. For 2GKU, 12 poses were deposited. For 2HY9, 2JPZ, 2JSL, 2JSM, and 2KF8, 10 poses were deposited. The experimental values were determined as previously described. The calculated values were fitted to the experimental values using a global-fit approach (eq 1) described in the latest HYDROPRO calibration report.[50]

$$\Delta^2 = \frac{1}{N_{\text{G-Quadruplex}}} \sum^{N_{\text{G-Quadruplex}}} \left[\frac{\text{calculated} - \text{experimental}}{\text{experimental}} \right]^2 \quad (1)$$

The Δ value is the root-mean-square relative difference between the calculated values and the experimental values, with 100Δ representing the percent difference typically used to characterize the goodness of prediction.

Hierarchical Agglomerative Cluster Analysis. Cluster analysis was performed using the *cpptraj* module of the AmberTools 13 Package. In hierarchical agglomerative clustering, each data point began in its own cluster and the two closest clusters were merged into a new cluster following after one run of the clustering iteration. The clustering process stopped when a certain number of clusters remained. In order to determine the optimum number of clusters, cluster analyses were performed until one to nine clusters remained. For each cluster analysis, ANOVA was employed to calculate the sums of squares. The percent of variance explained by the clustering was determined by dividing the regression sum of squares by the total sum of squares. The cluster number corresponding to the greatest

dx.doi.org/10.1021/jp502213y | J. Phys. Chem. B 2014, 118, 5390−5405

increase in the percent of variance explained was taken as the optimum number of clusters. For instance, if the greatest increase in percent variance occurred when the number of clusters increased from two to three, three was designated the optimum cluster number. Following cluster analysis, the "snapshot" structures from each cluster were extracted for further investigation.

Free Energy Calculation. Free energy calculations were performed using the MMPBSA module of the AmberTools 13 Package. Solvation free energies were estimated with a nonlinear Poisson−Boltzmann electrostatic continuum method with a hydrophobic component from a surface-area-dependent term. The solute dielectric constant was $\kappa = 1$. A cubic lattice with linear dimensions ~50% larger than the longest dimension was applied with 0.25 Å grid spacing; potentials at the boundaries of the finite-difference lattice were set to a sum of Debye−Huckel potentials. Salt effects were set to be 0.400 M, which corresponded to the salt concentration of sodium and potassium used in the sedimentation velocity experiments. To estimate the nonpolar contributions to solvation, $\Delta G_{nonpolar}$, the solvent accessible surface area (SASA) algorithm of Sanner was used in a parametrization where $\Delta G_{nonpolar} = \gamma(SASA) + \beta$, where $\gamma = 0.00542$ kcal/Å^2 and $\beta = 0.92$ kcal/mol.

Principal Component Analysis (PCA). PCA was performed using the *cpptraj* module of the AmberTools 13 Package. The eigenvectors and eigenvalues were calculated from the diagonalization of the covariance matrix which contained the atomic positional fluctuation, about the average structures, in Cartesian coordinate spaces for all three coordinate axes. The covariance matrix is a $3N \times 3N$ matrix with $3N - 6$ eigenvectors possible, where N is the number of atoms in the system.[51] The eigenvectors describe the nature of the fluctuation, while the eigenvalues describe the contribution of each eigenvector to the overall atomic fluctuation. The results were the partition of the atomic positional fluctuations reported in the previous section into individual components in a way that all components are orthogonal to each other and that the first component accounts for the most variance possible and that each subsequent component, in turn, accounts for the highest variance possible while remaining orthogonal to the preceding components.

Radial Distribution Functions (RDFs) and Grid Mapping. RDF and grid mapping were performed using the *cpptraj* module of the AmberTools 13 Package. The number of water molecules bound to the G-quadruplex structure was calculated by integrating the RDF function up to the first minimum, which represents the boundary of the primary hydration shell. Prior to grid mapping, the trajectory was prepared using the autoimage command. The structures were RMS fit to the bases of the G-tetrads. Grid mappings of water and cation distributions were calculated by binning atom positions at 100 ps intervals into 0.5 × 0.5 × 0.5 Å^3 grids over 1 μs duration of the trajectories. In other words, the value of each grid element represents the number of times the coordinates of the center of a particular atom of interest (i.e., water oxygen) were within the 0.5 × 0.5 × 0.5 Å^3 represented by that particular grid element. For 10 000 frames, the expected number of waters per grid element, assuming bulk water density (55.5 M), was 42. The reference density was 6 M for sodium and 4 M for potassium, which represents the solubility of NaCl and KCl in water, respectively. This corresponded to an expected 4.5 sodium atoms and 3.0 potassium atoms per grid element.

RESULTS AND DISCUSSION

Molecular Dynamics Simulations of Telomeric G-Quadruplex Structures. The aim of this research was the detailed evaluation and parametrization of HBM in the study of G-quadruplex structural polymorphism. To that end, MD simulations were performed to sample the conformational spaces around the human telomere G-quadruplex structures and to produce "snapshot" structures for subsequent hydrodynamic calculations. Simulations were performed on 10 different G-quadruplex structures representing the five folding topologies (Table 1): the antiparallel "basket" in sodium, the parallel "propeller", the mixed "hybrid-1", the mixed "hybrid-2", and the antiparallel "basket" in potassium. For the parallel "propeller" topology, the crystal structure originally reported by Parkinson et al.[20] was used. In addition, simulations were performed on two separate models of the 2KKA structure.[26] The 2KKA-I model, which contains an inosine substitution for guanine at position 14, is the model deposited in the PDB database. 2KKA-G is a model created by changing the inosine residue back to guanine in order to study the effect of inosine substitution on G-quadruplex formation.

The best-fit root-mean-square deviations (RMSDs) over the full 1 μs of the MD trajectories for the G-quadruplexes indicated that the stem structures are rigid and the loop and flanking structures are more flexible (Figure 1). The stem of two or three

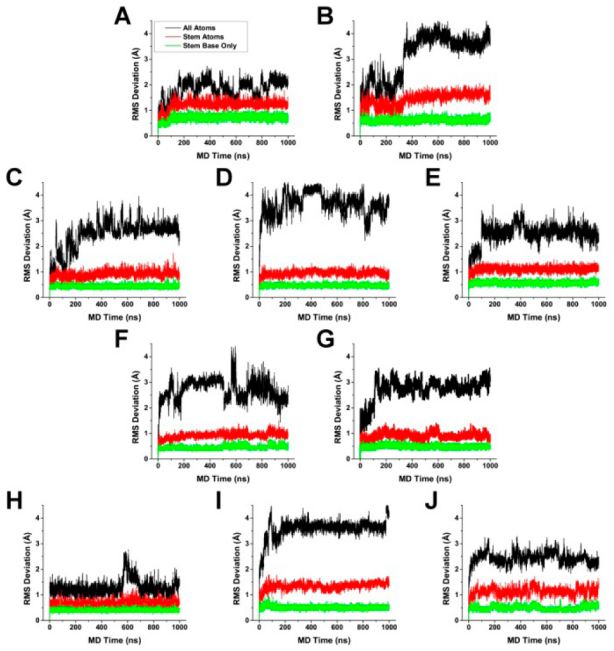


Figure 1. Root mean square deviations with fitting, compared to the first frame, for 143D (A), 1KF1 (B), 2GKU (C), 2HY9 (D), 2JSM (E), 2JPZ (F), 2JSL (G), 2KF8 (H), 2KKA-G (I), and 2KKA-I (J). RMSD calculations were carried out for all non-hydrogen atoms in the G-quadruplex structures (black), in the stacked G-tetrads only (red), and only for G-tetrad guanine bases excluding sugars and phosphate groups (green).

stacked G-tetrads, held together by Hoogsteen hydrogen bonds and supported by the $\pi-\pi$ stacking interaction between adjacent G-tetrads and the electrostatic interaction between the centrally coordinated cations, was the more rigid structural feature (average RMSD of 1.0 Å). The lower RMSD values that were observed for just the guanine bases alone (average RMSD of 0.5

dx.doi.org/10.1021/jp502213y | J. Phys. Chem. B 2014, 118, 5390−5405

Å) compared to the complete stem can be attributed to the more flexible phosphate backbone. Compared to just the stem alone, the RMSD for the G-quadruplex as a whole was much higher, with RMSD values as high as 4.5 Å. To highlight the mobility of different structural components in the G-quadruplex structures, the atomic positional fluctuations were calculated on a per-atom and per-residue basis (Supporting Information, Figures S1 and S2) and mapped to representative structures from the MD trajectories (Figure 2). The fluctuation calculation is a better

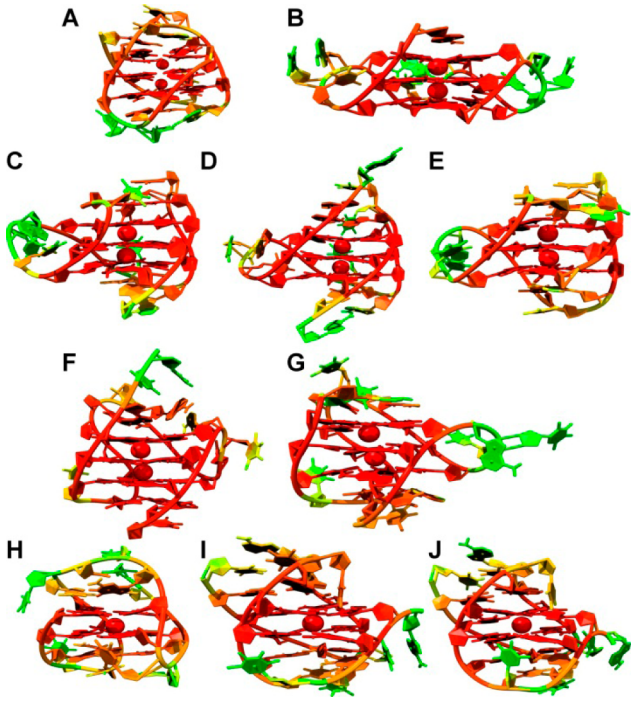


Figure 2. Representative structures from MD trajectories for 143D (A), 1KF1 (B), 2GKU (C), 2HY9 (D), 2JSM (E), 2JPZ (F), 2JSL (G), 2KF8 (H), 2KKA-G (I), and 2KKA-I (J). Residues and atoms are colored by atomic fluctuation units calculated by the AMBER *cpptraj* module with the color gradient progressing from low fluctuation values to high fluctuation values (red to green, respectively).

indicator of mobility, whereas the RMSD calculation, which measures the deviation from a reference set of coordinates, is a better indicator of individual substates. The stacked G-tetrad stems were observed to remain remarkably rigid during the course of the simulation (Figure 2, colored red and orange) with an average positional fluctuation of about 0.5 Å (Supporting Information, Figures S1 and S2). This observation agreed with previously reported MD results[52] and helps explain the stability of these structures *in vitro*.[42,53,54] In contrast to the rigid stems, the loop and flanking bases were found to be more flexible (Figure 2, colored yellow and green). The thymine loop residues were observed to be more mobile than the adenine loop residues, which have been previously shown to have possible stacking interactions with the G-tetrad stem.[52,55] In addition, different degrees of fluctuation were observed for different loop types. The diagonal loops, which transverse the G-tetrad and are more likely to stack with the G-tetrad bases, appeared more rigid (Figure 2A, H, and I) compared to the chain-reversal loops (Figure 2B–G). These findings regarding loop dynamics have significant implications for ligand recognition, as it is theorized that interaction with loop and flanking bases contributes to a ligand's G-quadruplex selectivity.

