Supplementary Information: Predicting site-binding modes of ions and water to nucleic acids using molecular solvation theory

George M. Giambaşu, David A. Case and Darrin M. York *,†,‡,¶

†*Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey*

‡*Laboratory for Biomolecular Simulation Research, Rutgers, The State University of New Jersey*

¶*Center for Integrative Proteomics Research, Rutgers, The State University of New Jersey*

E-mail: Darrin.York@rutgers.edu

S.1 List of X-ray structures used in the current work

Table S1: Summary of the 18 single crystal X-ray structures of *Oxy-GQ* available in the PDB.

PDB ID (Space Group)	GQ units	Channel Cations	Resolution (Å)	R Free
1JPQ[1] (P 32 2 1)	1	5 K^+	1.60	0.276
1JRN[1] (P 21 21 21)	2	5 K^+	2.00	0.276
2GWE[2](P 21 21 21)	7	5 K^+	2.20	0.318
2GWQ[3](P 21 21 21)	4	5 K^+	2.00	0.299
2HBN[4] (P 21 21 21)	2	5 Tl^+	1.55	0.248
4R45[5] (P -1)	1	5 K^+	1.90	0.340
4R47[5] (P 1 21/n 1)	1	5 K^+	1.85	0.304
5HIX[6] (P 41)	1	4 K^+	2.48	0.288

S.2 Force field recalibration details

The LJ potential for two atoms situated at distance r and that were assigned atom types i, j respectively is:

$$U_{LJ}^{ij}(r) = \epsilon_{ij}\left[\left(\frac{R_{min,ij}}{r}\right)^{12} - 2\left(\frac{R_{min,ij}}{r}\right)^{6}\right] \tag{S1}$$

ϵ_{ij} and $R_{min,ij}$ are computed using Lorentz-Berthelot mixing rules of atom type specific parameters. A comparison between the initial and recalibrated parameters is shown in Table S2 and the details and criteria used in the parametrization are presented in Fig. S1.

Table S2: Lennard-Jones potential parameters, $R_{min,ij}$ in Å and ϵ_{ij} in kcal/mol for cation-oxygen pairs.

		original		re-calibrated	
cation	oxygen	$R_{min,ij}$	ϵ_{ij}	$R_{min,ij}$	ϵ_{ij}
Na^+	carbonyl	2.8957	0.41324870	2.90	0.17
	ether	2.8732	0.45930025		
K^+	carbonyl	3.2767	0.45617345	3.32	0.20
	ether	3.2542	0.50700844		

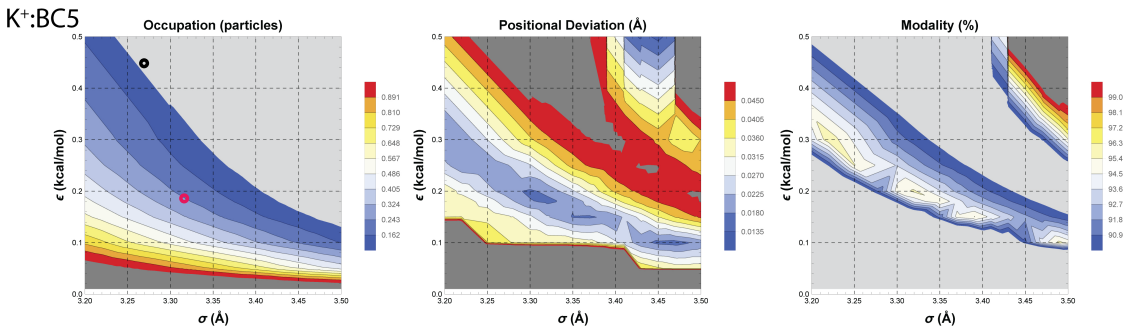

Figure S1: Parameters used in force field recalibration using complexes of Na^+ and K^+ with bis-crown-ethers. All the parameters were computed as integrals over the space identified as the binding locus in the bis-crown-ether cavity using the Laplacian criterion. To focus on the favorable values of the indexes, the plots were truncated with areas corresponding to values lower (light gray) or larger (dark gray) than those in the scale. Occupation represents the integrated cation density over the binding locus. Positional deviation represents the difference between the cation position resolved in the crystal and the average position of the calculated K^+ distribution. The modality index is computed as the ratio of the density at the average position of the binding mode and the maximum density within the binding mode and aims to maintain the constrain of a unimodal normal distribution across the binding locus, as resolved crystallographically. This is needed as for specific values of epsilon and sigma the distribution within the binding pocket acquires multiple peaks as the cation size decreases and the cation prefers to stick on the walls of the cavity. The original and refined parameter sets are marked with a black and red circle, respectively.

