CALTECH

New approaches to accurate predictions of free energies: from proteins to organic nanostructures

Thesis by
Tod A. Pascal

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California
2010
(Defended October 7th 2009)

To mom, who is no longer with me
To mom and dad, who always will be

ACKNOWLEDGEMENTS

I would like to acknowledge my advisor Bill Goddard for his patience and guidance throughout my tenure at Caltech. His passion for science is usually cited as his greatest asset. For my part however, it is his inherent kindness and common decency, his basic humanity, which makes him an exemplary advisor and teacher.

I have been blessed to learn from the diverse and extremely talented group of scientists in the WAG group. To my friend Prabal Maiti, whose help in the atomistic simulations of DNA nano-structures we invaluable. Many thanks to Valeria Molinero, Mario Blanco and Shiang-Tai Lin, for their thoughtful discussion and direction. Special thanks goes to Vaidehi Nagarajan, under whose stewardship and watchful eye I made the necessary early mistakes that is a graduate career.

A hallmark of our work has been the close interaction with experimentalist, on whose problems we have attempted to elucidate the underlying physics and chemistries. Prof. Nadrian Seeman (NYU) and Prof. Erik Winfree (Caltech) were invaluable for direction on the DNA nanotechnology project. For our work on DNA repair, Dr. Larry Sowers has acted as our voice of reason, providing tremendous insights into the biology of protein/DNA. Dr. Nemani Prasadaro (Children's Hospital LA/USC) provided the passion and expertise that was required on the bacterial meningitis project.

To my family and friends, who have endured with me through the missed holidays and forgotten birthdays. To Lydia: you been with me through this entire journey and this is as much your triumph as it is mine.

My tenure at Caltech was funded by fellowships from the National Science Foundation and the Department of Energy Computational Science Graduate Fellowship (DOE-CSGF).

ABSTRACT

Much of our understanding of solvation thermodynamics arise from explicit atom computer simulations. For condensed phase systems, this approach in principle presents the most accurate and realistic model for such investigations. Monte-Carlo (MC) and Molecular-Dyanamics (MD) trajectories, subjected to the machinery of statistical thermodynamics, can in principle provide all the necessary thermodynamics quantities of interest, provided that the interatomic potentials are adequate and the system being considered is properly equilibrated.

These explicit-atom simulations are inherently computationally intensive, and are intractable for investigations of realistic biomolecular systems on a timescale (usually from micro-seconds to hours) necessary to observe critical transitions. For many problems, the full atomic model has to be abandoned and empirical or mean field models that require less computational power must be adopted instead. To this end, we've developed a coarse-grained model of DNA that is two orders of magnitude faster than standard explicit atom MD simulations.

Although computer simulations have been applied to the calculation of solvation free energies for a wide range of molecules, reliable calculations including explicit consideration of entropy and quantum effects (zero-point energy corrections) are less common, owing to the enormous computational effort required by standard perturbation methods. Accurate calculations of entropies are essential if computer simulations are to become more useful tools for obtaining molecular insights into solvation and ligand binding phenomena.

We report on the extension of a method of calculating exact entropies and quantum effects from standard MD simulations. This novel method is applied to the investigation of four

model cases: 1) the folding of a DNA three-way junction 2) stability of DNA nanostructures 3) the efficacy of binding in a protein-protein interaction, critical in the pathogenesis of bacterial meningitis in neonates and 4) the free energy of water molecules at two extreme surfaces as an investigation of the hydrophobic effect. By developing a scheme to partition the entropies and enthalpies into the per-atom components, we show that the water molecules in the first solvation shell have unique character, and are critical in understanding the underlying physics in these systems.

CONTENTS

Acknowledgements iii
Abstract v
List of Figures viii
Section I. Introduction 2
References 8
Section II. Thermodynamic stability of Bio-molecular systems 9
Chapter 1: Thermodynamics of the DNA 3WAY Junction 10
Chapter 2. Interaction between OmpA and Ecgp-96: Role in Bacterial Meningitis ... 79
Section III. DNA nanostructures 120
Chapter 3. The PX/JX DNA Crossover Structures 121
Section IV. Exploring The Hydrophobic Effect 143
Chapter 4. The QM-FF force field for Carbon 144
Chapter 5. The QM-FF force field for Graphite – water interactions 205
Chapter 6. Properties of water molecules at interfaces 220
Appendix I. Theoretical Framework 221
A. General Simulation scheme 222
B. 2PT-FE-MD method for accurate thermodynamics 223
C. Obtaining Interaction Free Energy of DNA from First Principles 228
Appendix II. Meso-scale DNA Dynamics 238
i) Description of Meso-scale Model 239
ii) Parameter Development 240
ii) Meso-scale simulation of B-DNA dodecamer 243
iii) Conclusions 248