Determination of Hydrodynamic Values by Sedimentation Velocity Experiments. MD simulations provided a detailed picture of the G-quadruplex structures *in silico*. Sedimentation velocity experiments were carried out by AUC to obtain experimental information about the G-quadruplex structures. The 2KKA-I sequence was excluded because of the inosine substitution in order to limit experimental measurements to sequences containing only canonical bases. AUC experiments were performed for the unsubstituted 2KKA-G sequence, which served as an experimental reference for both the 2KKA-G and 2KKA-I MD models. As the data will show, the calculated hydrodynamic values for the 2KKA-I model agreed with the experimental values for the 2KKA-G sequence, suggesting that the inosine substitution did not substantially alter the hydrodynamic behavior of the 2KKA-G sequence in solution. Sedimenting mixtures of polyelectrolytes often exhibit nonideal behaviors due to the electrostatic interaction between different charged components, resulting in smaller measured molecular weights and sedimentation coefficients.[56] To reduce the effect of nonideality, AUC experiments were carried out in high salt (400 mM NaCl/KCl) buffers to "swamp out" the electrostatic interactions. The distributions of sedimenting species are shown in Figure 3 (black lines). The Stokes radius (a_T),

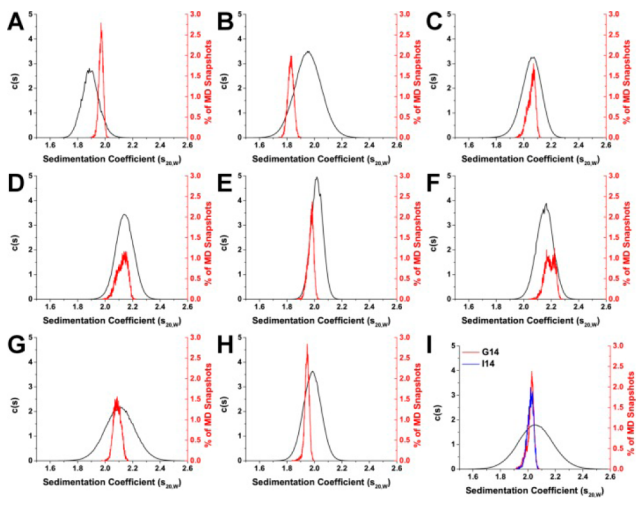


Figure 3. Comparison of experimentally determined and HYDROPRO calculated sedimentation coefficient distributions for 143D (A), 1KF1 (B), 2GKU (C), 2HY9 (D), 2JSM (E), 2JPZ (F), 2JSL (G), 2KF8 (H), and 2KKA (2KKA-G, red; 2KKA-I, blue) (I). For each G-quadruplex structure, sedimentation coefficients ($s_{20,W}$) were determined experimentally by AUC (black) and calculated from MD "snapshots" using HYDROPRO (red and blue).

molecular weight (MW), and frictional ratios (f/f_0) were calculated from the diffusion ($D_{t20,W}$) and sedimentation coefficients ($s_{20,W}$) (Table 2). The frictional ratio is a dimensionless value comparing the observed translational diffusion coefficient of a macromolecule with the translational diffusion coefficient of an equivalent sphere of the same molecular weight with the higher frictional ratios being indicative of a more asymmetric and less spherical shape.[57,58] The value for f is determined from the experimental measurement of $D_{t20,W}$ using eq 2:

$$f = \frac{RT}{N_A D_{t20,W}} \quad (2)$$

dx.doi.org/10.1021/jp502213y | J. Phys. Chem. B 2014, 118, 5390–5405

Table 2. Hydrodynamic Values Determined by Sedimentation Velocity Experiments

G-quadruplex	$s_{20,W}$ [a,b]	$D_{t20,W}$ [a,b]	a_T [b,c]	MW [d]	f/f_0 [e]
143D	1.90	1.42	1.50	7199.97	1.10
1KF1	1.97	1.45	1.48	7330.24	1.07
2GKU	2.05	1.45	1.48	7644.45	1.06
2HY9	2.14	1.44	1.49	8039.62	1.05
2JSM	2.02	1.61	1.33	6779.45	1.00
2JPZ	2.15	1.45	1.48	8049.02	1.05
2JSL	2.11	1.41	1.52	8113.66	1.07
2KF8	1.98	1.54	1.39	6975.38	1.03
2KKA-G	2.06	1.58	1.35	7021.38	1.02

[a] $s_{20,W}$ and $D_{t20,W}$ were calculated from sedimentation velocity data using the discrete noninteracting species model of Sedfit. [b] $s_{20,W}$ expressed in units of 10^{-13} s, $D_{t20,W}$ expressed in units of 10^{-6} cm^2/s, a_T (Stokes radius) expressed in units of 10^{-7} cm. [c] a_T calculated from $D_{t20,W}$, solvent viscosity, and temperature. [d] MW is calculated from $D_{t20,W}$, solvent density, temperature, and 0.55 mL/g value for partial specific gravity. [e] f/f_0 calculated using 0.3 g/g water bound and 0.55 mL/g value for partial specific gravity.

where R is the gas constant and N_A is Avogrado's number. The value for f_0 is determined from the molecular weight using eq 3:

$$f_0 = 6\pi\eta_{20,W}\left(\frac{3M\left(\bar{v} + \frac{\delta}{\rho_{20,W}}\right)}{4\pi N_A}\right)^{1/3} \quad (3)$$

where $\eta_{20,W}$ is the viscosity of water at 20 °C, $\bar{v}$ is the partial specific volume, δ is a uniform expansion factor to account for hydration, $\rho_{20,W}$ is the density of water at 20 °C, and M is the molecular weight determined by the Svedberg equation (eq 4):

$$M = \frac{s_{20,W}RT}{D_{20,W}(1 - \bar{v}\rho_{20,W})} \quad (4)$$

From these equations, it becomes apparent that the frictional ratio of a macromolecule is dependent on its shape, flexibility, and the amount of hydration associated. In the current work, the amount of associated hydration or δ was assumed to be 0.3 g of water/g of G-quadruplex. For nucleic acids, a value of 0.3−0.35 g/g is typically used, although this parameter can be quite difficult to determine with accuracy.[59]

For each of these sequences, the G-quadruplex structures sedimented as one species (Figure 3), including the hTel22 sequence in potassium and the 2KKA-G sequence, which are known to exist as a mixture of multiple G-quadruplex species.[26,27] The inability of AUC to resolve the existence of different G-quadruplex species in solution can be attributed to the low-resolution nature of the technique. Overall, the G-quadruplex structures formed from different telomeric sequences all assumed similar spherical shapes (f/f_0 = 1.00−1.10). The G-quadruplex structures formed by the hTel22 sequence in sodium (143D) sedimented at a lower rate ($s_{20,W}$ = 1.90) than G-quadruplex structures in potassium (1KF1) ($s_{20,W}$ = 1.97), which indicated the sodium form is less compact than the potassium form. Since the sequences are identical in both cases, the difference in sedimentation might be attributed to a change in shape. In fact, the sodium form appeared more elongated (f/f_0 = 1.10) when compared to the potassium form (f/f_0 = 1.07). It is important to note, however, that the difference between the two sequences might not be distinguishable, as it falls within the precision limit of 5% for the experimental technique. For the 2JSM, 2KF8, and 2KKA-G sequences, the changes in sedimentation were also attributed to differences in shape. These sequences have lower mass compared to the hTel22 sequence but are more compact in shape (f/f_0 = 1.00−1.03) and thus sedimented at a higher rate ($s_{20,W}$ = 1.98−2.06). The differences in frictional ratios of these sequences compared to the 1KF1 sequence were distinguishable, as these differences are greater than the 5% precision limit of the experimental technique. In contrast, the changes in sedimentation for the 2GKU, 2HY9, 2JPZ, and 2JSL sequences compared to the hTel22 sequence can be attributed mainly to a difference in sequence sizes. These sequences appeared to be of similar shape or more compact (f/f_0 = 1.05−1.07) yet sedimented at a higher rate ($s_{20,W}$ = 2.02−2.15) due to the increased sequence molecular weight. Taken together, these findings indicated that AUC can be informative when used to study G-quadruplex structures.

Calculation of Hydrodynamic Properties Using the Program HYDROPRO. To determine if the *in silico* models can reproduce the hydrodynamic values observed experimentally, HYDROPRO[50] was used to calculate the hydrodynamic values (i.e., $s_{20,W}$, D_{t20W}) for each "snapshot" structure obtained from MD simulations. Prior to calculating the hydrodynamic properties, the optimum size of the beads in the primary hydrodynamic model was determined using the procedure previously described.[50] In brief, the size of the beads (termed the atomic element radius or AER in the HYDROPRO input file) was varied and the AER that yielded the smallest difference between Stokes radius calculated by HYDROPRO (a_T) and the Stokes radius determined by AUC experiments was accepted as the optimum bead size (Table 3 and Supporting Information, Figure S3). For the atomic-level model, where each non-hydrogen atom is replaced with the beads, the physical meaning of AER can be thought of as the radius of a typical non-hydrogen atom plus a uniform expansion to account for hydration. For the residue-level calculations, where each residue (or nucleotide) is replaced with a bead, the physical meaning of AER is the size of the nucleotide plus a uniform expansion to account for hydration. The calibration was done using previously reported NMR structures (PDB ID: 143D, 2GKU, 2HY9, 2JPZ, 2JSL, 2JSM, and 2KF8). All structural poses deposited in the PDB database were used for HYDROPRO calibration. For the first calculation mode where

Table 3. Results of HYDROPRO Global Fit Analysis

calculation mode	AER (Å)	% difference a_T	% difference $s_{20,W}$	% difference $D_{t20,W}$
atom, shell model	2.19	2.35	1.19	2.24
atom, shell model	2.84	6.32	4.27	5.83
residue, shell model	3.98	2.50	1.39	2.42
residue, shell model	4.84	7.78	5.59	7.12
residue, bead model	5.04	2.60	1.41	2.50
residue, bead model	6.11	8.10	5.86	7.38

dx.doi.org/10.1021/jp502213y | J. Phys. Chem. B 2014, 118, 5390−5405

the non-hydrogen atoms were replaced by a collection of overlapping spheres and the hydrodynamic values were calculated using a shell-model methodology,[60] the best-fit AER was determined to be 2.19 Å (Table 3) with a difference of 2.35% between the predicted hydrodynamic values and the experimental values. For a non-hydrogen atom, the atomic radius is approximately 1.8 Å.[50] The determined best-fit AER suggests a hydration sphere of about 0.4 Å which is less than the 1.1 Å typically used in hydrodynamics calculations to estimate hydration.[61] When the default AER (2.84 Å) was used instead of the best-fit AER, the difference increased from 2.35 to 6.32% (Table 3). For the second calculation mode where each residue is replaced by a sphere instead of each atom and the hydrodynamic values are also calculated using a shell-model methodology, the best-fit AER was 3.98 Å with an difference of 2.50% compared to the 4.84 Å for the standard AER (7.78% difference). For the third calculation mode, which calculates the hydrodynamic values directly on the primary bead models from the second calculation mode, the best-fit AER was 5.04 Å with a difference of 2.60% compared to 6.11 Å for the standard AER (8.10% difference). Overall, the calibrated AERs were able to reproduce the experimental hydrodynamic values better than the standard default AERs. While the three modes to calculate hydrodynamic properties provide the same general information about the structure being modeled, the residue-level models were much faster in the amount of time required for calculation and could be useful for preliminary analysis of large MD simulations where thousands of structures are sampled.