S.3 Solvent distributions obtained using the Non-Linear Poisson Boltzmann (NLPB) equation

The approach for solving the NLPB equation used here was previously described in.[7] Briefly, the NLPB equation has the form:

$$\begin{aligned} \nabla \cdot [\epsilon(\mathbf{r}) \nabla \phi(\mathbf{r})] &+ 4\pi \sum_{j=1}^{m} \rho_j^{bulk} q_j \exp\left[-q_j \phi(\mathbf{r}) - V_j(\mathbf{r})\right] / kT \\ &= 4\pi \rho\left(\mathbf{r}\right) \end{aligned} \tag{S2}$$

where $\epsilon\left(\mathbf{r}\right)$ is an inhomogeneous dielectric medium of relative permittivity, $\rho(\mathbf{r})$ is the charge distribution of the solute, ρ_j^{bulk} is the bulk number density of the j^{th} species of mobile ion in solution have charge q_j, $V_j\left(\mathbf{r}\right)$ is the steric interaction energy term between the mobile ion and the solute and $\phi\left(\mathbf{r}\right)$ is the resulting electrostatic potential. $\phi\left(\mathbf{r}\right)$ is numerically solved for on a 3D grid, which gives the following expression for the pair distribution function of ion species j:

$$\rho_j\left(\mathbf{r}\right) = \rho_j^{bulk} \exp\left[-q_j \phi(\mathbf{r}) - V_j(\mathbf{r})\right] / kT. \tag{S3}$$

NLPB calculations were performed with Adaptive Poisson Boltzmann Solver (APBS)[8] using a grid size of 1 Å, an internal dielectric set to 1.0 for the DNA and 78.54 for the solvent where the standard molecular surface definition was used with a solvent radius of 1.4 Å. Single Debye-Huckel boundary conditions were used. Temperature was set to 298.15K. Specific cation radii needed to define the steric interaction energy term, $V_{cation}\left(\mathbf{r}\right)$, are chosen to reproduce the solvation free energies at infinite dilution of the cation. Here, the input cationic radii were varied over an interval of 1.0 to 3.6 to explore their effect on cationic population along the GQ internal channel (see Figure S2). This range of values includes values corresponding of effective ionic radii of K^+, Na^+ and ionic radii augmented by a water solvation layer. The effective cationic size of Na^+ and K^+ varies between 1.13 - 1.53 Å and 1.51 - 1.78 Å respectively depending on the cations coordination environment.[9] Augmenting the radii with an increment of 1.4 Å (the approximate width of a water layer) expands the radii interval to ~3.2 Å. As the anion is highly unlikely to make direct contact with the surface of the DNA, its excluded volume term, $V_{anion}\left(\mathbf{r}\right)$ is chosen to be equal to that of the cation, which is the standard behavior in APBS.