LIST OF FIGURES

Chapter 1

Figure 1 ***(a)*** **Schematic of DNA TWJ1 showing each of the three arms (A, B and C in the shaded circles) and the 3 characteristic orientation angles (Φ, ψ and θ) used to determine the helical stacking preference. A/B stacking corresponds to ψ ≤ 60°, and Φ, θ ≈ 90°, whereas A/C stacking corresponds to Φ ≤ 60°, and ψ, θ ≈ 90°. Note that there is a third stacking configuration (θ ≤ 60°, and ψ, Φ ≈ 90°) that is termed B/C stacking.** ***(b)*** **TWJ1 nucleotide sequence with unique arms indicated as before (capital letter in shaded circles). Base sequence numbers are indicated with red numbers next to the nucleotide. The unpaired thymines at branch point and capping arms A and B are underlined. As drawn, this structure corresponds to A/B stacked T-shaped conformer** ***(c)*** **Ribbon structure of the A/B stacked starting structure (conformer 1) from the NMR ensemble. This conformer (used as the starting structure for the TWJ1_Na and MG_A MD simulations) was calculated to have had the second lowest average CRMS to the other conformers. Further, the stacking angle of unique arms A and C is 53°, closest to the experimentally determined average of 53° ± 11°. 53**

Figure 2 From left to right: distribution of ϕ, ψ, θ angles (degrees) for various TWJ1 simulations during 100 ns NVT MD simulation (25ns for MG_B). Note the multi-modal distribution in the ϕ and θ angles of TWJ1_Na, representing the two conformers: Na_aP (ϕ ≈ 60°, θ ≈120°) and Na_P (ϕ ≈ 150°, θ ≈ 25°), with both conformers showing A/B stacking (ψ ≈ 20° ± 7.3°). On the other hand, MG_A exhibits A/C stacking (ψ/ θ ≈ 90°) with Gaussian like distributions for all three dihedral angles. The relative flexibility of TWJ1_Na (compared to the structurally more rigid MG_A) is reinforced by comparing the width of the various distributions. MG_B converged to the average MD structure of MG_A within 20ns of NVT simulations, and was therefore only simulated for 50ns. Analysis of the dihedral angle show similar distributions to that of MG_A, as expected, with half the frequency. 54

Figure 3 Plot of cross correlation function of each set of the three unique dihedral angles characterizing the stacking preference of TWJ1. Cross correlations of ϕ/θ (red) , ϕ/ψ (black) and ψ/θ (blue) are plotted as a function of correlation time (ns). (a) TWJ1_Na dihedral angle cross-correlation. The ϕ/θ angles move in a highly correlated fashion. This indicates an A/B helix exploring its conformation spaced as a rigid body stacked on arm C (A/B stacking). (b) The dihedral angle cross-correlation of MG_A. Here, unlike TWJ1_Na, the ϕ/ψ angles are highly correlated, indicative of the underlying A/C stacking configuration of MG_A. 55

Figure 4 Best MD structure for of TWJ1_Na (a – c), MG_A (d) and MG_B (e and f) (a) NMR/starting structure. Each snapshot is rotated so as to minimize the total CRMSD (excluding the unpaired thymines), thereby illustrating the differences in the stacking of helix 2 on helix 1. Conformer 1 of the 26 NMR ensemble was taken as representative and used as starting structure for TWJ1_Na and MG_A simulations. A/B stacking is observed, with an acute A/C Φ stacking angle of 53°. (b) Anti-parallel conformer Na_aP. It is structurally similar to NMR structure with acute Φ stacking angle of 64°. (c) Parallel conformer Na_P. It also shows A/B stacking with obtuse acute A/C Φ stacking angle of 143°. During dynamics, TWJ1_Na adopts the Na_aP conformation for the first 25ns. There is then a transition to Na_aP between 27 and 32 ns. Na_P is assumed to be the true solution structure (at 300K), as TWJ1_Na stays in this conformation for the remaining 60ns of dynamics. We later show that the experimentally similar Na_aP structure would be more favorable at 285K. (d) MG_A solution structure. This structure is more structurally compact than any of the TWJ1_Na structures and has a different topology: it exhibits A/C stacking. (e) Starting structure of MG_B. MG_B is the only simulated TWJ1 structure that was built from scratch, using standard B-DNA geometries. (f) MD Solution structure of MG_B, showing convergence to the solution structure of MG_A (d) after 25ns. 56

Figure 5 Cartesian root mean-squared deviation (CRMS) of (a) TWJ1_Na during dynamics, using the three best average md structures as reference: NMR/starting structure (black), Na_aP (red), Na_P (blue). The NMR/starting structure and Na_aP profiles are highly correlated during dynamics, further indicating the similarities of both structures. The transition from Na_aP to Na_P occurs between 37 and 42 ns, after which time, TWJ1_Na remains in the Na_P conformation (b) MG_A during dynamics, using the average md structure (red) and the starting NMR structure (black) as references. MG_A converges to it solution structure after 45ns of simulation, which is 6Å away from the starting NMR structure (c) Comparison of the

CRMSD of MG_A during dynamics using the TWJ1_Na solution conformers as reference: Na_aP (black) and Na_P (red). MG_A is topologically different from either TWJ1_Na conformers, exhibiting B/C stacking. It is however more similar to Na_aP (6.42Å) than Na_P (12.26Å). (d) CRMS of MG_B during dynamics using the starting (Namot2 built) structure (black), the MD solution structure (blue) and the MD solution structure of MG_A (red) as references. Note that MG_B converges to with 2.5 Å of the global MG_A structure after 30ns of dynamics, after starting 7Å away. 57