Ten thousand "snapshot" structures were sampled from each MD trajectory, and hydrodynamic properties were calculated for each structure using the calibrated AER values. The average calculated hydrodynamic values for each of the MD simulations are shown in Table 4. The hydrodynamic values calculated using MD snapshots agreed with previously calculated values using reported NMR structures with a difference of 1.85% for the atomic-level/shell-model mode, 1.89% for the residue-level/shell-model mode, and 1.88% for the residue-level/bead-model mode (Supporting Information, Table S1). In general, all three modes of calculation by HYDROPRO were able to accurately predict the hydrodynamic values of the G-quadruplex structures (Table 5). The distributions of the $s_{20,W}$ values calculated using the HYDROPRO atomic-level model shell-model calculation mode (Figure 3, colored lines) are shown with the concentration-dependent $c(s)$ experimental $s_{20,W}$ distributions (Figure 3, black line). For the $c(s)$ experimental $s_{20,W}$ distributions, the width of the distribution curve is related to the homogeneity of species in solution.[62,63] Samples containing multiple species that sediment at similar rates will have broad $s_{20,W}$ distributions, while highly homogeneous samples with a single predominant species (accounting for more than 90% of the total ensemble) will have narrow $s_{20,W}$ distributions. In general, it was observed that sequences (143D, 2GKU, 2HY9, 2JPZ, 2JSL, and 2KF8) for which there was enough enrichment of a major species for NMR structure elucidation have narrower distributions compared to sequences for which NMR structure elucidation was not possible (1KF1 and 2KKA). The only exception was the 2JSM sequence, which was highly enriched (>70% of the total ensemble)[23] in a major species yet has a broad $s_{20,W}$ distribution. Electrostatic interaction, which contributed to the nonideal behaviors of the DNA at lower salt concentrations, does not affect the width of the distributions but will shift the distribution to the left (lower apparent $s_{20,W}$). At higher concentrations of ions, the hydrodynamic behavior became more ideal, the distribution shifted to the right, and the apparent $s_{20,W}$ approached the true $s_{20,W}$. Compared to the experimental distributions, the calculated $s_{20,W}$ distributions were narrower and in certain cases were asymmetric. The shape of the calculated distributions is determined by the sampling capability of MD simulations which is on a much shorter time scale (1 μs) compared to the sampling capability of AUC experiments (~6 h). The time-course graph of $s_{20,W}$ values for the 2JPZ model highlights the effect of sampling time on the distribution peak shape (Figure 4F). The distribution of calculated $s_{20,W}$ values across the full 1 μs of MD simulation was bimodal (Figure 3F). However, had the simulation completed within the first 500 ns, the distribution would have been narrower and unimodal.

With the exception of the hTel22 sequence in potassium (1KF1), the HYDROPRO calculated hydrodynamic values agreed with the experimental values (Table 5). As hydrodynamic

Table 4. Summary of HYDROPRO Calculated Hydrodynamic Values

G-quadruplex	atom/shell	residue/shell	residue/bead
143D	$s_{20,W}$[a] = 1.97 ± 0.02	1.94 ± 0.01	1.92 ± 0.01
	$D_{t20,W}$[a] = 1.51 ± 0.01	1.49 ± 0.01	1.48 ± 0.01
1KF1	$s_{20,W}$ = 1.83 ± 0.02	1.89 ± 0.01	1.87 ± 0.03
	$D_{t20,W}$ = 1.40 ± 0.02	1.45 ± 0.03	1.44 ± 0.03
2GKU	$s_{20,W}$ = 2.06 ± 0.03	2.03 ± 0.01	2.02 ± 0.01
	$D_{t20,W}$ = 1.45 ± 0.02	1.44 ± 0.01	1.43 ± 0.01
2HY9	$s_{20,W}$ = 2.12 ± 0.04	2.15 ± 0.02	2.14 ± 0.02
	$D_{t20,W}$ = 1.38 ± 0.02	1.40 ± 0.01	1.39 ± 0.01
2JSM	$s_{20,W}$ = 1.97 ± 0.02	1.95 ± 0.01	1.94 ± 0.01
	$D_{t20,W}$ = 1.45 ± 0.02	1.44 ± 0.01	1.43 ± 0.01
2JPZ	$s_{20,W}$ = 2.19 ± 0.04	2.17 ± 0.02	2.17 ± 0.02
	$D_{t20,W}$ = 1.43 ± 0.02	1.42 ± 0.01	1.42 ± 0.01
2JSL	$s_{20,W}$ = 2.09 ± 0.03	2.08 ± 0.01	2.08 ± 0.01
	$D_{t20,W}$ = 1.42 ± 0.02	1.41 ± 0.01	1.41 ± 0.01
2KF8	$s_{20,W}$ = 1.95 ± 0.02	1.91 ± 0.01	1.89 ± 0.01
	$D_{t20,W}$ = 1.49 ± 0.01	1.46 ± 0.01	1.45 ± 0.01
2KKA-G[b]	$s_{20,W}$ = 2.02 ± 0.03	2.00 ± 0.01	1.99 ± 0.01
	$D_{t20,W}$ = 1.49 ± 0.02	1.47 ± 0.01	1.47 ± 0.01
2KKA-I[b]	$s_{20,W}$ = 2.02 ± 0.03	2.00 ± 0.01	1.99 ± 0.01
	$D_{t20,W}$ = 1.49 ± 0.02	1.47 ± 0.01	1.46 ± 0.01

[a] $D_{t20,W}$ expressed in units of 10^{-6} cm^2/s and $s_{20,W}$ expressed in units of 10^{-13} s. [b] Experimental values for 2KKA-G and 2KKA-I were obtained with the oligonucleotide sequence $dAG_3(T_2AG_3)_3T$.

Table 5. Comparison of HYDROPRO-Calculated and Experimental $s_{20,W}$

structure	% diff. atom/shell	% diff. residue/shell	% diff. residue/bead
143D	3.76	1.96	1.19
2GKU	0.30	0.83	1.23
2HY9	0.90	0.35	0.16
2JPZ	1.70	0.95	0.61
2JSL	1.15	1.46	1.68
2JSM	2.23	3.23	3.60
2KF8	1.95	3.95	4.59
1KF1	7.23	3.80	4.91
2KKA-G[a]	1.88	2.67	3.11
2KKA-I[a]	1.98	2.98	3.45

[a] Experimental values for 2KKA-G and 2KKA-I obtained with the oligonucleotide sequence $dAG_3(T_2AG_3)_3T$.

dx.doi.org/10.1021/jp502213y | J. Phys. Chem. B 2014, 118, 5390−5405

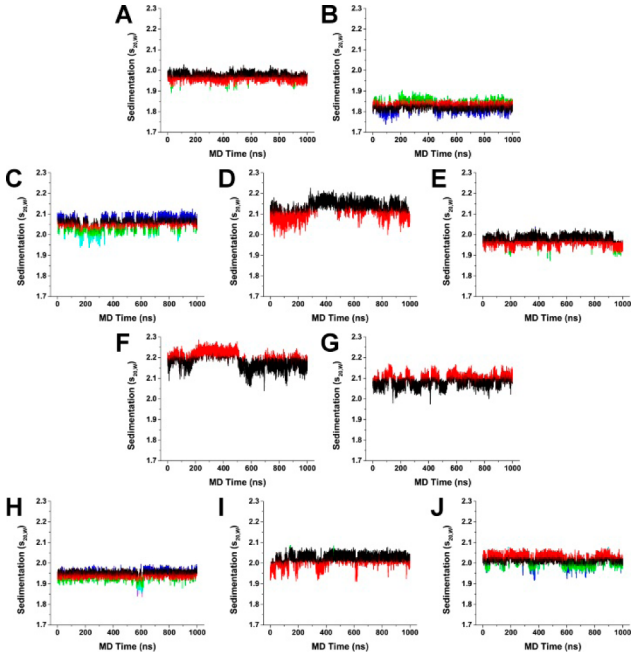


Figure 4. Hydrodynamics substates identified by clustering of HYDROPRO calculated sedimentation coefficients (S20,W) for 143D (A), 1KF1 (B), 2GKU (C), 2HY9 (D), 2JSM (E), 2JPZ (F), 2JSL (G), 2KF8 (H), and 2KKA (I). The most populated cluster for each model was colored black, and the second most populated cluster was colored red. Additional clusters were colored accordingly (green, blue, cyan, and magenta) in the order of decreasing population density.

experiments are essentially low-resolution measurements, a difference between calculated and experimental values by 5% (the precision limit for hydrodynamic measurements[59]) can be considered acceptable. The % difference in Table 5 was determined by comparing the HYDROPRO calculated values from Table 4 with the experimental values in Table 2 using eq 1. For the seven models used in the HYDROPRO calibration procedure, the calculated distributions agreed with the experimental distributions (0.30−3.76% difference), as shown in Table 5. With the 143D model, it appeared initially that the model did not agree with the experimental data (Figure 3A); however, the difference of 3.76% (Table 5) was still within the experimental precision limit.

Of the three models which were excluded from the calibration procedure, the calculated distribution of 1KF1 did not agree with the experimental distribution with a difference of 7.23% (Table 5). The current findings agreed with previously reported data, which demonstrated that the parallel form of 1KF1 has different hydrodynamic behaviors compared to the ensemble of structures in solution.[39] The conclusion that the 1KF1 structure was distinct from the 1KF1 sequence in solution was made on the basis of the calculated $s_{20,W}$ value alone and not the $D_{t20,W}$ value, which appeared to agree better with the experimental value than the $s_{20,W}$. The accurate experimental determination of $D_{t20,W}$ can be difficult in a solution containing multiple species as the differential migrations of different species can result in a broadening of the sedimentation boundary and an incorrect apparent$D_{t20,W}$.[62,63] For the 1KF1 structures in solution, there is a mixture of at least three to four G-quadruplex species in solution.[33] For that reason, the apparent $D_{t20,W}$ value determined by experimental measurement might not be representative of the structures in solution and was reported only to complete the data even though it might not be appropriate to use it to rule in or rule out a certain structure in solution. For the other two models (2KKA-G and 2KKA-I), the calculated hydrodynamic values differed from the experimental values only slightly (1.88−1.98% difference). It should be noted that the unsubstituted 2KKA-G (1.88%) model agreed better with experimental results than the inosine-substituted 2KKA-I (1.98%) model. However, this difference is too small to draw any definitive conclusion regarding a possible difference between the two models. These findings highlighted a major limitation of HBM, which is that hydrodynamic measurements are inherently low resolution. The model that accurately predicts the hydrodynamic parameters measured in solution is not necessarily representative of the definitive structure for that molecule but rather one of the possible conformations among many others.[44] Thus, it can be concluded that the 2KKA-G and 2KKA-I models represent a possible set of conformations for the G-quadruplex structures in solution. However, this prediction should be confirmed by nonhydrodynamic experiments (e.g., 2-aminopurine fluorescence spectroscopy to probe the solvent accessibility of adenine bases or DMS footprinting to probe guanine base-pairing interactions). Such experiments were done by Li et al.[39] Overall, the present findings demonstrated that HYDROPRO calculations can accurately predict the correct hydrodynamic properties and can be used as a screening tool to initially rule out "incorrect" structures associated with other G-quadruplex-forming sequences.