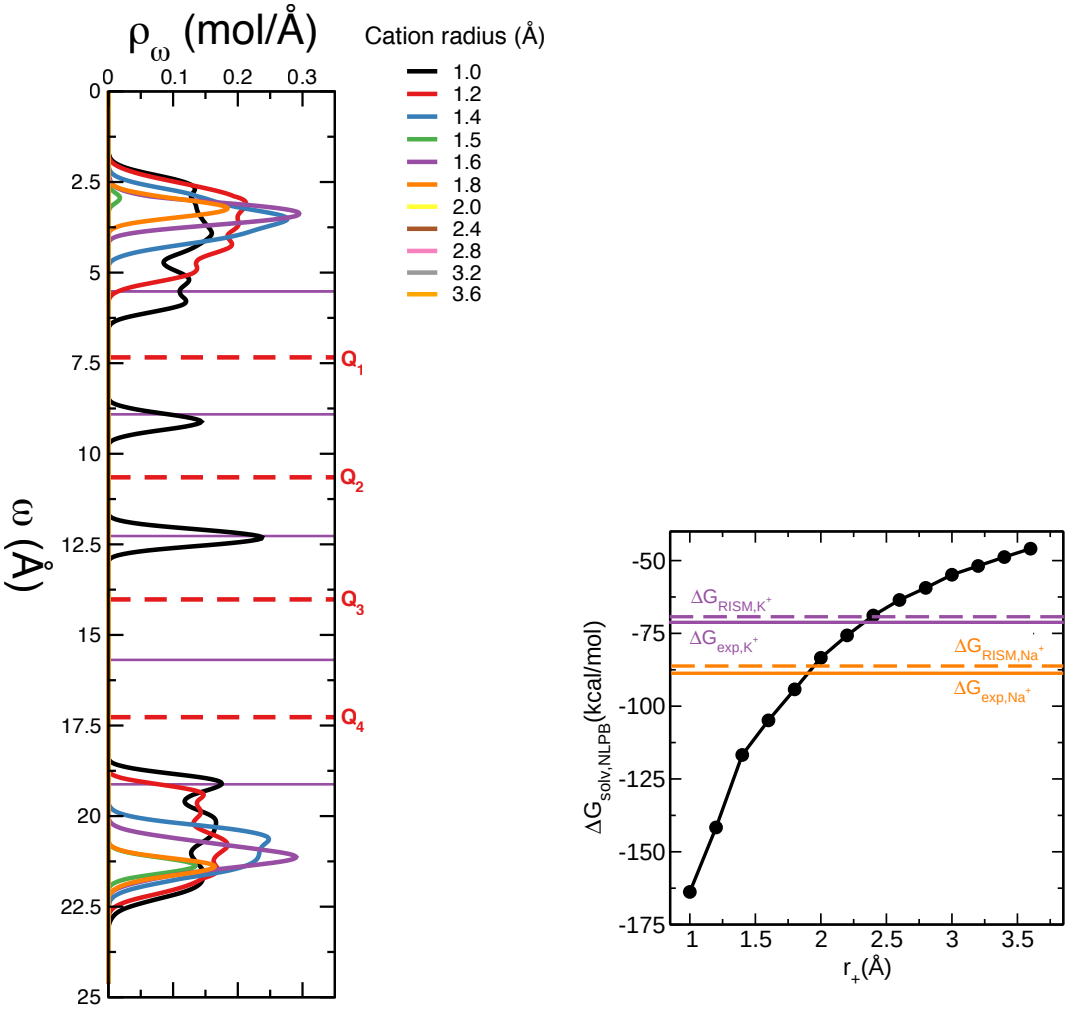

Figure S2: Details of NLPB calculation. (Left) Worm plots obtained from solving NLPB for *Oxy-GQ* in the presence of 100mM monovalent salt. The input cationic radii were varied over an interval of 1.0 to 3.6. This range of values includes values corresponding of effective ionic radii of K^+, Na^+ and ionic radii augmented by a water solvation layer. The effective cationic size of Na^+ and K^+ varies between 1.13 - 1.53 Å and 1.51 - 1.78 Å respectively depending on the cations coordination environment.[9] Augmenting the radii with an increment of 1.4 Å (the approximate width of a water layer) expands the radii interval to ~3.2 Å. The position of the G-quartets along the worm are marked with red dashed lines. (Right) Solvation free energies as function of cationic radii obtained from NLPB. For comparison, overlaid are solvation free energies at infinite dilution of Na^+ (orange) and K^+ (violet) obtained from experiment[10] (horizontal solid lines) and using RISM[11] (horizontal dashed lines). NLPB calculations were carried out using APBS[8] are are detailed in.[7]

