SUPPLEMENTARY MATERIAL

R. Štefl, T. E. Cheatham III, N. Špačková, E. Fadrná, I. Berger, J. Koča, and J. Šponer
"Formation pathways of a guanine-quadruplex DNA revealed by molecular dynamics and thermodynamic analysis of the substates." *Biophysical Journal* (2003).

Generation, equilibration and simulation of the model structures.

A variety of G-DNA quadruplex models were built and simulated ranging from standard G-DNA quadruplexes to models with one or two strands shifted and various potential dimer and trimer intermediates. These are described in greater detail below than in the original manuscript. PDB structures of the average MD structures for the various models are available from the corresponding authors.

Native G-DNA and strand slipped structures: A variety of alternative models of the $d(GGGG)_4$ quadruplex stem structure were investigated in MD simulation (shown in the main text in Figure 3 and described briefly in Table 1). The initial models structures were based on the 0.95 Å resolution crystal structure of the $d(TGGGGT)_4$ quadruplex (Phillips et al., 1997) omitting the terminal thymines. The alternate models include the native $d(GGGG)_4$ quadruplex stem with three Na^+ ions bound into the central channel (G-DNA native) and the native parallel stranded $d(GGGG)_4$ quadruplex stem with no ions in the central channel (G-DNA vacant).

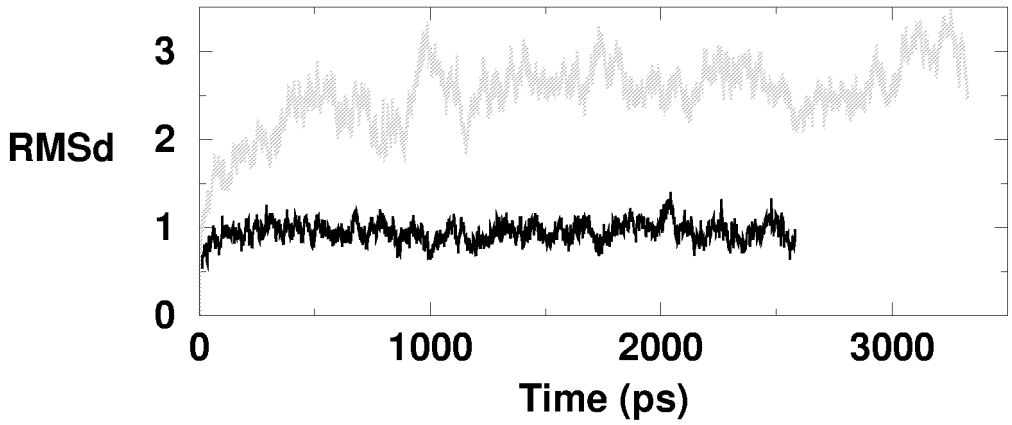

Figure S1: Plot of the RMSd (Å) vs. time for all DNA atoms of the native G-DNA quadruplex (black) and ion-vacant G-DNA quadruplex (gray) from the initial structures.

A variety of other models were investigated which had one or more of the strands slipped, or shifted up or down, by one base pair. The impetus for investigating these models came from earlier simulations where spontaneous strand slippage was observed during nanosecond-scale simulation of the G-DNA vacant model structure (Spackova et al., 1999). These single slipped strand structures represent the most probable geometry for strand slippage initiated from the fully paired four-quartet quadruplex stem (Figure 3b). Two different slipped-strand structures involving a shift of one of the strands are discussed in the current work. The first slipped strand model structure (slipped_D_2ions) is a quadruplex stem with one strand (D) shifted in the 5' direction out of register by one base pair step (~3.4 Å) with respect to the three other strands. This model structure was taken as a snapshot (1.9 nanoseconds) from the G-DNA vacant simulation and was, after adding two ions into the channel, re-solvated and equilibrated. An additional simulation of the same structure was performed that had three ions bound in the central channel (slipped_D_3ions).