Clustering of Molecular Dynamics Trajectories into Hydrodynamic Substates. The results of HYDROPRO calculations revealed that sedimentation rates can differ dramatically between different "snapshot" structures within the same MD trajectory (Figure 4). To identify different hydrodynamic substates within the MD trajectories, hierarchical cluster analysis was performed using HYDROPRO calculated $s_{20,W}$ values as the clustering criterion. The results of the cluster analysis indicated that heterogeneous mixtures of G-quadruplex structures were present in simulations (Supporting Information, Figures S4−S13). Cluster analysis was able to identify key differences between the 2KKA-G model and the 2KKA-I model which were not apparent by $s_{20,W}$ calculation alone (Figure 3I). The 2KKA G-quadruplex structures consist of a two-stack G-tetrad stem that is capped by a triple-base cap and a double-base cap on either end. In the 2KKA-I model, the G-G-I triple-base cap was slightly offset from the G-tetrad bases (Supporting Information, Figure S13), while, in the 2KKA-G model, the G-G-G triple-base cap was directly on top of the G-tetrads (Supporting Information, Figure S12). It was unclear whether the direct triple-base cap conformation is simply a new stable conformation or if it is an indication of the G-quadruplex structure transitioning to a new folding topology, as it is beyond the current capacity of MD to model such a phenomenon. The current findings clearly demonstrated that the substitution of inosine for guanine played a critical role in stabilizing the selected G-quadruplex structures. In addition, these findings raised the need for additional experimental investigation into the role of these noncanonical bases in selecting for a G-quadruplex topology from an ensemble of structures.

In addition to the HYDROPRO calculated hydrodynamic values, free energy values (ΔG) and the number of water bound were also determined for each cluster (Table 6). Free energy was calculated using the MMPBSA method and is the sum of all the molecular mechanics energies (bond, angle, torsion, van der Waals, and electrostatic) and the solvation energy (calculated

dx.doi.org/10.1021/jp502213y | J. Phys. Chem. B 2014, 118, 5390−5405

Table 6. Results of $s_{20,W}$ Cluster Analysis on MD Trajectories

structure	cluster no.	N	$s_{20,W}$[a]	f/f_0	ΔG[a]	no. of water
143D	1	6170	1.98	1.12	−4421.89	165.26
	2	3802	1.95	1.14	−4421.62	166.92
	3	28	1.91	1.16	−4416.97	169.89
1KF1	1	5244	1.82	1.22	−4360.05	174.63
	2	3503	1.84	1.20	−4363.84	173.62
	3	768	1.87	1.19	−4368.20	173.43
	4	485	1.78	1.24	−4357.99	176.97
2GKU	1	4694	2.07	1.14	−4664.48	174.66
	2	2482	2.04	1.15	−4664.43	176.61
	3	1413	2.01	1.17	−4664.29	178.65
	4	1167	2.09	1.13	−4663.50	172.86
	5	244	1.98	1.19	−4664.01	181.28
2HY9	1	6374	2.15	1.16	−5094.00	190.93
	2	3626	2.08	1.19	−5091.86	197.55
2JSM	1	6852	1.98	1.15	−4536.80	175.92
	2	3115	1.95	1.17	−4533.68	177.01
	3	32	1.90	1.19	−4525.48	178.38
	4	1	2.04	1.13	−4545.97	172.00
2JPZ	1	6009	2.16	1.14	−4933.18	188.59
	2	3991	2.23	1.12	−4930.52	180.50
2JSL	1	6681	2.07	1.16	−4811.53	186.55
	2	3319	2.12	1.14	−4817.04	184.74
2KF8	1	5471	1.96	1.14	−4241.88	166.86
	2	3892	1.93	1.15	−4239.20	168.79
	3	411	1.90	1.16	−4224.43	170.41
	4	186	1.98	1.13	−4247.06	164.19
	5	36	1.87	1.18	−4215.53	172.75
	6	4	1.84	1.20	−4205.31	171.50
2KKA-G	1	7863	2.03	1.13	−4434.13	171.78
	2	2135	1.98	1.15	−4427.76	175.58
	3	2	2.09	1.11	−4409.13	179.00
2KKA-I	1	4673	2.01	1.14	−4347.30	172.24
	2	4146	2.04	1.12	−4347.99	170.08
	3	1068	1.98	1.15	−4343.95	173.96
	4	111	1.94	1.17	−4340.41	176.68
	5	2	2.08	1.11	−4331.61	162.00

[a] $s_{20,W}$ is expressed in units of 10^{-13} s; ΔG is expressed in units of kcal/mol.

using a numerical solution of the Poisson−Boltzmann equation).[64] The free energy values did not include an estimate of solute entropy, which is the least reliable component of MMPBSA calculations.[65−67] The results of the cluster analysis demonstrated that G-quadruplex structures remained highly polymorphic even with the limited sampling of MD simulations. In general, it was observed that the more populated substates were more energetically favorable (i.e., lower ΔG). The G-quadruplex structures in these substates were more spherical (i.e., lower f/f_0) and compact (i.e., higher $s_{20,W}$), and were characterized by favorable interactions between the loops and flanking bases with the G-tetrad stems (Supporting Information, Figures S4−S13). For example, in the case of the 2HY9 sequence, which contained a three-residue long 5′ flanking sequence and a two-residue long 3′ flanking sequence, the more compact structures featured stacking interactions between the flanking bases and the G-tetrads, while the less compact structures did not. Previous studies have reported that the folding of a single-stranded DNA structure into the G-quadruplex form was associated with the release of water molecules from the DNA into the surrounding environment.[68,69] It was observed that the more compact structures were also associated with a reduced number of bound water molecules. As hydration plays an important role in ligand binding and recognition,[70−73] for instance, the binding of groove-binding small molecules to duplex DNA is driven by water being displaced from the minor grooves.[74] Therefore, clustering by $s_{20,W}$ values can be used in the process of drug design to identify substates that could interact more favorably with small-molecule inhibitors.

Principal Component Analysis of Molecular Dynamics Trajectories. PCA was employed to identify the major patterns of motions in the MD models and investigate the underlying mechanism for the formation of the different substates identified by clustering analysis. Motions from MD can appear random and chaotic; however, most of the fluctuations associated with the MD models can usually be reduced to several low-frequency eigenvectors with large eigenvalues.[75] The first, second, and third eigenvectors are depicted using "porcupine" plots[76,77] with the arrows representing the magnitude and direction for the eigenvector for each atom (Figure 5 and Supporting Information, Figures S14 and S15). In general, it was observed that the overall dynamics of the G-quadruplex structures were composed of only one to five major movements, as indicated by the large

dx.doi.org/10.1021/jp502213y | J. Phys. Chem. B 2014, 118, 5390−5405

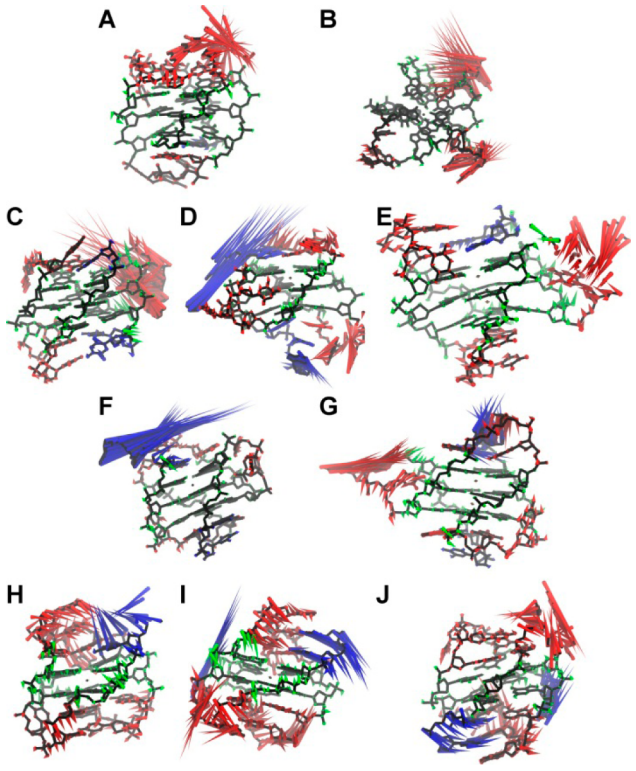


Figure 5. Porcupine plots of the first eigenvectors for 143D (A), 1KF1 (B), 2GKU (C), 2HY9 (D), 2JSM (E), 2JPZ (F), 2JSL (G), 2KF8 (H), 2KKA-G (I), and 2KKA-I (J). Principal component analysis was carried out on MD trajectories in order to determine the major patterns of motions. Motions associated with stem residues are colored green, loop residues are red, flanking residues are blue, and central ions are yellow.

eigenvalues associated with the first several eigenvectors compared to subsequent eigenvectors (Supporting Information, Figure S16). In fact, the first three eigenvectors accounted for 30−80% of the variance in the models (Supporting Information, Figure S17). The porcupine plots indicated that motions related to the more rigid green colored stems (Figure 5 and Supporting Information, Figures S14−S15) tend to be small and localized, while the motions related to the more flexible red loops (Figure 5 and Supporting Information, Figures S14 and S15) and blue flanking bases (Figure 5 and Supporting Information, Figures S14 and S15) were more prominent. Each structural component was defined by characteristic movements. Most of the movements associated with the stem structures were the twisting motions of the phosphate backbone around the quadruple helix stack. The major movements associated with the loop structures were rotation of the phosphate backbone which moved the bases either toward the G-tetrad stem or away from it. In agreement with previous atomic positional fluctuation results, higher magnitude movement was seen with the chain-reversal loops compared to the lateral and diagonal loops. Lastly, the transition between flanking bases stacking on the G-tetrad bases to the unstacking of the flanking bases was observed as the major movement associated with the flanking bases. The movements associated with structures containing longer flanking sequences (i.e., 2HY9 and 2JPZ) were much higher in magnitude. The movements identified by PCA correlated with the structural differences between the different substates previously identified by cluster analysis.