S.4 Testing convergence of explicit solvent MD simulations

To test the ability of MD simulations to provide a converged image of the water and ion binding modes we run two sets of simulations of *Oxy-GQ* using explicit solvent and salt (NaCl and KCl, respectively) using a salt composition be close to 0.2M. Each set of simulations was comprised of 4 independent trajectories where ions were placed at random initial positions under the constraint of being at least 3 Angstroms away of any nucleic acid atom. The GQ, water and ions force-field were the same as the ones used in the 3D-RISM calculations discussed in the main text. The simulation box was chosen to have truncated octahedron symmetry, the size of the simulation system was chosen to assure a clearance of at least 35 Angstroms between any atom of the GQ and the margin of the box, having as such a reasonable space to represent the diffuse and territorial bound water and ion binding modes. The total number of water molecules in the system was 30,849, the total number of cations was 133, while the number of anions was 111, resulting in total number of 93,573 atoms. In the first MD equilibration step the DNA solute was restrained to its crystallographic structure while the water and salt was allowed to equilibrate at a temperature of 298K and pressure of 1bar for 20ns in the NPT ensemble using the Langevin thermostat and a the Monte Carlo barrostat. The second equilibration stage (50 ns) was run with the positional restraints on the DNA molecule removed in the NVT ensemble. Production of each trajectory was run for 300ns, resulting in a total of 1.2 micro-seconds of simulation for each set of simulations. In all simulations water and cations were able to permeate along the GQ central channel during the equilibration stage. During the production most the binding sites were occupied by either a water or a cation.

A general feature of all simulations was the long-lifetime of cation and water coordination patterns along the GQ central channel. Each simulation converged to a well defined and distinctive coordination pattern, with the particles bound in the 3 intercalated binding sites showing the longest lifetimes and fewest exchanges (if any) with neighboring sites. In the case of simulations run the presence of KCl, 3 of the 4 simulations converged to a coordination pattern where K^+ occupied two (out of three) intercalated binding sites with the other intercalated site being occupied by a water molecule. The 4th simulation in the presence of KCl shown a K^+ in the central intercalated binding mode and on the marginal binding sites; in this case the cation entered the quadruplex laterally, not through the channel. In the case of GQ simulations in the presence of NaCl, the majority of simulations adopted a coordination pattern whereby two cations occupy two of the intercalated binding sites, with one simulation having cation in the central intercalated site and two others in the marginal sites. As shown by the worm plots, water occupancy of the binding sites is inversely correlated with cation occupancy.(see Figure S3)