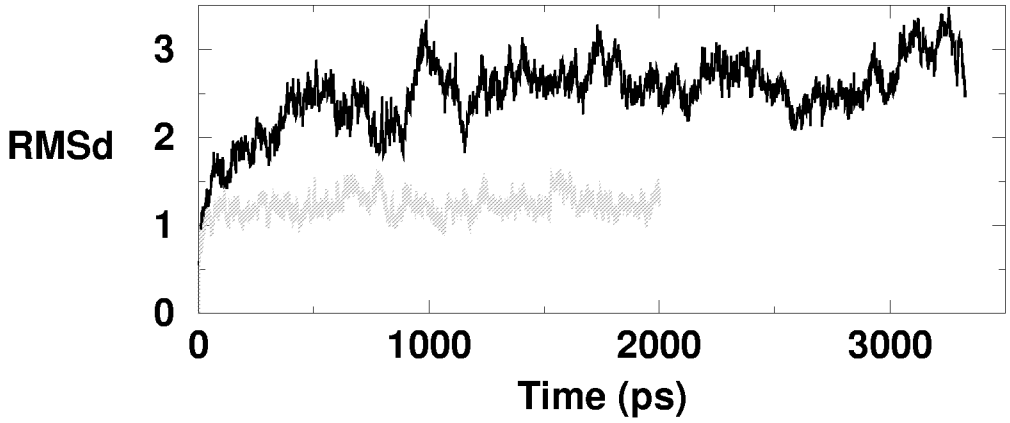

Figure S2: Plot of the RMSd (Å) vs. time for all DNA atoms of the slipped_D_2ions (black) and slipped_D_2ions (gray) models compared to the initial structure.

Two additional models were built by hand which had one strand shifted in the 5' direction and another strand shifted in the 3' direction. These simulations are denoted as shift_AB (Figure 3c) and shift_AD (Figure 3d) and represent shifting of the A and B or A and D strands in the 5' and 3' directions respectively.

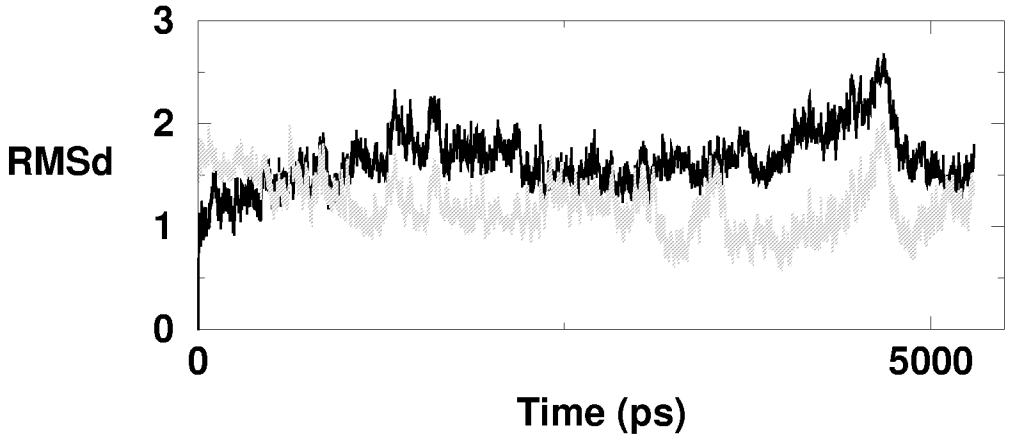

Figure S3: Plot of the RMSd (Å) vs. time for all DNA atoms of the shift_AB model to the initial structure (black) and the 3-5ns average structure (gray).