Figure 6 (a) Comparison of the radius of gyration of TWJ1_Na (black) and MG_A (red) during dynamics. After equilibrations, MG_A has a radius of gyration of 14.9Å, whereas Na_aP (12 – 35 ns) is 17.5Å and Na_P (42 – 100ns) is 16.3Å. (b) Comparison of the SASA of TWJ1 during dynamics. Na_P has a smaller SASA (6747 $Å^2$) than Na_aP (6901 $Å^2$), indicating a more compact structure. This is consistent with the radius of gyration measurements, and the fact that Na_P has a smaller A/C stacking angle than Na_aP. MG_A is the most compact (6588 $Å^2$) of the TWJ1 structure, which when combined with having the smallest radius of gyration indicates that MG_A is the most structurally rigid of all the TWJ1 conformers. 58

Figure 7 RDFs of Na_aP (black), Na_P (red) and MG_A (blue) during dynamics. (a) PO_4^- — water. A magnification of the first hydration shell is shown in the upper left corner. TWJ1_Na has slightly more water in the first solvation shell than the more compact MG_A. Two distinct hydration shells are found, the 1st from 2.5 – 4.6Å and the 2nd from 4.6 – 6.5Å (b) PO_4^-—ion RDFs using the different counterions. The RDFs for the TWJ1_Na structures have been multiplied x10 for visualization purposes. Note that the double charged counter-ion Mg^{2+} is strongly coordinated to the DNA backbone (13 of the 17 ions in the 1st shell), whereas TWJ1_Na has as many Na^+ ions in the 1st hydration shell (3 and 5 for Na_aP and Na_P respectively) as in the second, with approximately 90% of the ions at least 1 solvation shell away from the DNA. 59

Figure 8 Two-dimensional projection of water in different groups around the starting NMR TWJ1 structure, used in the entropy calculation. Waters in the solvation shells are obtained from observing the water-phosphate RDF. Here all waters within 4.6Å of the phosphate backbone (249 molecules) are considered in shell 1, whereas all waters within 6.8Å of the phosphate group (380 molecules) are considered in shell 2. Water molecules centered at distances greater than 7.0Å from the TWJ1 backbone (7780 molecules) are considered as bulk water. The unit cell is shown by the enclosing blue box. 60

Figure 9 Translational (a – c) and rotational (d - f) components of the density of state for various types of water molecules in the Na_aP (a,d), Na_P (b,e) and MG_A (c,f) ensembles respectively. These calculations are done at 300K. We see an increase in the translational entropy of the waters in the 2nd hydration shell compared to bulk, whereas the waters in the 1st hydrations shell are more constrained, both translationally and rotationally. Shown are the contributions obtained from decomposition of the total atomic velocities (equations 11 – 13) for water molecules in the 1st solvation shell (solid black line), 2nd solvation shell (red dashed line) and bulk waters (dashed blue line). The vibrational contributions to the velocity are all zero, since the SHAKE algorithm was used to constrain the bonds and angles of the water molecules to their equilibrium values. 61

Chapter 2

Figure 1 Overview of OmpA structure used in this study. 100

Figure 2 Structure of the chitobiose/OmpA (WT) complex after 30 ns of MD in explicit solvent, counterions, and membrane. These MD studies started with the two structures from docking: Pose L (for Loops) with chitobiose bound to Loops 1, 2 and 3) Pose B (for Barrel) with chitobiose bound to Loop 4. 101

Figure 3 Snapshots of OmpA/chitobiose complexes during the MD, with calculated thermodynamic quantities (kcal/mol) relative to the starting structure (A=Helmholtz free energy, E_0=Zero-Point corrected energy and TS = temperature times entropy). *(a)* WT OmpA. The chitobiose in both L and B regions are flexible but remain strongly bound. The largest change in free energy is < 1.6% of the total. *(b)* 1b mutant. The chitobiose in region L gets ejected from the structure after 6 ns: the calculated energy (enthalpy) is 3.1% less favorable than for WT, but the total entropy (the TS term) is more favorable by 2.3%, due mainly to the increased motions of the chitobiose. The net result is that the free energy is only 0.7% less favorable than the WT. After 10ns, the system has a less favorable free energy than WT by 16% due to the less favorable energy (20%), but more favorable TS (3.7%). Similar results are found for the 1a mutant (increase in A of 26%, becoming unbound after 6.7 ns). Indeed our experiments show no invasion of the bacteria experimentally for mutants 1b and 1a. *(c)* Helmholtz free energy profile for WT and 10 mutants during dynamics. The energies are reported relative to the free energy of the system at 2ns. The profile of

the 1a and 1b mutants is unique: increasing until 6ns then falling dramatically once the chitobiose becomes unbound. 103

Figure 4 *(a)* Correlation between free energies from theory and experimental invasion activity. Abscissa: experimental invasion activities (% relative to WT) of the nine OmpA mutants. Ordinate: Filled red squares – Calculated free energies (kcal/mol) relative to WT, Filled blue diamonds – calculated energies (enthalpies). The special cases of WT and the 1a (filled red and blue circles), 1b (filled red and blue triangles) and 2c (open symbols) mutants are shown. For the 2c mutant open symbols are shown since there are no experimental data (the abscissa value chosen based on the fit, which constitutes a prediction). The lines are fitted to the 8 cases for which there are experimental data. The 1a and 1b mutants are predicted not to invade since the chitobiose in region L becomes unbound during dynamics. This is consistent with observation. We find excellent correlation of the free energy to the experimental invasion (92%). Ignoring the entropy (energy curve) leads to a poorer correlation (79%). *(b)* Comparison between experimentally observed invasions, and that predicted from the free energy best-fit line. The error bars are shown. We correctly predict no invasion for the 1a and 1b mutants since they are unbound. Overall we find excellent agreement between the experimentally observed and predicted invasion values. 104