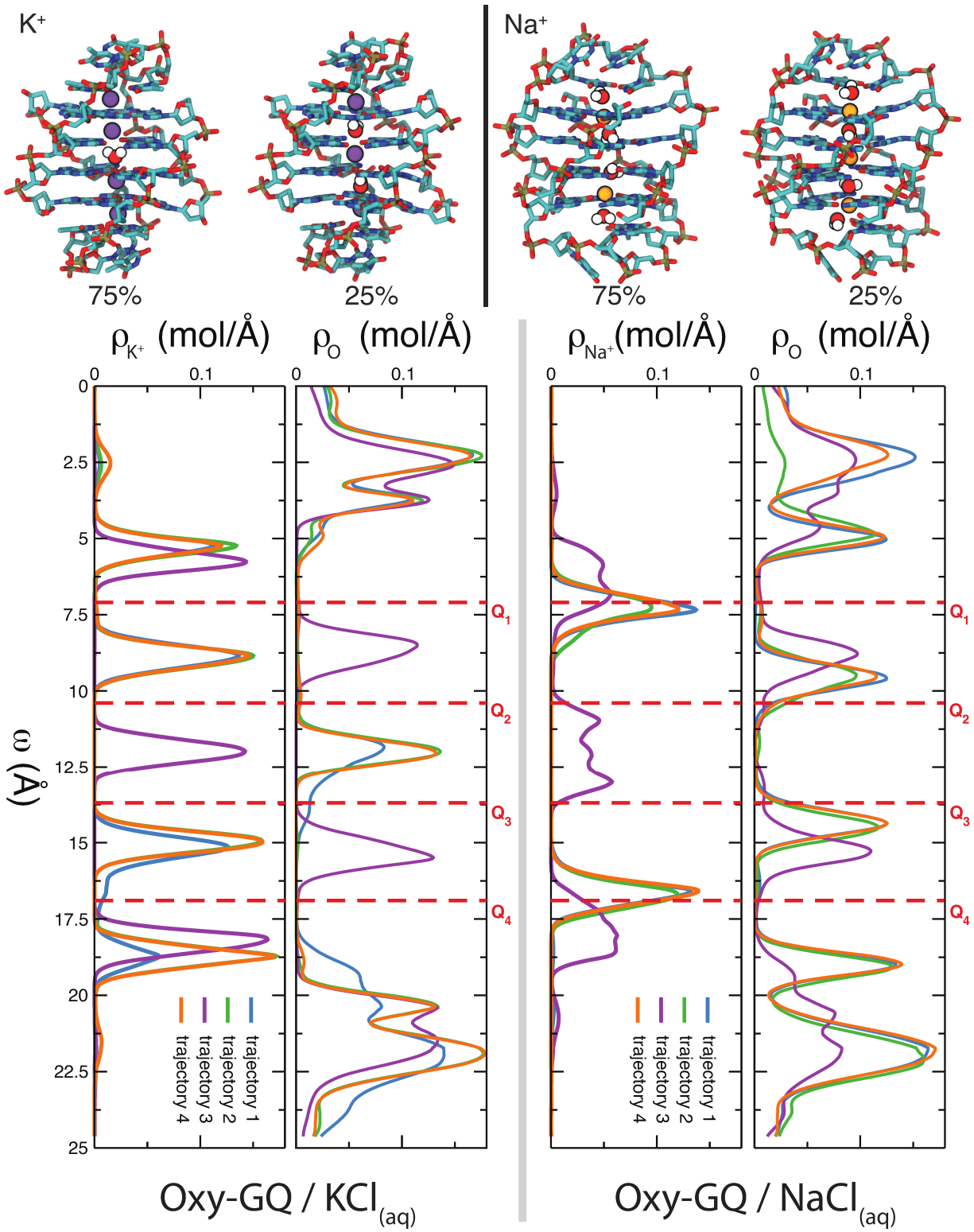

Figure S3: Representative equilibrated structures taken from MD simulation of *Oxy-GQ* in the presence of KCl (left), and NaCl(right). Particles bound to the channel are shown using spheres: K^+ in violet, Na^+ in orange, water oxygen and hydrogen using red and white, respectively.

S.5 Energy minimization using 3D-RISM and XMIN

Energy minimization was carried out using an L-BFGS preconditioned truncated Newton conjugate gradient algorithm as implemented in the XMIN approach available in AMBER which is a technique that can minimize molecular structures to lower energy and gradient than other strategies and and requires an order of magnitude fewer minimization cycles. An initial minimization stage was carried out at an intermediary salt concentration (0.1M NaCl and KCl, respectively) for 200 cycles (for X-ray refined structures) or 400 cycles (for structures from NMR ensembles). 50 more energy minimization cycles were carried out at each salt conditions, although the structural changes were minimal. Convergence criteria for 3D-RISM calculations were stricter than those intended for estimating thermodynamic quantities such as chemical potentials or preferential interaction parameters and used threshold residual of 10^{-11} rather than 10^{-6}.These requirement is necessary to assure an accurate estimation of forces and estimates of second order derivatives but comes with a significant computational cost. Nevertheless, the minimization protocols lead o a gradient norm is less than 10^{-1} kcal/mole-Å. The NMR ensemble (10 structures) used here was refined based on structural data measured in the presence of Tl^+ (an electron rich K^+ surrogate) presented in.[12]