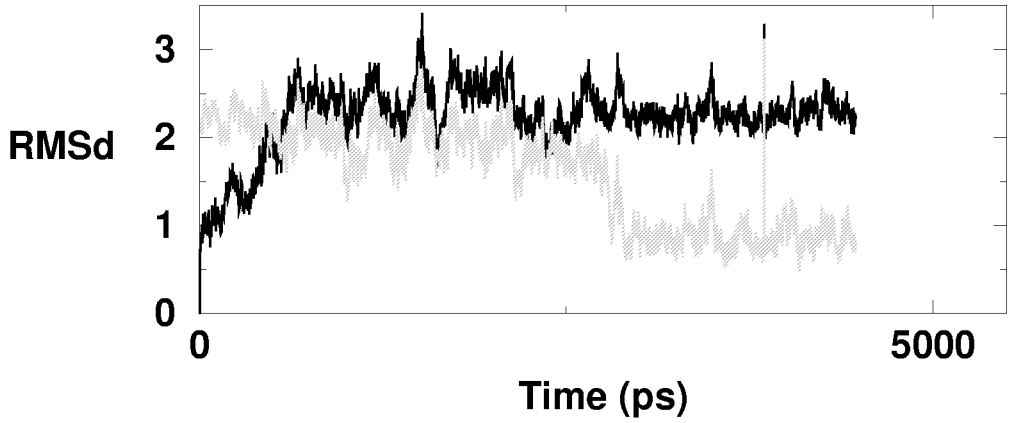

Figure S4: Plot of the RMSd (Å) vs. time for all DNA atoms of the shift_AD model to the initial structure (black) and the 3-5ns average structure (gray).

Spiral stem: As an extreme example of a slipped or shifted structure, a model denoted spiral stem (Figure 3e) was built from the 0.95 Å resolution crystal structure (Phillips et al., 1997). The strands were consecutively shifted by one base pair step towards the 5'-end. The second, third, and fourth strands were shifted with respect to the first strand by one, two and three steps. This spiral stem model thus contains only one complete quartet layer in the center of the molecule, enveloped in both directions by a base triad, a G-G Hoogsteen pair, and an isolated guanine residue.

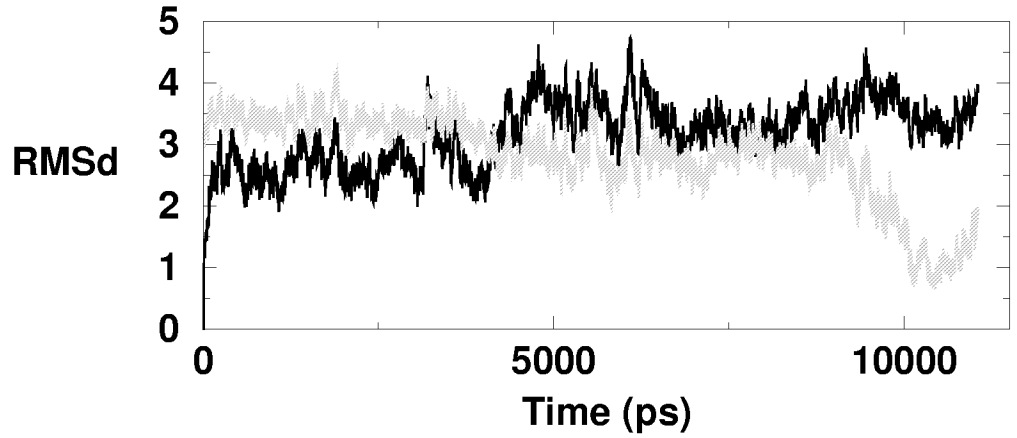

Figure S5: Plot of the RMSd (Å) vs. time for all DNA atoms of the spirala model to the initial structure (black) and the 9-10 ns average structure (gray).

G-DNA triplex: A parallel d$(GGGG)_3$ triplex was created by removing one guanine strand (D-strand, see Fig. 1) from the crystal structure (Fig. 3f). The sodium ions in the channel were kept in the original positions.