Chapter 3

Figure 1 The base pair sequences used in the generations of PX55, PX65, PX75 PX85, and PX95. 134

Figure 2 Computer generated built initial structure for PX55, PX65, PX75 PX85, and PX95. 135

Figure 3 structures for various PX molecules averaged over the last 10 ns of molecular dynamics (a) in presence of Na(+) and (b) in presence of Mg(+2). For clarity water molecules and counter ions are not shown. 136

Figure 4 Variation of the coordinate rms displacement (CRMSD) of all atoms of various snapshots from the MD simulation run with respect to the average dynamics structure (a) in presence of Na+ and (b) in presence of Mg(+2) 137

Figure 5 CRMSD for individual bases for PX55, PX65 PX75, PX85 and PX95 from the starting dynamics structure, in (a) Na+ and (b) Mg(+2). The data has been averaged for last 200 ps of the dynamics run. 138

Figure 6 Average Rise, Tilt, Roll, and twist for for PX55, PX65 PX75, PX85 and PX95 from the starting dynamics structure, in (a) Na+ and (b) Mg(+2). Solid line is for helix1 and broken line is for helix2. The vertical lines corresponds to the crossover points. The data has been averagred over last 200 ps of the 3ns long dynamics. The horizontal solid lines gives the upper bound and lower bound for the corresponding quantities expexted for the helices in their B-DNA form (non-crossover form) during the dynamics. This shows that PX65 retains the B DNA much better than PX95.. 139

Figure 7 Average major groove width and minor groove width for PX65 in presence of Mg(+2) Solid black and blue lines represent major groove and minor groove width respectively for helix1. Broken black and blue lines is for helix2. The data have been averagred over last 200 ps of the 3ns long dynamics 140

Figure 8 Strand shortening as defined in text for various PX molecules in (a) Na+ and (b) Mg(+2). The data have been averaged over last 200 ps of the dynamics run. 141

Figure 9 Global bend angle calculated for each double helix of each PX structure. The error bars indicate the fluctuations that occur in the molecular dynamics simulation. 141

Figure 10 Solvent accessible surface area (SASA) per base pair for various PX molecules. The area has been avergaed over last 200 snapshots of the dynamics. The area has been calculated by a very fast and accurate AVGB alogorithm [1] 142

Figure 11 Number of water in the first solvation shell (within a shell of 2 Å). 142

Chapter 4

Figure 1. Side projection of the lowest energy structure for each set of coronene structure (PD-X – (a) and PD-Y – (c)) used to obtain the van der Waals parameters for QM-FF. The top view of each structure is shown to the right. Also shown are corresponding grapite structure (c) and the high energy eclipsed structure (d). For the PD-X and PD-Y structures, the displacement vectors values are given. The dispersion curves are obtained by varying the z coordinate of the top coronene structure while holding the bottom fixed. 185

Figure 2. PD-X van der Waals curves for QM-FF (X6, Morse and LJ) compared to M06-2X. The optimization procedure minimized the differences for the 5 points around the minima, combined with those of the PD-Y structure (the curve for the PD-Y structure is similar and not shown). 186

Figure 3. (a) Potential energy surface and (b) contour plot of sliding for the top layer of graphite, obtained using the QM-FF X6 potential. The unit cell has two layers of 800 carbons each. The low energy (graphite like) PD-X and PD-Y structures are indicated. 187

Figure 4. Phonon dispersion curve for all vibrational modes of (hexagonal) graphite at 0K using the X6 potential. Experimental results (300K) from inelastic x-ray scattering are shown: red triangles from reference[122], black circles from reference[112] 188

Figure 5. Phonon dispersion curves for the low frequency modes of (hexagonal) graphite at 300K. Experimental data (symbols) from reference[79-80,112,122]. 189

Figure 6. (a) Vibrational entropy (S) of graphite (b) Vibrational energy (V) and free energy (F) of graphite 190

Figure 7. Specific heat of hexagonal graphite as computed with the QM-FF X6 potential. Low temperature (< 10K) results obtained from the thin-plate approximation, other results obtained from the uniform grid method. Experimental results from different sources are indicated, as reported in references[88,92].). Values for rhombohedral graphite is not plotted, since the lines would be essentially super-imposed on the hexagonal graphite lines. 190

Figure 8. (a) In-plate (a) lattice parameter of graphite as a function of temperature, calculated with the QM-FF X6 potential. The experimental results[115] (red squares) are compared the calculated values (black triangles). The solid black line is the least squares line by cubic spline regression. (b) Calculated in-plane coefficient of thermal expansion. The experimental results (red and blue squares) are shown, as reported in references[77,117]. The best fit of the data (black line) is used to compute the coefficient of thermal expansion in the *ab* and *c* directions. 191

Figure 9. Phonon dispersion curves for the low frequency modes of rhombohedral graphite at 0K (using the X6 potential with ζ = 17.6) 192

SECTION I. INTRODUCTION

Accurate prediction of the entropy and enthalpy are an important complement to free energy, providing additional information to help understand and interpret the connection between molecular perturbations and thermodynamic changes in condensed phases[2]. They are thermodynamic state functions and are independent of the thermodynamic path connecting two thermodynamic states of the system. The decomposition of the free energy into enthalpic and entropic components is, therefore, always thermodynamically meaningful.