The first 10 eigenvalues are reported (Supporting Information, Figure S16) as calculated for the G-quadruplex structures as a whole and for various structural components (i.e., stem, loops, flanking bases). The eigenvalues were slightly reduced when the more mobile hydrogen atoms were removed from the analysis. Because PCA is sensitive to the scaling of the original data, the eigenvectors and eigenvalues of similar motions should be similar to one another. An overlap of the eigenvalues of loop and flanking bases with the eigenvalues of the overall G-quadruplex structures indicated that overall dynamic motions of G-quadruplex structures are mainly due to motions of loop and flanking bases. To demonstrate this correlation between the dynamics motions of individual structural components and the overall dynamic motions of the G-quadruplex structures, the first principal components were plotted as a function of the second principal components for all 10 MD models for overall G-quadruplex structure and the individual stem, loop, and flanking structures (Supporting Information, Figures S18−S27). From the PCA results, each model can be classified as having loop-dominated dynamics (i.e., 143D, 1KF1, 2GKU, 2JSM), flanking bases-dominated dynamics (i.e., 2HY9 and 2JPZ), and mixed dynamics with contributions from both loop and flanking bases (i.e., 2JSL, 2KF8, 2KKA-G, and 2KKA-I). Overall, the PCA findings help explain the differences in hydrodynamic values for individual substates identified by previous cluster analysis. In addition, the differences observed between the dynamic motions of different loop and flanking bases could have significant implications in drug interaction and the choices of sequence chosen for *in vitro* drug binding studies.

Investigation of Water and Cation Distribution around G-Quadruplex Structures. Given the importance of hydration and cations in influencing G-quadruplex formation, as well as the role of water and ions in determining hydrodynamic properties of a macromolecule, density mapping and radial distribution functions (RDFs) were used to visualize the distribution of these molecules around the G-quadruplex structures and provide a general overview of hydration and cation binding. The RDFs of the distance between oxygen atoms of the water molecules and non-hydrogen atoms on the surface of the G-quadruplex structures indicated the formation of a well-defined first hydration shell between 2.4 and 3.2 Å around the G-quadruplex structures (Figure 6). The second and third hydration shells can also be observed, although these peaks were not as sharp as the peaks for the first hydration shell. To visualize these hydration shells, density mapping of the water molecules over the duration of the trajectory was performed. Water density is shown contoured at equivalent levels (about 2 times the expected bulk water density) around the average structures from the MD trajectories (Supporting Information, Figure S28). The condensation of the water around the G-quadruplex structure was evident. While the water appeared to be evenly distributed around the G-quadruplex structures, a higher density of water was observed around the rigid stem structures and a lower density was observed around the more mobile loops and flanking structures. This observation agreed with a previous study, which reported that rigid structures are associated with more well-defined water positions and higher apparent density.[78] When contoured at a higher level (about 3 times the expected bulk water density), preferential binding of water to the grooves of the G-quadruplex structures became more apparent. At this higher contour level, a "spine of hydration" which traverses the grooves of the G-quadruplex structures was observed (Figure 7). These findings demonstrated the interaction of water molecules with DNA and suggested a role for hydration in maintaining the G-quadruplex structures.

dx.doi.org/10.1021/jp502213y | *J. Phys. Chem. B* 2014, 118, 5390−5405

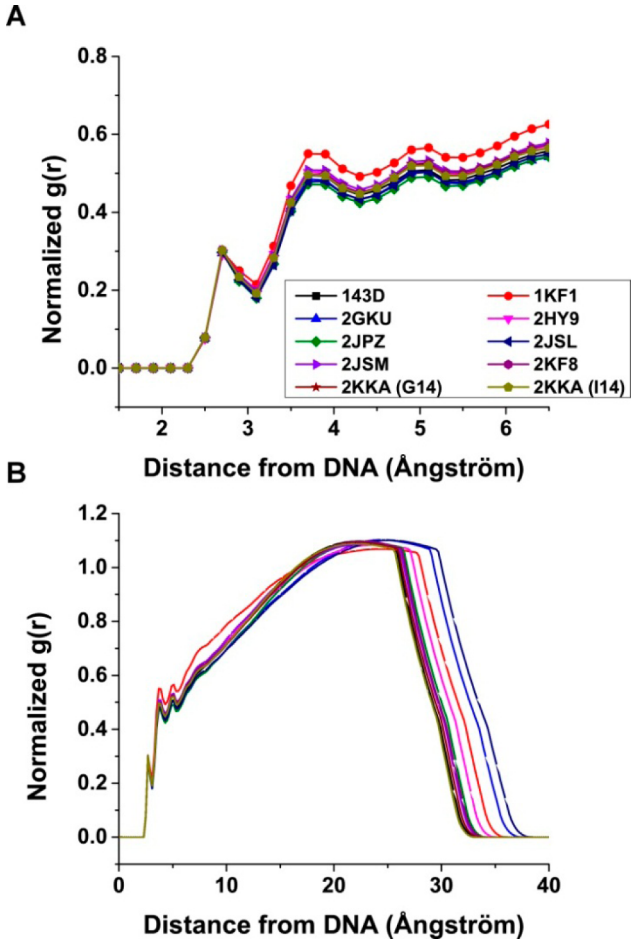


Figure 6. Radial distribution functions (RDFs) between G-quadruplex surface heavy atoms (no hydrogen) and water oxygen atoms. RDFs were calculated for distances from 1.5 to 6.5 Å (A) and for distances from the G-quadruplex surface to the edge of the periodic box.

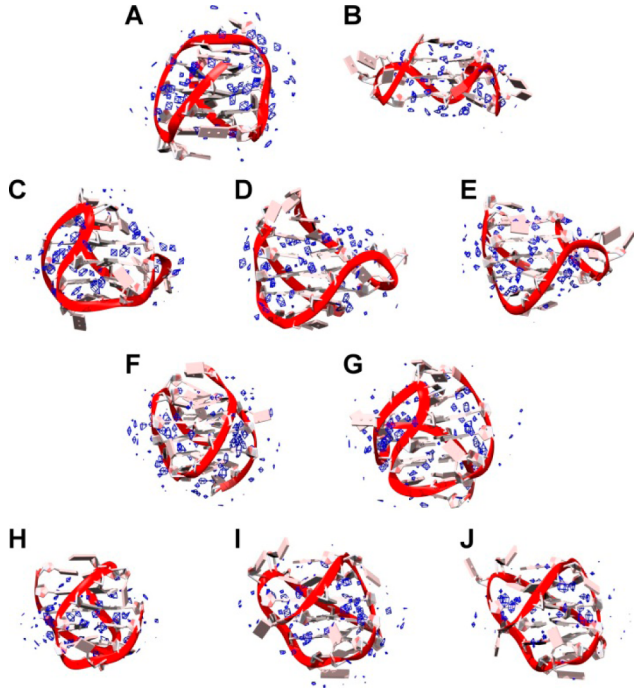


Figure 7. Pseudodensity grid maps of water oxygen atoms for 143D (A), 1KF1 (B), 2GKU (C), 2HY9 (D), 2JSM (E), 2JPZ (F), 2JSL (G), 2KF8 (H), 2KKA-G (I), and 2KKA-I (J). Water density (blue) was contoured at 3× the density (55.5 M) of bulk water. The average structure of each G-quadruplex over the trajectory is shown.

The RDF plots of the distance between cations and G-quadruplex structures indicated cations interacted with G-quadruplex structures differently from water (Figure 8). The RDF plots showed a clear attraction of positively charged cations to the negatively charged nucleic acids with a high density of cations closer to the DNA and decreased density further out. The interaction shells for the sodium ions were closer to the DNA than those for the potassium ions. This observation was expected given that within the force field the sodium ions are smaller (radius = 1.369 Å vs 1.705 Å)[79] and thus can be closer to the G-quadruplex structures. Four cation interaction shells were observed for each structure. The first two interaction shells were well-defined, while the second two interaction shells were more diffused. Density mapping analyses were performed to visualize these cation shells. The density was contoured at the reference density (Supporting Information, Figure S29). At this contour level, two different modes of interaction were observed. The first mode was the coordination of cations within the central G-tetrads. It was observed that the cations remained positioned between the stacked G-tetrads throughout the simulation with no exchange between the G-quadruplex stem and the bulk solvent as observed in other simulations.[80−82] The 143D model of the hTel22 sequence in sodium containing solution is noteworthy, as the starting structure contained three sodium ions placed within the plane of the G-tetrads. However, when the system was equilibrated, one of the sodium ions were ejected from the G-quadruplex stem into the surrounding solvent and the remaining two sodiums assumed positions between the G-tetrads and remained so for the course of the MD trajectory. In addition to the central coordination, diffused interaction of the cations along the grooves of the G-quadruplex structures was also observed. These "spines of cations" were not deep within the grooves as were the "spines of hydration". Instead, they were closer to the backbones and interacted with the backbone phosphate groups. When the cations were contoured at a higher level, these electrostatic interactions began to disappear and a third mode of interaction, external coordination with loops and flanking bases, was observed (Figure 9). These externally coordinated cations were critical in maintaining loop and flanking base positions. For instance, potassium binding was responsible for stabilizing the capping structures in the 2KF8 model. The binding site at the top of the G-quadruplex structure remained 100% occupied with the potassium ion either coordinated within the plane of the triple-base cap or stacked between the planes of the inner triple-base cap and the outer double-base cap. Cation stabilization of loop structures has also been observed in previous simulations[37] as well as experimentally in the crystal structure of the human c-Kit DNA promoter sequence.[83] In that study, one Mg^{2+} and two K^+ ions were observed in the G-quadruplex loops and grooves in addition to the K^+ ions within the canonical central ion channel. In the crystal structure, all three external cations are believed to play a role in maintaining the c-Kit G-quadruplex structure. It should be noted that one of the K^+ ions appeared to be transient and capable of adopting one of several distinct positions, while the other K^+ ion appeared to be more static, suggesting the existence of a high affinity binding site. While the K^+ ions are believed to have a primary role in stabilizing the G-quadruplex structure by

dx.doi.org/10.1021/jp502213y | J. Phys. Chem. B 2014, 118, 5390−5405

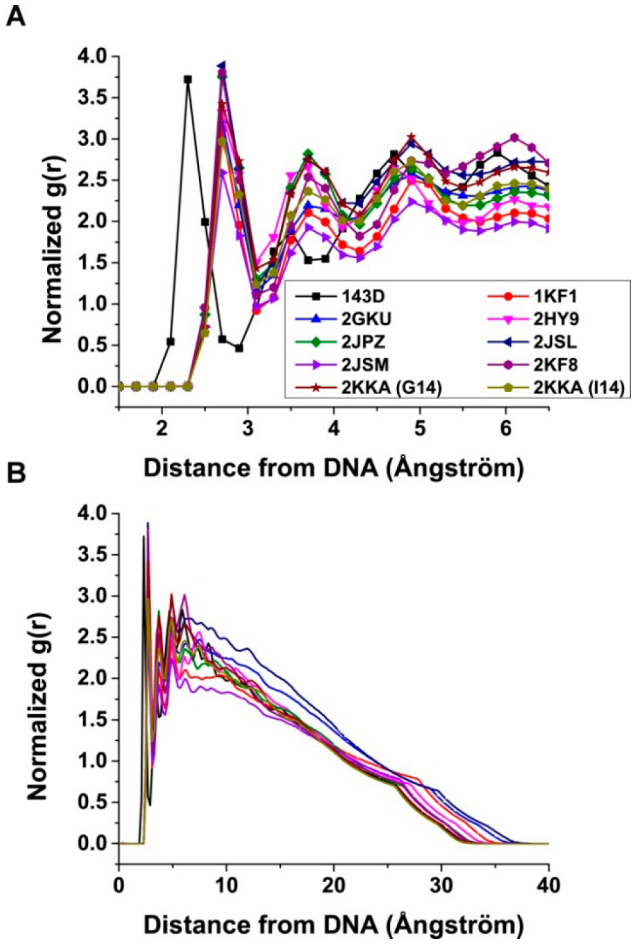


Figure 8. Radial distribution functions (RDFs) between G-quadruplex surface heavy atoms (no hydrogen) and cations. RDFs were calculated for distances from 1.5 to 6.5 Å (A) and for distances from the G-quadruplex surface to the edge of the periodic box.