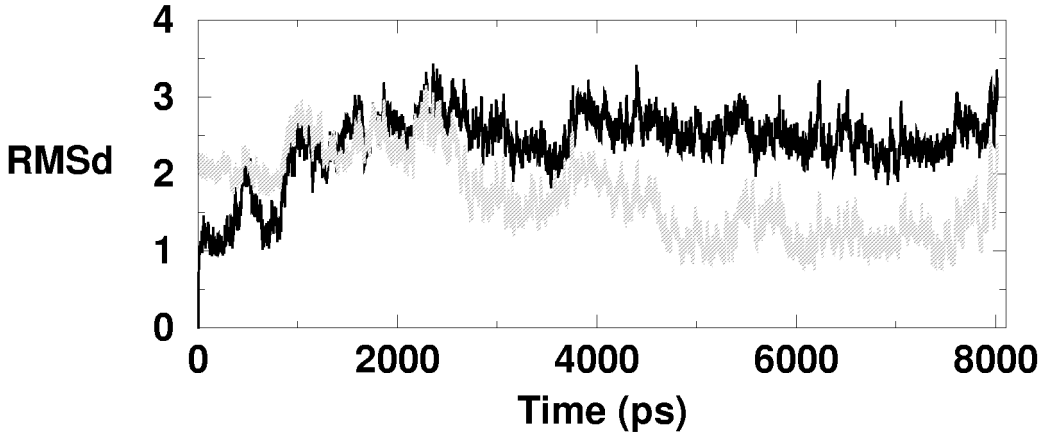

Figure S6: Plot of the RMSd (Å) vs. time for all DNA atoms of the triplex model to the initial structure (black) and the 6-8 ns average structure (gray).

G-DNA dimers: Two dimer structures were investigated. The first parallel $d(GGGG)_2$ edge (Hoogsteen) duplex (G-DNA dimer_AB) was generated based on the crystal structure by removing strands C- and D- (see Fig. 1, 3G). The three Na^+ ions in the channel were retained in their original positions. The two remaining guanine strands (A- and B-strands) thus form one edge of the G-DNA and maintain a duplex base paired by N7•H21 and O6•H1 hydrogen bonds (Fig. 1 and 3).

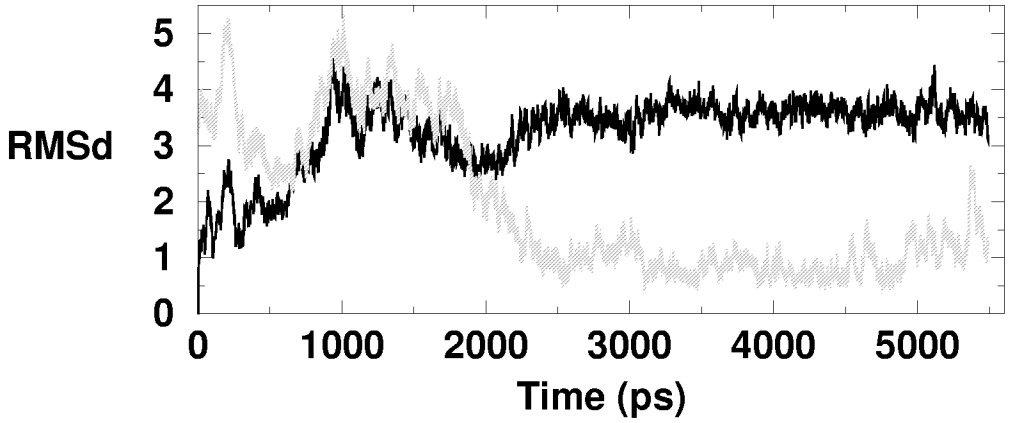

Figure S7: Plot of the RMSd (Å) vs. time for all DNA atoms of the dimer_AB model to the initial structure (black) and the 3-5 ns average structure (gray).