More importantly, enthalpies and entropies are experimentally measurable quantities. The comparison between calculated and experimental enthalpies and entropies may provide additional physical insights and may be used as benchmarks to optimize force fields for condensed-phase simulations.

In this thesis, a fast, accurate method of calculating entropies of condensed phase system is validated by several calculations on biological systems and organic nanostructures. In this method[3-6] (the two phase thermodynamic method – 2PT), the vibrational entropy of the system is described completely and efficiently from the Fourier transform of the atomic velocity autocorrelation function (VAC), with corrections for fluidic effects. Statistics for this method are collected over 20ps of a molecular dynamics (MD) simulation, in which the velocities and coordinates are saved every 4 fs (must be shorter than the fastest vibrational levels which have periods of ~10 fs for a 3000 cm-1 vibration). From this 20 ps trajectory we calculated the VAC for each atom,

$$C(t) = \sum_{j=1}^{N} \sum_{k=1}^{3} m_j \left[\lim_{t \to \infty} \frac{1}{2\tau} \int_{-\tau}^{\tau} v_j^k(t' + t) v_j^k(t') dt' \right] \quad (1)$$

where $v_j^k(t)$ is the *k-th* component of the velocity of atom *j* at time *t*.

Then we obtained the vibrational density of states (DoS) $\hat{C}(v)$ (also referred to as the power spectrum or spectral density) from a fast Fourier Transform (FFT) of equation (1). Correction to the low frequency, anharmonic modes are performed by splitting the total DoS into a component related to a "gas-like" diffusion term $\hat{C}(v)_{diff}$ (modeled as a hard sphere) and a "solid-like" vibrational term $\hat{C}(v)_{vib}$ (modeled as a vibrating crystal):

$$\hat{C}(v) = f\hat{C}(v)_{diff} + (1-f)\hat{C}(v)_{vib} \quad (2)$$

where *f* is the fraction of the 3N total modes corresponding to the fluid or diffusional parts of the dynamic system.

The partition function *Q* (from which all thermodynamic quantities are calculated) is obtained by applying quantum statistical mechanics, for example the entropy is given by

$$S = k\,lnQ + \beta^{-1}\left(\frac{\partial lnQ}{\partial T}\right)_{N,V} = k\int_0^{\infty} dv\left[\hat{C}_{diff}(v)W_{diff}^S(v) + \hat{C}_{vib}(v)W_{vib}^S(v)\right] \quad (2)$$

In this 2PT paradigm, the entropy is intrinsic and quantifiable, completely described by the vibrational modes of the system, without the need for an external reference state. This last point is crucial, and represents a departure from the fundamental theories underlying other methods of calculating free energies. Details of this method are presented in appendix I.3. We present the details of this method as well as our overall computational framework in Appendix I.

In Section II, we apply these methods to studying the free energy of bio-molecules. We investigate the folding thermodynamics of a DNA Three-way Junction (TWJ1) in chapter 1. TWJ1 is a single strand of DNA that forms a "T-shaped junction" [8-10] characterized by three separate arms, whose relative orientations determine the three dimensional structure. Two of

the arms form a coaxially stacked helix, connected to the other arm by a bulge region of two unpaired thymine nucleotides. Our analysis of the three characteristic orientation angles during long term (more than 100ns), explicit solvent MD simulations shows that TWJ1 adopts a configuration with arms A and B forming a continuous helix (A/B stacking) for solvents with a low NaCl concentration, while the same system with Na^+ replaced with Mg^{2+} system shows A/C stacking. These results are consistent with the stacking behavior observed in RNA/DNA TWJs and DNA four-way (Holliday) junctions. Our analysis indicates that TWJ1 folds so as to maximize the entropy of the water molecules in the first solvation shell, decreasing the interfacial free energy. This interfacial entropic contribution, combined with the minimized DNA backbone electrostatic repulsion, determines the optimal packing. Replacing Na^+ counter ions with divalent Mg^{2+} makes electrostatic interactions dominant: condensation of the Mg^{2+} ion on the DNA backbone provides a much more compact structure than for Na^+. For the Na^+ case we find a second, more compact structure conformer with better entropy than the experimental structure. We predict this structure is stabilized at higher temperatures and could be observable experimentally.

In Chapter 2, we present a free energy model for predicting the efficiency of Escherichia coli K1 invasion into Human Brain Microvascular Endothelial Cells and experimental validation. We describe here a new approach for predicting the free energy of binding of ligands to proteins, showing that it correlates quite well with new experiments on the invasion efficiencies of *E. coli* OmpA and 9 of its mutants with Ec-gp96, a human brain microvascular endothelial cell (HBMEC) receptor. OmpA is known to interact with the GlcNAc1-4GlcNAc (chitobiose) epitopes on Ec-gp96 to facilitate bacterial meningitis in neonates. The excellent correlation between

experiments and the total free energy of chitobiose bound to OmpA and its mutants validates the use virtual screening against predicted free energies to identify new leads for experimental synthesis and testing. Detailed thermodynamic analysis of the residues in the binding sites suggests the specific of amino acid in OmpA critical in binding, a crucial step toward designing small molecules that can inhibit bacterial meningitis in neonates.