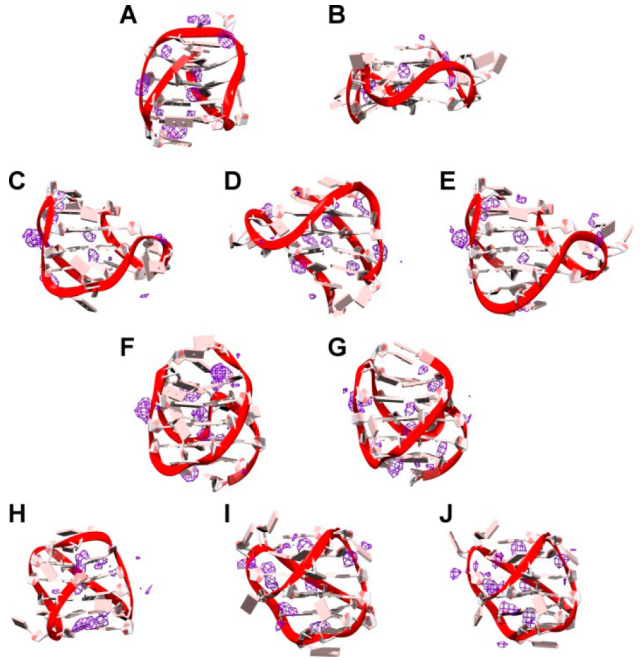


Figure 9. Pseudodensity grid maps of cations for 143D (A), 1KF1 (B), 2GKU (C), 2HY9 (D), 2JSM (E), 2JPZ (F), 2JSL (G), 2KF8 (H), 2KKA-G (I), and 2KKA-I (J). Cation density (purple) was contoured at 3× the density of sodium (6 M) for 143D and at 3× the density of potassium (4 M) for the other G-quadruplex structures. The average structure of each G-quadruplex over the trajectory is shown.

direct electrostatic interaction with the DNA, the Mg^{2+} ion is thought to assume a secondary role of shielding the anionic charge of the phosphate groups. In fact, it is well-known that polyanions, like DNA, can attract a shell of cations to help partially neutralize the negatively charged backbone phosphates.[84,85] Together, these findings suggest that, in addition to stabilizing the G-tetrads within the G-quadruplex core, cations can also play other roles in promoting G-quadruplex formation and stability.

Effects of Cation Binding on Hydrodynamic Calculations. HBM was used to provide information about the number of cations bound to the G-quadruplex structures. This use of HBM is based on the premise that the hydrodynamic properties are dictated by the size (volume and mass) and shape of a macromolecule in solution. When accurate experimental information is known about the volume and shape of the G-quadruplex structure, HBM can be used to estimate the mass of the G-quadruplex and, by extension, the number of cations bound to it. To illustrate this, a series of HYDROPRO calculations were performed. The additions of potassium or sodium to the G-quadruplex structure were accounted for by changing the molecular weight value in the HYDROPRO parameter file. For example, for a G-quadruplex structure with no potassium bound the molecular weight value used was the molecular weight of the DNA only. Whereas, for a G-quadruplex structure with two potassium ions bound, the molecular weight value used was the sum of the molecular weight of the DNA and two times the molecular weight of potassium. The calculations were done using all three modes of HYDROPRO calculation with the default AER instead of the calibrated AER. In order to display the data for the sequences on the same scale, the number of ions bound was normalized using eq 5:

$$\%\ \text{charge saturation} = \frac{\text{no. of potassiums bound}}{\text{sequence length(no. of bases)} - 1} \quad (5)$$

The results of the HYDROPRO calculations are shown in Figure 10. For the atomic-level hydrodynamic bead models, the lowest errors were observed at about 50% charge neutralization (i.e., about 10−13 cations), while, for the coarser residue-level models, the lowest errors were observed at about 60% charge saturation (i.e., about 12−15 cations). The findings agreed well with previous studies which reported 10−11 potassium ions bound to the hTel22 sequence in 30 mM KCl buffer.[86] This demonstrates that HBM may have application beyond structural prediction.

CONCLUSION

HBM was demonstrated to be a powerful tool for studying G-quadruplex structures when atomic-level structural representations are unavailable, ambiguous, or cannot be determined experimentally. In such cases, the use of low-resolution techniques such as hydrodynamics, combined with readily accessible biophysical measurements (e.g., CD spectroscopy, fluorescent spectroscopy), can be used to obtain general information regarding the G-quadruplex structure, size, and shape. HBM can bridge low-resolution hydrodynamic measurements and high-resolution molecular modeling to provide

dx.doi.org/10.1021/jp502213y | J. Phys. Chem. B 2014, 118, 5390−5405

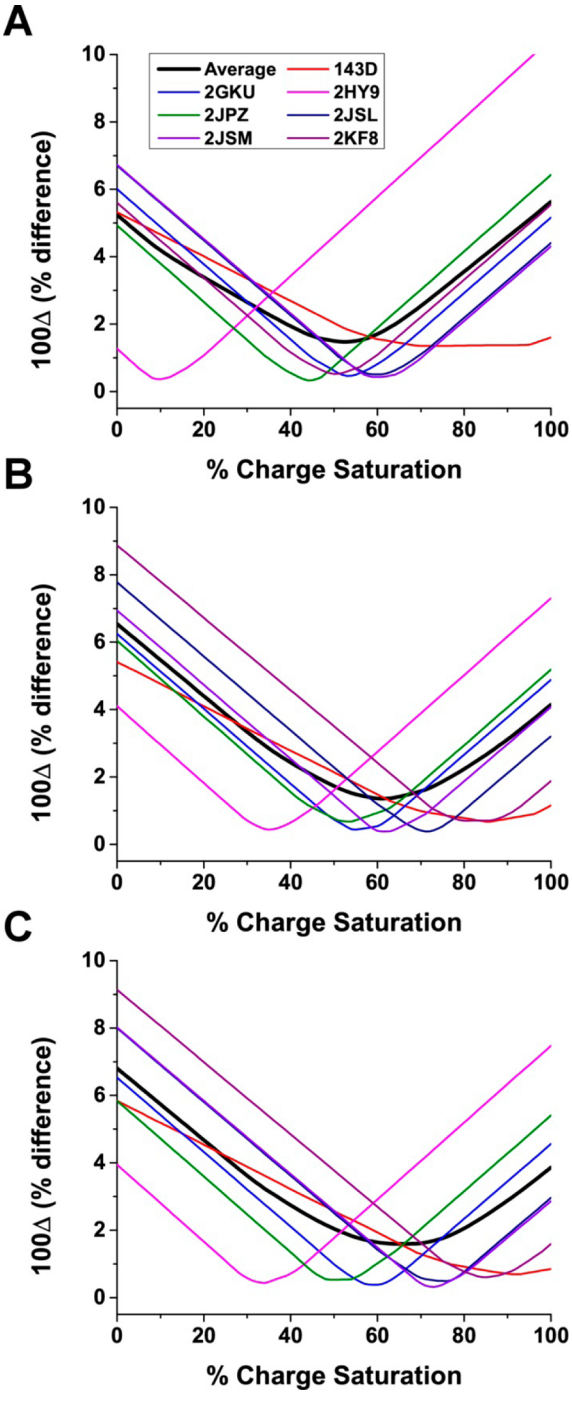


Figure 10. $s_{20,W}$ values as a function of the number of bound cations. The values for 100Δ as a function of percent charge saturation for the primary hydrodynamic model calculated using the seven G-quadruplex structures formed from the human telomere sequence. Hydrodynamic properties of G-quadruplexes were calculated using atomic-level shell-model calculation (A), residue-level shell-model calculation (B), and residue-level bead-model calculation (C).

further information regarding these structures. For example, HBM can be used to estimate the number of cations bound to the G-quadruplex structure. When applied to the study of novel G-quadruplex-forming sequences, HBM can be used to rule out structures that are not representative of the ensemble as was the case with 1KF1 whose calculated values differed greatly from the experimental values. However, hydrodynamics remains a low-resolution technique and molecular models for which calculated hydrodynamic values agreed (or differed only slightly) with experimental values will need to be confirmed with additional hydrodynamic and biophysical measurements, as was the case with the 2KKA-G and 2KKA-I models.

Another limitation of hydrodynamics and HBM is that the calculation is typically performed on one structure giving a static look at an otherwise dynamic system. HYDROPRO can only calculate hydrodynamic values for a single structure at any given time. However, as demonstrated, HBM can be used in tandem with MD simulations to provide a more dynamic representation of the macromolecule. In addition, the high-resolution nature of molecular dynamics can help complement the low-resolution nature of hydrodynamic measurements. As was observed with the 2KKA-G and 2KKA-I models, hydrodynamics was not able to distinguish between the two sequences, as both gave rise to identical sedimentation and diffusion values. However, cluster analysis of the MD trajectories revealed that the 2KKA-G model was more heterogeneous than the 2KKA-I model and demonstrated the effect of inosine substitution on selecting for a particular G-quadruplex topology. The findings demonstrated that molecular dynamics can supplement HBM in the study of G-quadruplex structures.

For G-quadruplexes, it is recommended to use either the atomic-level model with the shell-model calculation mode (AER = 2.19 Å) or the residue-level model with the bead-model calculation mode (AER = 5.04 Å) for HYDROPRO calculations, as both modes can predict the hydrodynamic properties accurately with a reasonable estimate of the size of the macromolecule. The atomic-level model with shell-model calculation is considered the gold standard and should be used whenever the most rigorous calculation is required. The residue-level model with the bead-model calculation mode has the advantage of being significantly faster (by several orders of magnitude) but with a slightly higher error in predicting hydrodynamic properties. For molecular weight, the recommended value is the molecular weight of the DNA and internal coordinating cations. For Na^+, because these ions are smaller and can fit within the G-quartet plane, the number of internal coordinating ions is equal to the number of G-quartets. For K^+, which is bigger and thus can only fit in between the planes of the G-quartets, the number of internal coordinating ions is one less than the number of G-quartets. The parameters for HYDROPRO presented can be used for hydrodynamic calculation of G-quadruplexes or can be further optimized by fitting with additional hydrodynamic values (i.e., rotational diffusion coefficients, NMR relaxation time, intrinsic viscosity, etc.). Although this work was conducted on the human telomere sequence, the experimental approach outlined can be easily adapted for other G-quadruplex-forming sequences to propose relevant G-quadruplex structures that can be used as a basis for drug design.