An additional dimer model of G-DNA (G-DNA dimer_AC) was built manually based on an average structure obtained from the previous parallel stranded $d(GGGG)_4$ four-quartet stem simulations averaged over the period of 1-3 ns (Spackova et al., 1999). The B- and D-strands (Fig. 1, Figure 3H), along with all of the counter-ions were deleted and then the strands hand shifted and rotated to reduce the separation between strands while attempting to avoid overlap. Here, we did not position any cations initially at the G-G bases (as justified by our G-DNA triplex simulation, see below). The resulting dimer model of two d(GGGG) strands (A- and C-strands, Fig. 1) forms a "diagonal duplex" base paired by two symmetric O6•H1 hydrogen bonds (Fig. 3). Note, that this symmetrical G•G base pair (abbreviated in the literature as GG1) is the intrinsically most stable arrangement of a guanine dimer, with a gas phase enthalpy of formation of around -24 kcal/mol (Sponer et al., 2000; Sponer et al., 1996) comparable to that of the regular G•C Watson-Crick base pair. This remarkable hydrogen bonding is attributed to an optimal antiparallel alignment of the molecular dipole moments of both guanines correlated with high acidity of the H1 hydrogen site. This favorable electrostatic interaction is counterbalanced by hydration effects, thus the free energy of association of an isolated base pair of this type in water has recently been estimated to be around -0.4 kcal/mol (Florian et al., 1999).

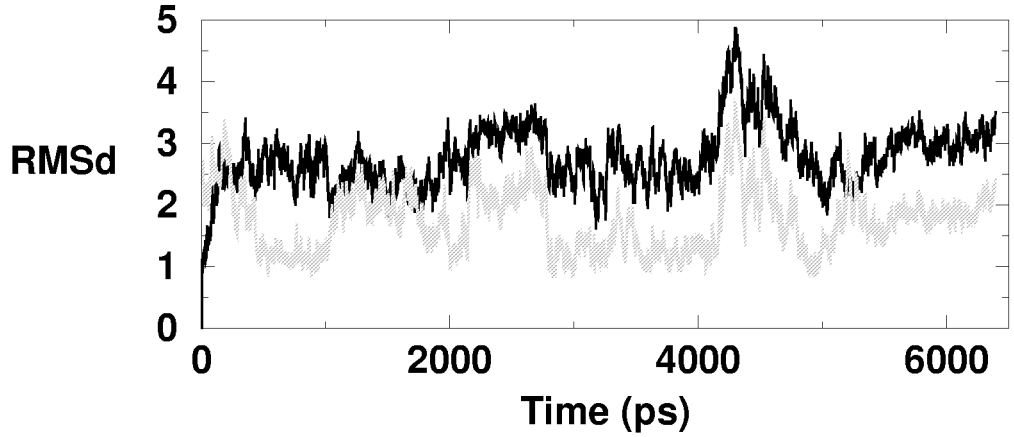

Figure S8: Plot of the RMSd (Å) vs. time for all DNA atoms of the shift_AC model to the initial structure (black) and the 3-6 ns average structure (gray).

Equilibration of the Starting Structures.

The goal of the equilibration phase of the MD simulation is to relax the initial solvent and ion positions and to begin to equilibrate the model nucleic acid structure. This relaxation of the (TIP3P) solvent is relatively rapid and occurs on the 50-100 ps time scale (as judged by convergence of the density, volume and overall potential energy of the system). Complete relaxation of the nucleic acid structure and ionic environment takes longer and there has been some concern, particularly with specific force fields and also with specific nucleic acid systems, that complete equilibration is not obtained even in 1-10 nanosecond-length simulation (Cheatham et al., 2001; Feig et al., 1998). Fortunately, with the Cornell *et al.* force field, relaxation of a nucleic acid model is relatively rapid with complete A-DNA to B-DNA transformations observed in simulations on a 500 ps to 1 ns time scale (Cheatham et al., 1996). The relative rigidity of the native G-DNA quadruplex allows rapid structural relaxation, although ion binding may take longer. However, despite the considerable stability of G-DNA quadruplex structures in MD simulation (Spackova et al., 1999; Spackova et al., 2001), in this study we are able to see convergence on a 2-5 ns time scale of two distinct dimer G-DNA models to a common intermediate and immediate collapse of the trimer model to a more stable form (see below) and this suggests that the present time scale is sufficient for our purposes.