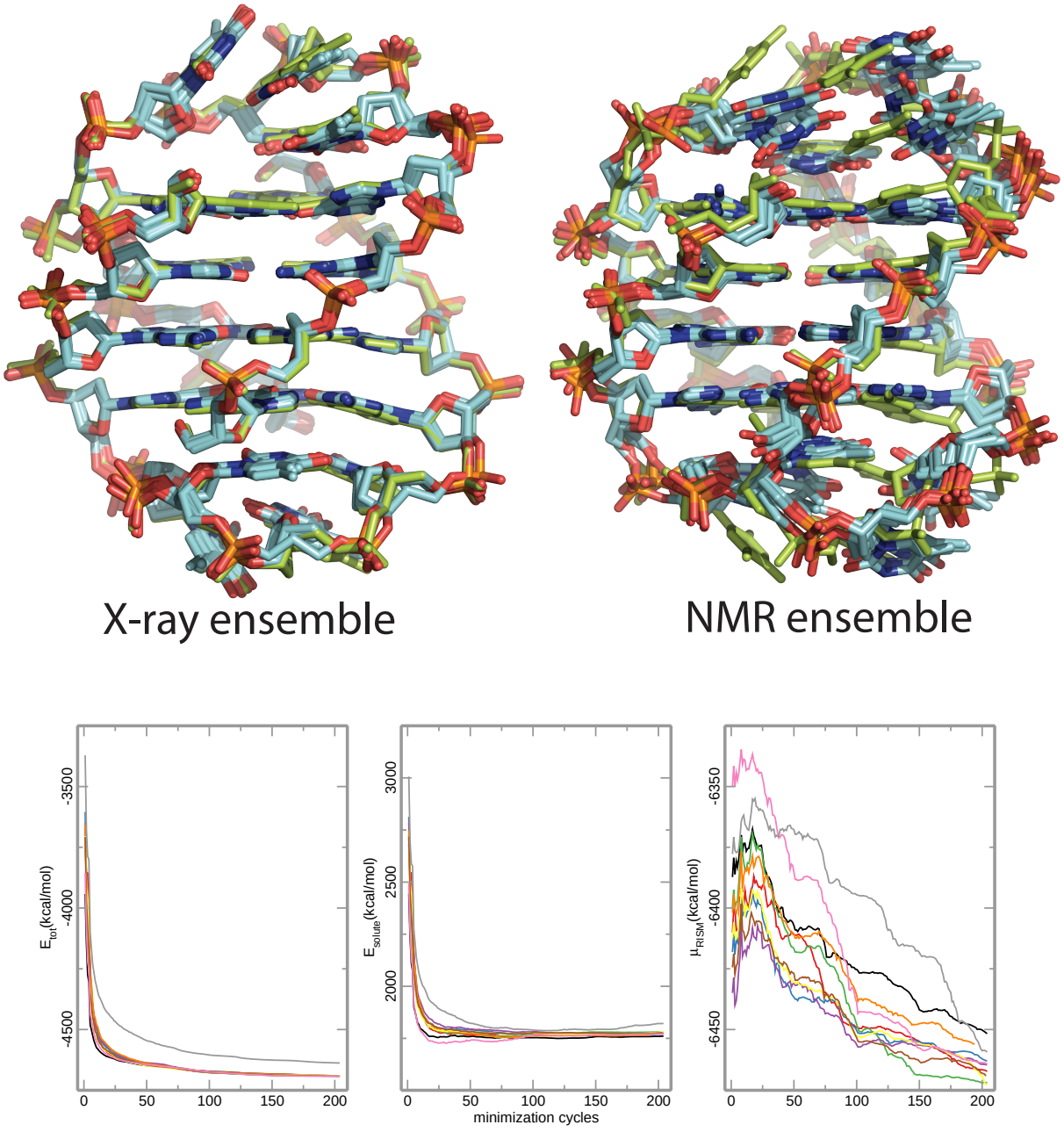

Figure S4: (Top) Sample 3D-RISM Free-Energy minimized crystallographic ensemble (top, left) and NMR ensemble (top, right) overlapped with one of the initial X-ray refined structures (shown in light green, PDB ID: 4r45). (Bottom) Summary of a typical energy minimization using 3D-RISM. (Bottom Left) Evolution of the total energy of the system and of its (bottom, middle) solute internal energy and (bottom, left) solvation chemical potential components, respectively.