In Section III, we present an overiew of the thermodynamic stability of DNA based nanostructures. We use MD to understand the structure, and stability of various Paranemic crossover (PX) DNA molecules and their topoisomer JX molecules, synthesized recently by Seeman and coworkers at New York University (NYU). Our studies include all atoms (4432 to 6215) of the PX structures with an explicit description of solvent and ions (for a total of up to 42,000 atoms) with periodic boundary conditions. We report the effect of divalent counterions Mg(+2) on the structural and thermodynamic properties of these molecules and compare them to our previously reported results in presence of monovalent Na+ ions.

The dynamic structures averaged over the 3-nanosecond simulations preserves the Watson-Crick hydrogen bonding as well as the helical structure. We find that PX65 is the most stable structure both in Na+ and Mg(+2) in accordance with the experimental results. PX65 has helical twist and other helical structural parameters close to the values for normal B-DNA of similar length and sequence. Our strain energy calculations demonstrate that stability of the crossover structure increases with the increase in crossover points. Included in Appendices II and III are two other papers we've published on this topic.

In Section IV we turn our attention to evaluating the entropy and free energy of water molecules at interfaces. We look at two extreme surfaces 1) a 1-palmitoyl-2-oleoyl-

phosphatidylcholine (POPC) membrane, as a model for a hydrophilic surface (Chapter III.3) and 2) graphene as a model for a hydrophobic surface (Chapter III.4). We observe differences in the calculated properties (entropy, diffusion constants, dielectric constants, surface tension) of water molecules at these interfaces, with the water molecules in the first hydration shell displaying dramatically different character than water molecules in the subsequent hydration shells and in the bulk.

Since there are few experimental results to compare to our results, we validate our results by developing *ab-initio* force fields (QM-FF) basely on quantum mechanical electronic structure calculations. We show in Chapter IV.1 that the QM-FF forcefield fitted to results from M06-2X DFT calculations on the coronene dimer system leads to accurate predictions of available experimental mechanical and thermodynamics data of graphite (lattice vibrations, elastic constants, Poisson ratios, lattice modes). Parameters for carbon are obtained, suitable for accurately describing various graphitic systems with standard molecular mechanics, monte-carlo codes. The carbon nonbonded interactions are modeled as an analytic Exponential-6 potential. QM-FF is used to predict the phonon dispersion curves, specific heat, thermal expansion and other properties of graphite, leading to good overall agreement with available experimental data.

In Section IV.2, we develop a QM-FF forcefield suitable for describing the interaction of water molecules on graphite. This forcefield is the first to distinguish between the three binding modes on water on graphite, with the alpha orientation (both protons pointing into the graphite surface) most stable with a binding energy of 2.47 kcal/mol, while the beta orientation (both protons pointing away from the surface) the least stable with a binding energy of 2.27

kcal/mol. Accurate description of the water-graphite interactions is critical in reproducing the observed contact angle on water on graphite. Our potential predicts a contact angle of 85^{o}, within the range of the best experimental results (xx to xx).

Efficient codes, employing parallel computing and fast methods of computing the non-bonded interactions present in these molecules are to be credited for the increase in simulation time recently. However, in order for realistic simulations to be performed in a reasonable amount of time, there is a need to have an alternate description of these molecules, a description that is preferably based on atomistic simulations, just as most force fields for atomistic simulations are based on quantum mechanical information.

We present our first attempt at development of a coarse grained DNA forcefield, based on MD simulations of solvated DNA molecules. We show that this simplified model reproduces much of the structural characteristics of atomistic DNA molecule, allowing simulations two orders of magnitude faster than standard MD. This forcefield and the associated methods are presented in Appendix II.

References

(1) Levy, R. M.; Gallicchio, E. *Annu. Rev. Phys. Chem.* **1998**, *49*, 531.
(2) Lin, S. T.; Blanco, M.; Goddard, W. A. *J. Chem. Phys.* **2003**, *119*, 11792.
(3) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187.
(4) Kirkwood, J. G. *The Journal of Chemical Physics* **1935**, *3*, 300.
(5) Kong, X.; Brooks Iii, C. L. *The Journal of Chemical Physics* **1996**, *105*, 2414.
(6) Darve, E.; Pohorille, A. *The Journal of Chemical Physics* **2001**, *115*, 9169.
(7) van Buuren, B. N. M.; Overmars, F. J. J.; Ippel, J. H.; Altona, C.; Wijmenga, S. S. *J. Mol. Biol.* **2000**, *304*, 371.
(8) Ouporov, I. V.; Leontis, N. B. *Biophys. J.* **1995**, *68*, 266.
(9) Overmars, F. J. J.; Pikkemaat, J. A.; vandenElst, H.; vanBoom, J. H.; Altona, C. *J. Mol. Biol.* **1996**, *255*, 702.
(10) Altona, C.; Pikkemaat, J. A.; Overmars, F. J. J. *Curr. Opin. Struct. Biol.* **1996**, *6*, 305.
(11) Holliday, R. *Genetical Research* **1964**, *5*, 282.
(12) Kallenbach, N. R.; Ma, R. I.; Seeman, N. C. *Nature* **1983**, *305*, 829.