As these models are not based on atomic-resolution X-ray data, we paid considerable attention to a properly relaxing the initial environment to avoid "hot-spots" that may artificially move a structure away from its initial geometry. To support this, flatwell restraints were initially added to enforce hydrogen bonds in the individual structures. The penalty functions were set up as follows. *Donor and hydrogen*: harmonic between 0.0 and 1.8 Å, flat between 1.8 Å and 2.2 Å, harmonic between 2.2 Å and 6.0 Å, and a linear penalty beyond this with a force constant of 15.0 kcal/(mol•Å^2). *Donor and acceptor*: harmonic between 0.0 and 2.8 Å, flat between 2.8 Å and 3.2 Å, harmonic between 3.2 Å and 6.0 Å, and a linear penalty beyond this with a force constant of 15.0 kcal/(mol•Å^2). Minimization was first performed for 500 steps (250 steepest descent followed by 250 conjugate gradient steps) with the restraints in-vacuo applying the Generalized-Born implicit solvent model as implemented in AMBER 6.0. A cutoff of 100 Å was applied (no cutoff). Neutralizing sodium ions were added by the AMBER LEaP program (based on a crude electrostatic potential calculation) and then the whole complex was immersed into a box of pre-equilibrated TIP3P waters. All waters overlapping the solvent were

removed along with any water molecules further than 12 Å for the "spiral stem" model and 10 Å for the trimer and dimer G-DNA model structures.
Equilibration of these solvated systems was performed in cycles of minimization and dynamics as discussed below. During the equilibration, the flatwell "hydrogen bonding" restraints discussed above were maintained. In all of these simulations, the particle mesh Ewald method in AMBER 6.0 was applied with a heuristic pairlist update (and 1.0 Å non-bonded pairlist buffer), a 9 Å cutoff, an Ewald coefficient (kappa) of 0.30768, a FFT grid size of the next largest product of powers of 2,3 or 5 larger than the box size and a uniform bulk density long range van der Waals correction. During the dynamics, the center of mass translation was removed every 5 ps (Chiu et al., 2000; Harvey et al., 1998) and a 2 fs time step was applied. In the MD simulations, SHAKE was applied with a tolerance of 10^{-8} (Ryckaert et al., 1977). Initially 500 steps of minimization (250 steepest descent steps followed by 250 conjugate gradient steps) at constant volume were performed. This was followed by 250 ps of MD simulation where the temperature was ramped from 50K to 300K during the first 25 ps (with a Berendsen temperature coupling (Berendsen et al., 1984) time of 0.2 ps) and then maintained at 300K (with a coupling time of 0.2 from 25 to 50 ps and 2.0 ps after). Constant pressure was maintained with Berendsen coupling (Berendsen et al., 1984) at 1 atm with a coupling time of 0.2 for the first 100 ps and a coupling time of 2.0 thereafter. Despite an overall shrinking of the periodic box size by a small amount, the simulated structures were sufficiently solvated, even after free rotation, to avoid artifacts from direct interaction with the periodic images.
The equilibration of the "spiral stem", triplex and diagonal G-DNA dimer model structures was performed using a similar protocol (Cheatham et al., 1997) and results in similar energy profiles. Subsequent short production runs show the same trends as the systems equilibrated using the protocol above, consistent with the notion that the precise details of the equilibration procedure are not important as long as the solvent and ions are well-equilibrated (Cheatham et al., 2001; Norberto de Souza et al., 1997; Young et al., 1997). Thus we assume that all structures are reasonably well equilibrated.