■ ASSOCIATED CONTENT

Supporting Information

Figures showing the following: (Figures S1 and S2) atomic fluctuation calculations per atom and per residue over the MD trajectories, (Figure S3) a_T calibration plot for HYDROPRO, (Figures S4−S13) representative structures from cluster analysis, (Figures S14 and S15) porcupine plots of second and third eigenvectors for MD trajectories, (Figure S16) plots of the first 10 eigenvalues for each MD trajectory, (Figure S17) plots of percent of variance explained by the first 10 eigenvectors,

dx.doi.org/10.1021/jp502213y | J. Phys. Chem. B 2014, 118, 5390−5405

(Figures S18−S27) plots of the first eigenvectors vs the second eigenvectors for each G-quadruplex structure, stems, loop, and flanking bases, (Figure S28) pseudodensity grid maps of water at 2× bulk water density, and (Figure S29) pseudodensity grid maps of cations at reference density. Also, Table S1 compares hydrodynamic values computed from MD simulations versus NMR structures. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: john.trent@louisville.edu. Fax: +1 (502) 852-7979. Phone: +1 (502) 852-2194.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Computing resources provided by the Cardinal Research Cluster at the University of Louisville. Molecular graphics images were produced using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR001081). Sedimentation velocity data were plotted using GUSSI from the University of Texas Southwestern Medical Center. This work was supported by NIH Grants CA35635 (J.B.C.), GM077422 (J.B.C. and J.O.T.) and University of Louisville grant, CTSPGP 20058 Award (J.B.C. and J.O.T).

ABBREVIATIONS

AUC, analytical ultracentrifugation; CD, circular dichroism; EDTA, ethylenediaminetetraacetic acid; MD, molecular dynamics; MMPBSA, molecular mechanic Poisson−Boltzmann surface area; NMR, nuclear magnetic resonance; RDFs, radial distribution functions; RMSD, root-mean-square deviation; SV, sedimentation velocity

REFERENCES

(1) 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.

(2) Huppert, J. L.; Balasubramanian, S. Prevalence of Quadruplexes in the Human Genome. *Nucleic Acids Res.* **2005**, *33*, 2908−2916.

(3) Huppert, J. L.; Balasubramanian, S. G-quadruplexes in Promoters Throughout the Human Genome. *Nucleic Acids Res.* **2007**, *35*, 406−413.

(4) Verma, A.; Halder, K.; Halder, R.; Yadav, V. K.; Rawal, P.; Thakur, R. K.; Mohd, F.; Sharma, A.; Chowdhury, S. Genome-Wide Computational and Expression Analyses Reveal G-Quadruplex DNA Motifs as Conserved cis-Regulatory Elements in Human and Related Species. *J. Med. Chem.* **2008**, *51*, 5641−5649.

(5) Wright, W. E.; Tesmer, V. M.; Huffman, K. E.; Levene, S. D.; Shay, J. W. Normal Human Chromosomes Have Long G-rich Telomeric Overhangs at One End. *Genes Dev.* **1997**, *11*, 2801−2809.

(6) Riou, J. F.; Guittat, L.; Mailliet, P.; Laoui, A.; Renou, E.; Petitgenet, O.; Mégnin-Chanet, F.; Hélène, C.; Mergny, J. L. Cell Senescence and Telomere Shortening Induced by a New Series of Specific G-quadruplex DNA Ligands. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2672−2677.

(7) Cuesta, J.; Read, M. A.; Neidle, S. The Design of G-quadruplex Ligands as Telomerase Inhibitors. *Mini-Rev. Med. Chem.* **2003**, *3*, 11−21.

(8) De Cian, A.; Lacroix, L.; Douarre, C.; Temime-Smaali, N.; Trentesaux, C.; Riou, J. F.; Mergny, J. L. Targeting Telomeres and Telomerase. *Biochimie* **2008**, *90*, 131−155.

(9) Lopes, J.; Piazza, A.; Bermejo, R.; Kriegsman, B.; Colosio, A.; Teulade-Fichou, M.-P.; Foiani, M.; Nicolas, A. G-Quadruplex-Induced Instability During Leading-Strand Replication. *EMBO J.* **2011**, *30*, 4033−4046.

(10) Rodriguez, R.; Miller, K. M.; Forment, J. V.; Bradshaw, C. R.; Nikan, M.; Britton, S.; Oelschlaegel, T.; Xhemalce, B.; Balasubramanian, S.; Jackson, S. P. Small-Molecule−Induced DNA Damage Identifies Alternative DNA Structures in Human Genes. *Nat. Chem. Biol.* **2012**, *8*, 301−310.

(11) Shay, J. W.; Bacchetti, S. A Survey of Telomerase Activity in Human Cancer. *Eur. J. Cancer* **1997**, *33*, 787−791.

(12) Balasubramanian, S.; Neidle, S. G-quadruplex Nucleic Acids as Therapeutic Targets. *Curr. Opin. Chem. Biol.* **2009**, *13*, 345−353.

(13) Han, H.; Hurley, L. H. G-Quadruplex DNA: A Potential Target for Anti-Cancer Drug Design. *Trends Pharmacol. Sci.* **2000**, *21*, 136−142.

(14) Neidle, S.; Read, M. A. G-Quadruplexes as Therapeutic Targets. *Biopolymers* **2000**, *56*, 195−208.

(15) Ou, T. M.; Lu, Y. J.; Tan, J. H.; Huang, Z. S.; Wong, K. Y.; Gu, L. Q. G-Quadruplexes: Targets in Anticancer Drug Design. *ChemMedChem* **2008**, *3*, 690−713.

(16) Saretzki, G. Telomerase Inhibition as Cancer Therapy. *Cancer Lett.* **2003**, *194*, 209−219.

(17) White, L. K.; Wright, W. E.; Shay, J. W. Telomerase Inhibitors. *Trends Biotechnol.* **2001**, *19*, 114−120.

(18) Drygin, D.; Siddiqui-Jain, A.; O'Brien, S.; Schwaebe, M.; Lin, A.; Bliesath, J.; Ho, C. B.; Proffitt, C.; Trent, K.; Whitten, J. P.; et al. Anticancer Activity of CX-3543: A Direct Inhibitor of rRNA Biogenesis. *Cancer Res.* **2009**, *69*, 7653−7661.

(19) Wang, Y.; Patel, D. J. Solution Structure of the Human Telomeric Repeat d[AG3(T2AG3)3] G-Tetraplex. *Structure* **1993**, *1*, 263−282.

(20) Parkinson, G. N.; Lee, M. P. H.; Neidle, S. Crystal Structure of Parallel Quadruplexes from Human Telomeric DNA. *Nature* **2002**, *417*, 876−880.

(21) 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.

(22) Dai, J.; Punchihewa, C.; Ambrus, A.; Chen, D.; Jones, R. A.; Yang, D. Structure of the Intramolecular Human Telomeric G-Quadruplex in Potassium Solution: A Novel Adenine Triple Formation. *Nucleic Acids Res.* **2007**, *35*, 2440−2450.

(23) Phan, A. T.; Kuryavyi, V.; Luu, K. N.; Patel, D. J. Structure of Two Intramolecular G-Quadruplexes Formed by Natural Human Telomere Sequences in K+ Solution. *Nucleic Acids Res.* **2007**, *35*, 6517−6525.

(24) 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.

(25) Lim, K. W.; Amrane, S.; Bouaziz, S.; Xu, W.; Mu, Y.; Patel, D. J.; Luu, K. N.; Phan, A. T. n. 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.

(26) Zhang, Z.; Dai, J.; Veliath, E.; Jones, R. A.; Yang, D. Structure of a Two-G-Tetrad Intramolecular G-Quadruplex Formed by a Variant Human Telomeric Sequence in K+ Solution: Insights into the Interconversion of Human Telomeric G-Quadruplex Structures. *Nucleic Acids Res.* **2010**, *38*, 1009−1021.

(27) Dai, J.; Carver, M.; Yang, D. Polymorphism of Human Telomeric Quadruplex Structures. *Biochimie* **2008**, *90*, 1172−1183.

(28) Yang, D.; Okamoto, K. Structural Insights into G-Quadruplexes: Towards New Anticancer Drugs. *Future Med. Chem.* **2010**, *2*, 619−646.

(29) Sagi, J. G-Quadruplexes Incorporating Modified Constituents: A Review. *J. Biomol. Struct. Dyn.* **2013**, *32*, 477−511.

(30) Sannohe, Y.; Sugiyama, H. Overview of Formation of G-Quadruplex Structures. *Current Protocols in Nucleic Acid Chemistry*; John Wiley & Sons, Inc.: Hoboken, NJ, 2001; Unit 17.2.1-17.