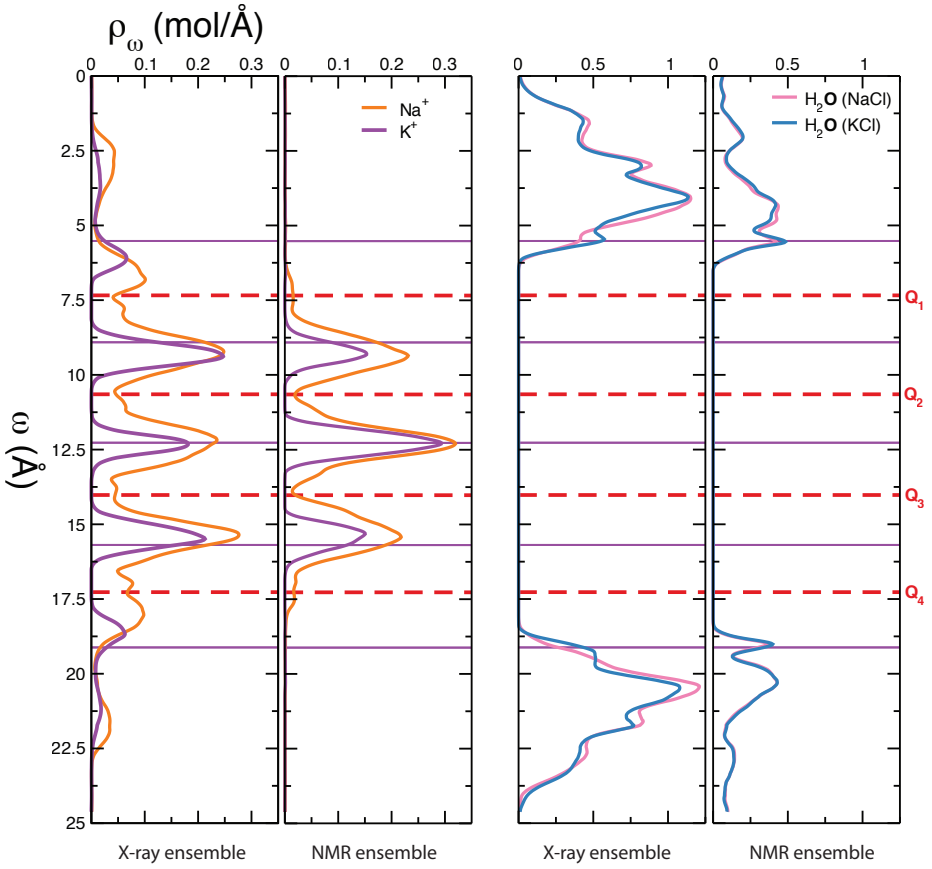

Figure S5: Worm plots for Na^+,K^+ (leftmost two panels) and water (rightmost two panels) obtained from minimized X-ray and NMR ensembles, respectively.

S.6 Ion counting profiles of helical double stranded 24bp DNA obatined using the original and refined force field