Production runs: After the initial equilibration, MD simulations were continued with the same parameters as above except that all restraints were removed. Coordinates were written to trajectory files after each picosecond. The lengths of unrestrained simulations are summarized in Table 1. In addition to runs at 300K, a supplementary simulation of the "spiral stem" structure at 400 K was run using NPT (constant pressure) ensemble in order to increase the sampling.

Reliability of the MM_PBSA results and sampling limitations:

A major point of the present study is finding that, for G-DNA molecules, meaningful MM_PBSA results are achieved *only* if ions are included in the analysis explicitly. To date, except for the work by Beveridge and co-workers (Jayaram et al., 1998; Sprous et al., 1998) and analysis of bound ions by Tsui & Case (Tsui et al., 2001), salt effects in this type of analysis have only been included implicitly. When this is done, the salt effects tend to be small. However, in the current work, if the bound ions are not explicitly included, the free energy numbers do not make sense. This finding should have serious implications for future studies of nucleic acid molecules with tightly bound ions.

The free energy data reported in Tables 2 and 3 of the manuscript should be considered as rather crude estimates of free energies since the current continuum model (with Cornell *et al.* charges (Cornell et al., 1995) and PARSE radii (Sitkoff et al., 1994)) is simple and not necessarily fully balanced. This model does give reasonable qualitative estimates, however, the absolute agreement for binding free energies is not perfect (Kollman et al., 2000). It is hoped that with further parameterization of the continuum model and better estimates of the entropy, that better quantitative agreement can be obtained even though it is

well established that continuous models of solvent are in general very sensitive to size and cavity shape definition. However, even with better estimates of the free energy, there are significant limits in conformational sampling that hinder this type of analysis and lead to cause for concern, particularly in longer simulation. The publications to date applying this technology clearly show that reasonable estimates of the free energies may be obtained from short simulations (~200 ps). Part of the reason for this is fortuitous since short simulations do not have the time to overcome significant energy barriers and will not sample all thermally accessible sub-states around the given "minimum" or representative structures. Although this can give an estimate of the free energy for a specific model, this will miss potentially important sub-states that may contribute to the overall free energy. However, this is sometimes a benefit, since larger, but still limited sampling in the 1-10 nanosecond time range, may hit some of the relevant nearby states while missing others. Moreover the time sampled in simulations for each substate may not be representative (i.e. not represent a true Boltzmann equilibria), leading to a potential bias. In other words, various structural fluctuations, such as changes in backbone angles or ion interaction, may lead to changes in the free energy which will need to be properly "weighted" by thermal sampling. In the current work, since the structures are relatively rigid and stable, this does not appear to be a major concern. The trends seen with our free energy estimates are reasonable, represent significant free energy differences, and substantially add to the information extracted from the trajectories. This gives a qualitative ranking of the relative importance of different structures and clear insights into the cation binding. On the other hand, although the methods for estimating the free energies for a given set of representative structures are fairly straightforward and not so dependent on the granularity of the analysis, it is somewhat qualitative and does critically depend on the extent of sampling or movement within that representative region. Ultimately, for proper energetic analysis it will be necessary to cluster the trajectories into representative sub-states and perform the energetic analysis across each sub-state, in addition to looking at the time evolution of the estimated free energies. We do not claim that this free energy technology is the panacea for modeling applications of this sort as there are numerous limitations in the methods and approximations applied.

REFERENCES:

Berendsen H. J. C., J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. 1984. Molecular dynamics with coupling to an external bath. *J. Comp. Phys.* 81:3684-3690.

Cheatham T. E., III, and P. A. Kollman. 1996. Observation of the A-DNA to B-DNA transition during unrestrained molecular dynamics in aqueous solution. *J. Mol. Biol.* 259(3):434-444.

Cheatham T. E., III, and M. A. Young. 2001. Molecular dynamics simulations of nucleic acids: Successes, limitations and promise. *Biopol. Nuc. Acids Sci.* 56:232-256.

Cheatham T. E., III., and P. A. Kollman. 1997. Molecular dynamics simulations highlight the structural differences in DNA:DNA, RNA:RNA and DNA:RNA hybrid duplexes. *J. Amer. Chem. Soc.* 119:4805-4825.