(31) Xue, Y.; Kan, Z.-y.; Wang, Q.; Yao, Y.; Liu, J.; Hao, Y.-h.; Tan, Z. Human Telomeric DNA Forms Parallel-Stranded Intramolecular G-Quadruplex in K+ Solution under Molecular Crowding Condition. *J. Am. Chem. Soc.* **2007**, *129*, 11185−11191.
(32) Miller, M. C.; Buscaglia, R.; Chaires, J. B.; Lane, A. N.; Trent, J. O. Hydration Is a Major Determinant of the G-Quadruplex Stability and Conformation of the Human Telomere 3′ Sequence of d(AG3-(TTAG3)3). *J. Am. Chem. Soc.* **2010**, *132*, 17105−17107.
(33) Buscaglia, R.; Miller, M. C.; Dean, W. L.; Gray, R. D.; Lane, A. N.; Trent, J. O.; Chaires, J. B. Polyethylene Glycol Binding Alters Human Telomere G-Quadruplex Structure by Conformational Selection. *Nucleic Acids Res.* **2013**, *41*, 7934−7946.
(34) Blume, S. W.; Guarcello, V.; Zacharias, W.; Miller, D. M. Divalent Transition Metal Cations Counteract Potassium-Induced Quadruplex Assembly of Oligo(dG) Sequences. *Nucleic Acids Res.* **1997**, *25*, 617−625.
(35) Miyoshi, D.; Nakao, A.; Sugimoto, N. Structural Transition of d(G4T4G4) from Antiparallel to Parallel G-Quartet Induced by Divalent Cations. *Nucleic Acids Res. Suppl.* **2001**, *1*, 259−260.
(36) Gray, R. D.; Li, J.; Chaires, J. B. Energetics and Kinetics of a Conformational Switch in G-Quadruplex DNA. *J. Phys. Chem. B* **2009**, *113*, 2676−2683.
(37) Gray, R. D.; Petraccone, L.; Trent, J. O.; Chaires, J. B. Characterization of a K+-Induced Conformational Switch in a Human Telomeric DNA Oligonucleotide Using 2-Aminopurine Fluorescence. *Biochemistry* **2009**, *49*, 179−194.
(38) Le, H. T.; Miller, M. C.; Buscaglia, R.; Dean, W. L.; Holt, P. A.; Chaires, J. B.; Trent, J. O. Not All G-quadruplexes are Created Equally: An Investigation of the Structural Polymorphism of the c-Myc G-Quadruplex-Forming Sequence and its Interaction with the Porphyrin TMPyP4. *Org. Biomol. Chem.* **2012**, *10*, 9393−9404.
(39) Li, J.; Correia, J. J.; Wang, L.; Trent, J. O.; Chaires, J. B. Not So Crystal Clear: The Structure of the Human Telomere G-Quadruplex in Solution Differs from that Present in a Crystal. *Nucleic Acids Res.* **2005**, *33*, 4649−4659.
(40) Hänsel, R.; Löhr, F.; Foldynová-Trantírková, S.; Bamberg, E.; Trantírek, L.; Dötsch, V. The Parallel G-Quadruplex Structure of Vertebrate Telomeric Repeat Sequences is not the Preferred Folding Topology under Physiological Conditions. *Nucleic Acids Res.* **2011**, *39*, 5768−5775.
(41) Heddi, B.; Phan, A. T. n. Structure of Human Telomeric DNA in Crowded Solution. *J. Am. Chem. Soc.* **2011**, *133*, 9824−9833.
(42) 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−%515.
(43) Dailey, M. M.; Miller, M. C.; Bates, P. J.; Lane, A. N.; Trent, J. O. Resolution and Characterization of the Structural Polymorphism of a Single Quadruplex-Forming Sequence. *Nucleic Acids Res.* **2010**, *38*, 4877−4888.
(44) Byron, O. Hydrodynamic Modeling: The Solution Conformation of Macromolecules and Their Complexes. In *Methods in Cell Biology*; Correia, J., Detrich, H. W., III, Eds.; Academic Press: Waltham, MA, 2008; Vol. *84*, pp 327−373.
(45) Niermann, M.; Bolten, M.; Eimer, W. Optimization of the Hydrodynamic Bead Model for the Analysis of DNA Conformations in Solution. *J. Phys. Chem. B* **1999**, *103*, 10065−10074.
(46) Petraccone, L.; Garbett, N. C.; Chaires, J. B.; Trent, J. O. An Integrated Molecular Dynamics (MD) and Experimental Study of Higher Order Human Telomeric Quadruplexes. *Biopolymers* **2010**, *93*, 533−548.
(47) Petraccone, L.; Spink, C.; Trent, J. O.; Garbett, N. C.; Mekmaysy, C. S.; Giancola, C.; Chaires, J. B. Structure and Stability of Higher-Order Human Telomeric Quadruplexes. *J. Am. Chem. Soc.* **2011**, *133*, 20951−20961.
(48) Petraccone, L.; Trent, J. O.; Chaires, J. B. The Tail of the Telomere. *J. Am. Chem. Soc.* **2008**, *130*, 16530−16532.
(49) Islam, B.; Sgobba, M.; Laughton, C.; Orozco, M.; Sponer, J.; Neidle, S.; Haider, S. Conformational Dynamics of the Human Propeller Telomeric DNA Quadruplex on a Microsecond Time Scale. *Nucleic Acids Res.* **2013**, *41*, 2723−2735.
(50) Ortega, A.; Amorós, D.; García de la Torre, J. Prediction of Hydrodynamic and Other Solution Properties of Rigid Proteins from Atomic- and Residue-Level Models. *Biophys. J.* **2011**, *101*, 892−898.
(51) Hayward, S.; Groot, B. Normal Modes and Essential Dynamics. In *Molecular Modeling of Proteins*; Kukol, A., Ed.; Humana Press: New York, NY, 2008; Vol. *443*, pp 89−106.
(52) Haider, S.; Parkinson, G. N.; Neidle, S. Molecular Dynamics and Principal Components Analysis of Human Telomeric Quadruplex Multimers. *Biophys. J.* **2008**, *95*, 296−311.
(53) Chaires, J. B. Human Telomeric G-Quadruplex: Thermodynamic and Kinetic Studies of Telomeric Quadruplex Stability. *FEBS J.* **2010**, *277*, 1098−1106.
(54) Lane, A. N. The Stability of Intramolecular DNA G-Quadruplexes Compared with Other Macromolecules. *Biochimie* **2012**, *94*, 277−286.
(55) Haider, S.; Neidle, S. Molecular Modeling and Simulation of G-Quadruplexes and Quadruplex-Ligand Complexes. *Methods Mol. Biol.* **2010**, *608*, 17−37.
(56) Yphantis, D. A.; Roark, D. E. Equilibrium Centrifugation of Nonideal Systems. Donnan Effect in Self-Associating Systems. *Biochemistry* **1971**, *10*, 3241−3249.
(57) Smith, C. A. Estimation of Sedimentation Coefficients and Frictional Ratios of Globular Proteins. *Biochem. Educ.* **1988**, *16*, 104−106.
(58) Harding, S. E. Frictional Coefficient, Ratio. *Encyclopedia of Molecular Biology*; John Wiley & Sons, Inc.: Hoboken, NJ, 2002.
(59) Fernandes, M. X.; Ortega, A.; López Martínez, M. C.; García de la Torre, J. Calculation of Hydrodynamic Properties of Small Nucleic Acids from Their Atomic Structure. *Nucleic Acids Res.* **2002**, *30*, 1782−1788.
(60) Bloomfield, V.; Dalton, W. O.; Van Holde, K. E. Frictional Coefficients of Multisubunit Structures. I. Theory. *Biopolymers* **1967**, *5*, 135−148.
(61) García de la Torre, J. Hydration from Hydrodynamics. General Considerations and Applications of Bead Modelling to Globular Proteins. *Biophys. Chem.* **2001**, *93*, 159−170.
(62) Schuck, P.; Perugini, M. A.; Gonzales, N. R.; Howlett, G. J.; Schubert, D. Size-Distribution Analysis of Proteins by Analytical Ultracentrifugation: Strategies and Application to Model Systems. *Biophys. J.* **2002**, *82*, 1096−1111.
(63) Dam, J.; Schuck, P. Calculating Sedimentation Coefficient Distributions by Direct Modeling of Sedimentation Velocity Concentration Profiles. In *Methods in Enzymology*; Michael, L. J., Ludwig, B., Eds.; Academic Press: Waltham, MA, 2004; Vol. *384*, pp 185−212.
(64) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; et al. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **2000**, *33*, 889−897.
(65) Štefl, R.; Cheatham, T. E.; Špačková, N. a.; 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.
(66) Fadrná, E.; Špačková, N. a.; 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.
(67) Cang, X.; Šponer, J.; Cheatham, T. E. Insight into G-DNA Structural Polymorphism and Folding from Sequence and Loop Connectivity through Free Energy Analysis. *J. Am. Chem. Soc.* **2011**, *133*, 14270−14279.
(68) Olsen, C. M.; Gmeiner, W. H.; Marky, L. A. Unfolding of G-Quadruplexes: Energetic, and Ion and Water Contributions of G-Quartet Stacking. *J. Phys. Chem. B* **2006**, *110*, 6962−6969.
(69) Miyoshi, D.; Karimata, H.; Sugimoto, N. Hydration Regulates the Thermodynamic Stability of DNA Structures under Molecular Crowding Conditions. *Nucleosides, Nucleotides Nucleic Acids* **2007**, *26*, 589−595.

(70) Poornima, C. S.; Dean, P. M. Hydration in Drug Design. 1. Multiple Hydrogen-Bonding Features of Water Molecules in Mediating Protein-Ligand Interactions. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 500−512.
(71) Poornima, C. S.; Dean, P. M. Hydration in Drug Design. 2. Influence of Local Site Surface Shape on Water Binding. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 513−520.
(72) Poornima, C. S.; Dean, P. M. Hydration in Drug Design. 3. Conserved Water Molecules at the Ligand-Binding Sites of Homologous Proteins. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 521−531.
(73) García-Sosa, A. T. Hydration Properties of Ligands and Drugs in Protein Binding Sites: Tightly-Bound, Bridging Water Molecules and Their Effects and Consequences on Molecular Design Strategies. *J. Chem. Inf. Model.* **2013**, *53*, 1388−1405.
(74) Chaires, J. B. A Thermodynamic Signature for Drug-DNA Binding Mode. *Arch. Biochem. Biophys.* **2006**, *453*, 26−31.
(75) Teeter, M. M.; Case, D. A. Harmonic and Quasiharmonic Descriptions of Crambin. *J. Phys. Chem.* **1990**, *94*, 8091−8097.
(76) Tai, K.; Shen, T.; Börjesson, U.; Philippopoulos, M.; McCammon, J. A. Analysis of a 10-ns Molecular Dynamics Simulation of Mouse Acetylcholinesterase. *Biophys. J.* **2001**, *81*, 715−724.
(77) Tai, K.; Shen, T.; Henchman, R. H.; Bourne, Y.; Marchot, P.; McCammon, J. A. Mechanism of Acetylcholinesterase Inhibition by Fasciculin: A 5-ns Molecular Dynamics Simulation. *J. Am. Chem. Soc.* **2002**, *124*, 6153−6161.
(78) Cheatham, T. E.; Kollman, P. A. Molecular Dynamics Simulations Highlight the Structural Differences among DNA:DNA, RNA:RNA, and DNA:RNA Hybrid Duplexes. *J. Am. Chem. Soc.* **1997**, *119*, 4805−4825.
(79) Joung, I. S.; Cheatham, T. E. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **2008**, *112*, 9020−9041.
(80) Reshetnikov, R. V.; Sponer, J.; Rassokhina, O. I.; Kopylov, A. M.; Tsvetkov, P. O.; Makarov, A. A.; Golovin, A. V. Cation Binding to 15-TBA Quadruplex DNA is a Multiple-Pathway Cation-Dependent Process. *Nucleic Acids Res.* **2011**, *39*, 9789−9802.
(81) Akhshi, P.; Acton, G.; Wu, G. Molecular Dynamics Simulations to Provide New Insights into the Asymmetrical Ammonium Ion Movement Inside of the [d(G3T4G4)]2 G-Quadruplex DNA Structure. *J. Phys. Chem. B* **2012**, *116*, 9363−9370.
(82) Pagano, B.; Mattia, C. A.; Cavallo, L.; Uesugi, S.; Giancola, C.; Fraternali, F. Stability and Cations Coordination of DNA and RNA 14-Mer G-Quadruplexes: A Multiscale Computational Approach. *J. Phys. Chem. B* **2008**, *112*, 12115−12123.
(83) Wei, D.; Parkinson, G. N.; Reszka, A. P.; Neidle, S. Crystal Structure of a c-Kit Promoter Quadruplex Reveals the Structural Role of Metal Ions and Water Molecules in Maintaining Loop Conformation. *Nucleic Acids Res.* **2012**, *40*, 4691−4700.
(84) Manning, G. S. The Molecular Theory of Polyelectrolyte Solutions with Applications to the Electrostatic Properties of Polynucleotides. *Q. Rev. Biophys.* **1978**, *11*, 179−246.
(85) Record, M. T.; Anderson, C. F.; Lohman, T. M. Thermodynamic Analysis of Ion Effects on the Binding and Conformational Equilibria of Proteins and Nucleic Acids: The Roles of Ion Association or Release, Ccreening, and Ion Effects on Water Activity. *Q. Rev. Biophys.* **1978**, *11*, 103−178.
(86) Gray, R. D.; Chaires, J. B. Linkage of Cation Binding and Folding in Human Telomeric Quadruplex DNA. *Biophys. Chem.* **2011**, *159*, 205−209.