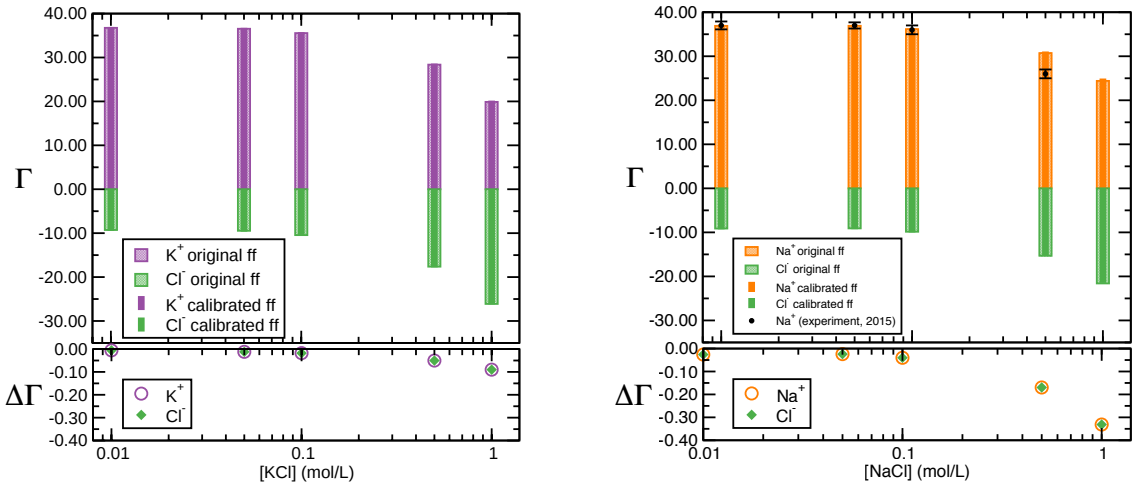

Figure S6: Comparison between ion counting profiles for a 24bp long DNA (24L[13]) computed using the original and calibrated force field. Ion counts (preferential interaction parameters) provide a measure of the extent to which counter-ion condensation and anion depletion contribute to electrostatic stabilization of charged polymers such as nucleic acids. Recent experimental measurements for Na^+ ion counting profiles in the presence of the Cl^- anion and 24L DNA are marked using black dots and error bars.[14]

References

(1) Haider, S.; Parkinson, G. N.; Neidle, S. Crystal Structure of the Potassium Form of an Oxytricha Nova G-Quadruplex. *J. Mol. Biol.* **2002**, *320*, 189–200.

(2) doi:10.2210/pdb2gwe/pdb.

(3) doi:10.2210/pdb2gwq/pdb.

(4) Gill, M. L.; Strobel, S. A.; Loria, J. P. Crystallization and Characterization of the Thallium Form of the Oxytricha Nova G-Quadruplex. *Nucleic Acids Res.* **2006**, *34*, 4506–4514.

(5) Mandal, P. K.; Collie, G. W.; Kauffmann, B.; Huc, I. Racemic DNA Crystallography. *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 14424–14427.

(6) Mandal, P. K.; Baptiste, B.; Langlois d'Estaintot, B.; Kauffmann, B.; Huc, I. Multivalent Interactions between an Aromatic Helical Foldamer and a DNA G-Quadruplex in the Solid State. *Chembiochem* **2016**, *17*, 1911–1914.

(7) Giambaşu, G. M.; Luchko, T.; Herschlag, D.; York, D. M.; Case, D. A. Ion counting from explicit-solvent simulations and 3D-RISM. *Biophys. J.* **2014**, *106*, 883–894.

(8) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 10037–10041.

(9) Miessler, G. L.; Fischer, P. J.; Tarr, D. A. *Inorganic Chemistry*, 5th ed.; Pearson, 2014.

(10) Schmid, R.; Miah, A. M.; Sapunov, V. N. A new table of the thermodynamic quantities of ionic hydration: values and some applications (enthalpy-entropy compensation and Born radii). *Phys. Chem. Chem. Phys.* **2000**, *2*, 97–102.

(11) Joung, I.; Luchko, T.; Case, D. A. Simple electrolyte solutions: Comparison of DRISM and molecular dynamics results for alkali halide solutions. *J. Chem. Phys.* **2013**, *138*, 044103.

(12) Gill, M. L.; Strobel, S. A.; Loria, J. P. 205Tl NMR Methods for the Characterization of Monovalent Cation Binding to Nucleic Acids. *J. Am. Chem. Soc.* **2005**, *127*, 16723–16732.

(13) Bai, Y.; Greenfeld, M.; Travers, K. J.; Chu, V. B.; Lipfert, J.; Doniach, S.; Herschlag, D. Quantitative and comprehensive decomposition of the ion atmosphere around nucleic acids. *J. Am. Chem. Soc.* **2007**, *129*, 14981–14988.

(14) Gebala, M.; Giambaşu, G. M.; Lipfert, J.; Bisaria, N.; Bonilla, S.; Li, G.; York, D. M.; Herschlag, D. Cation-Anion Interactions within the Nucleic Acid Ion Atmosphere Revealed by Ion Counting. *J. Am. Chem. Soc.* **2015**, *137*, 14705–14715.