Chiu S.-W., M. Clark, S. Subramaniam, and E. Jakobsson. 2000. Collective motion artifacts arising in long-duration molecular dynamics simulations. *J. Comp. Chem.* 21:121-131.

Cornell W. D., P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. 1995. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Amer. Chem. Soc.* 117(19):5179-5197.

Feig M., and B. M. Pettitt. 1998. Structural equilibrium of DNA represented with different force fields. *Biophys. J.* 75:134-149.

Florian J., J. Sponer, and A. Warshel. 1999. Thermodynamic parameters for stacking and hydrogen bonding of nucleic acid bases in aqueous solution: Ab initio/Langevin dipoles study. *Journal of Physical Chemistry B* 103(5):884-892.

Harvey S. C., R. K.-Z. Tan, and T. E. Cheatham, III. 1998. The flying ice cube: Velocity rescaling in molecular dynamics simulations leads to violation of equipartition. *J. Comp. Chem.* 19:726-740.

Jayaram B., D. Sprous, M. A. Young, and D. L. Beveridge. 1998. Free energy analysis of the conformational preferences of A and B forms of DNA in solution. *J. Amer. Chem. Soc.* 120:10629-10633.

Kollman P. A., I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case, and T. E. Cheatham, III. 2000. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.* 33:889-897.

Norberto de Souza O., and R. L. Ornstein. 1997. Effect of warmup protocol and sampling time on convergence of molecular dynamics simulations of a DNA dodecamer using AMBER 4.1 and particle mesh Ewald method. *J. Biomol. Struct. Dyn.* 14(5):607-611.

Phillips K., Z. Dauter, A. I. H. Murchie, D. M. J. Lilley, and B. Luisi. 1997. The crystal structure of a parallel-stranded guanine tetraplex at 0.95 angstrom resolution. *J. Mol. Biol.* 273:171-182.

Ryckaert J. P., G. Ciccotti, and H. J. C. Berendsen. 1977. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comp. Phys.* 23:327-341.

Sitkoff D., K. A. Sharp, and B. Honig. 1994. Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.* 98:1978-1988.

Spackova N., I. Berger, and J. Sponer. 1999. Nanosecond molecular dynamics simulations of parallel and antiparallel guanine quadruplex DNA molecules. *J. Amer. Chem. Soc.* 121:5519-5534.

Spackova N., I. Berger, and J. Sponer. 2001. Structural dynamics and cation interactions of DNA quadruplex molecules containing mixed guanine/cytosine quartets revealed by large-scale MD simulations. *J. Amer. Chem. Soc.* 123:3295-3307.

Sponer J., and P. Hobza. 2000. Interaction energies of hydrogen-bonded formamide dimer, formamidine dimer, and selected DNA base pairs obtained with large basis sets of atomic orbitals. *Journal of Physical Chemistry A* 104(19):4592-4597.

Sponer J., J. Leszczynski, and P. Hobza. 1996. Structures and energies of hydrogen-bonded DNA base pairs. A nonempirical study with inclusion of electron correlation. *Journal of Physical Chemistry* 100(5):1965-1974.

Sprous D., M. A. Young, and D. L. Beveridge. 1998. Molecular dynamics studies of the conformational preferences of a DNA double helix in water and an ethanol/water mixture: Theoretical considerations of the A->B transition. *J. Phys. Chem. B.* 102:4658-4667.

Tsui V., and D. A. Case. 2001. Calculations of the absolute free energies of binding between RNA and metal ions using molecular dynamics simulations and continuum electrostatics. *J. Phys. Chem. B* 105:11314-11325.

Young M. A., G. Ravishanker, and D. L. Beveridge. 1997. A 5-nanosecond molecular dynamics trajectory for B-DNA: Analysis of structure, motions and solvation. *Biophys. J.* 73:2313-2336.