SECTION II. THERMODYNAMIC STABILITY OF BIO-MOLECULAR SYSTEMS

Chapter 1: Thermodynamics of the DNA 3WAY Junction

This paper proposes and tests a new approach for atomistic predictions of the stability of DNA structures in solution. This is an important area since the 3D structure of DNA complexes in solutions play an essential role in many biological and chemical processes[2]. A great deal of progress has been made recently in classical molecular dynamics studies of biological molecules in solution, including our work on the PX and JX double crossover molecules[3-5]. However a potentially serious issue is whether such classical calculations properly account for the entropy of the DNA-water-ions system. In this paper we propose and validate the use of the Two-Phase Thermodynamics methods (2PT) to extract from the MD the vibrational density of state need for the proper quantum statistical mechanics prediction of the vibrational contributions to zero point energy, specific heat and entropy. Most important the 2PT approach includes the non-vibrational contributions to the free energy from diffusional processes. Our formulation requires only 20 picoseconds of MD to extract accurate entropies, making this a most practical procedure that could be done for any MD study on biological molecules in solution.

Our specific application here is to the DNA 3-way junction, a model for the Holliday junction that plays a role in DNA recombination and a model for the types of structures involved in the Seeman technology for self-organizing DNA into complex structures. Structures for this system are available only from NMR, but we are able to follow the system thermodynamics as it transition from the NMR structure to a different topology in Mg^{2+} salt, and to a different fold at an elevated temperature in Na^{+}. We provide the simple thermodynamic arguments to explain these observations.

DNA Three-way Junctions (TWJs) are members of a ubiquitous family of nucleic acid branch-points, with the more famous Holiday (4-way) junction[6-7] being the central intermediate in homologous recombination. These junctions are important in various processes for both DNA and RNA. In DNA they are essential intermediates in gene splicing and nucleic recombination[8-9]. Using electrophoretic and fluorescence approaches, the solution structure of the 4-way Holiday junction was determined[10] and shown to be the central intermediate in gene splicing and nucleic recombination: the mechanism that facilitates evolution. Three-way junctions occur frequently in RNA, playing roles in many cellular processes ranging from translation[11] to programmed frame shifting[12]. Immobilized branched DNA junctions have been used to create complex nano-scale structures[13-14] and molecular dynamics (MD) simulations have confirmed that these structures are more structurally rigid than regular B-DNA[4,15].

Branch migration[16] – the step-wise exchange of base-pairs between different arms of branched nucleic acids – was recently investigated using FRET techniques and shown to be highly non-uniform and governed by two types of sequence-dependent barriers[17]. Additionally, it was shown that in the absence of a multivalent cation, the holiday junction is a dynamic motif, existing in quasi-equilibrium between two structures: the open structure with the four arms directed towards the corner of a square and the X-stacked structure with arms coaxially stacked[18]. Although the timescale of these structural transitions can be micro- to milliseconds, a recent MD study at elevated temperatures mapped the potential energy surface[19], uncovering key intermediates along the reaction path.

Three-way junctions are commonly depicted as comprising three distinct arms A, B and C (figure 1a) with the arms stacked in 3 unique configurations:

a) **B/C stacking**: a *parallel* configuration with arms B and C forming a quasi-continuous helix and stacked on arm A

b) **A/B stacking**: a *perpendicular* configuration with the A/B helix stacked on arm C

c) **A/C stacking:** an anti-parallel configuration with helix A/C stacked on arm B.

Various studies have shown that RNA three-way junctions exist in one of these three conformations, depending on the salt environment [10,20-21]: A/B stacking is present in systems of low ionic strength in the absence of multivalent cations[4], A/C stacking is present in systems containing multivalent cations[15,16], whereas B/C stacking occurs in systems with low concentrations of multivalent cations and in their absence[21]. Recently, it was found that DNA TWJs are stabilized under molecular crowding conditions[22], where it was shown that this stability is influenced by the hydration level of the DNA.

We are interested in predicting the structures of DNA platforms for applications to nanotechnology, involving crossovers to form PX/JX[3,23] systems with multiple sequences coupled into origami. For such systems there is little in the way of detailed experimental structural and thermodynamic data to validate the accuracy of the theory. There are several high resolution RNA NMR structures of three-way junctions, and a high resolution DNA structure – TWJ1[24]. Since the conformational transition in TWJs does not rely on branch migration, we expect that it should occur on the nanosecond timescale; making it suitable for molecular dynamics (MD) investigations into the fundamental physics of the system at normal temperatures. Obtaining consistent results from the MD simulations of TWJ1 would serve to validate the computational protocol and the forcefield, providing an indication of the reliability for similar predictions on other systems.

We started with the high resolution NMR structure[24] for TWJ1 in 0.1M NaCl, in which the A/B quasi-helix was observed to form an acute angle of $53^{o} \pm 11^{o}$ with the anti-parallel arm C (figure 1a – c). The unpaired thymines at the junction site resemble the folding of hairpin loops. The A/B stacking preference is consistent with the predicted stacking preference observed in RNA three-way junctions in the absence of multivalent counter-ions[9,25-26]. The distance-geometry analysis of the NMR experiments led to 26 3D conformers (poses) of TWJ1 considered as compatible with the experimental NOE distances (PDB codeTWJ1 in the Protein DataBank[27]). We carried out MD and energy minimization calculations on all 26 NMR poses leading to one (conformer 1) we consider most representative of the ensemble (see table S1 of supplementary materials). We then immersed this structure into a box of explicit waters plus 37 Na^{+} counter-ions (to achieve charge neutrality) and (after an initial equilibration) carried out 100ns of MD simulation (denoted as NVT for fixed temperature, volume, and number of particles) to obtain the structure denoted as TWJ1_Na. Our structure leads to A/B stacking for TWJ1, as expected, but during the MD, we found a transition to a second distinct stacking mode with arm C:

- the anti-parallel conformer (denoted Na_aP) has an average acute stacking angle of 63.6^{o} and, corresponding to the TWJ1 structure
- the parallel conformer (denoted Na_P) with an average obtuse stacking angle of 142.6^{o}

Thus TWJ1_Na transitioned from the anti-parallel state to the parallel conformer after 37ns of MD simulation. The perpendicular “T-shaped” conformer was observed only briefly as the intermediate in the transition. The stacking angle of the Na_aP conformer is close to that

observed experimentally, and averaging over the 32ns lifetime of the conformer we find good agreement with the experimentally measured NOE distances[24] (figure S2 of the supplementary materials). At 300K, our MD simulations find the Na_P conformer to be the equilibrium structure, a result that seems inconsistent with the experimentally observed structure.

We extracted from the MD trajectories, the entropy and zero point energies of both conformers, as well as the T-shaped transition state (denoted Na_T). These results suggest that both conformers have similar free energies with a relatively low free energy barrier for transition between them. Indeed we find that lowering the temperature from 30^{o}C to 13^{o}C changes the equilibrium structure, showing that the structures have similar energies.

Having validated our computations for TWJ1 in sodium salt, we next investigated the effect of the multi-valent cation Mg^{2+} on the structure. Using the same starting structure as TWJ1_Na, we replaced 34 of the 35 Na^{+} ions with 17 Mg^{2+} ions to obtain the structure denoted MG_A and carried out 100 ns of MD. Analysis of the stacking conformation of MG_A during dynamics revealed a single conformer with a B/C dihedral angle of $31^{\underline{o}}$. This structure, obtained after only 45ns of MD, has A/C stacking, which is consistent with experimentally observed[19-20] RNA TWJs in the presence of divalent cations.

To learn what factors control these structural preferences between Na^{+} and $Mg2^{+}$ ions, we examined the ion-phosphate radial distribution function (RDF) (figure 7b). This shows a large percentage (75%) of the divalent cations condensed on the DNA backbone, whereas the majority (90%) of the Na^{+} ions prefer to be solvated by an shell of six waters located at least one water shell away from the DNA. This ion condensation with Mg^{2+} effects a partially shielding of the DNA backbone self repulsion, leading to a structure for the MG_A more

compact by 10 - 15% than for either conformer of TWJ1_Na (as measured by the radius of gyration – figure 6a). Consistent with this, MG_A has a solvent accessible surface area (SASA) (figure 6b) 5% smaller than Na_aP, releasing 21 additional water molecules (5% of first layer) to the solvent, increasing the entropy of the system. The net result is that the free energy of the entire system is decreased, as will be discussed below.

The above calculations were all performed at 300K and a pressure of 1 bar. To examine the temperature dependence of TWJ1_Na folding dynamics, we evaluated the system thermodynamics at 285K and 315K. We find that the relative stability of the Na_P conformer and Na_aP are reversed at the lower temperature, with the free energy of the Na_aP conformer more stable by 16 kcal/mol rather than 23 kcal/mol less stable (as calculated at 300K). At 315K we find that the Na_P conformer is further stabilized, being 44kcal/mol more stable than the Na_aP. This change in stability is dominated by the change in entropy, particularly the entropy of the water molecules in the first solvation shell. Although Na_P is observed experimentally to be stable at 300K, our calculations indicate that that the Na_aP will be stabilized by increasing the temperature by ~20C, which could be tested with NMR experiments as a function of temperature. To our knowledge, this is the first complete, computational mapping of the free energy surface of DNA molecules evaluated self consistently while the molecule undergoes a structural transition.

Finally, we tested how well we could have predicted the structure of TWJ1 *without* using the NMR structural data to initialize the calculations. Here we used the NAMOT2[28] DNA building program to construct the starting configuration for TWJ1 as an A/B stacked structure with a stacking angle of 90^{o} to arm C (rather than 53^{o} from the NMR structure), using standard

B-DNA geometry (denoted MG_B). As with the MG_A structure, we solvated the DNA in a pre-equilibrated box of TIP3 water molecules and neutralized the system with Mg^{2+} counter ions. We find that MG_B converges to within 2Å of the equilibrium MG_A structure after 30 ns, and then remains in this structure for the remainder of the 20 ns MD simulation. This convergence to the previous calculated MG_A structure validates that our MD protocol is capable of reproducing the global minimum of these DNA structures on the tens of ns timescale at room temperature.

The remainder of the paper is organized as follows: section II describes the details of the MD simulations and various trajectory analyses. Section III, presents the results from our simulations and compares them with the available experimental data. Finally, section IV summarizes the results.

I. Theoretical Section

a) The initial structures

The NOE values from the NMR experiments[24] determine ranges of distances between various protons, that are combined with distance-geometry methods to generate an ensemble of 26 NMR conformers that together were considered to provide a reasonable fit to the NMR distance constraints [Protein Databank[27] (PDB) code TWJ1]. This is necessary because the distances sampled by NMR are averaged dynamically so that no single 3D structure can describe the ensemble of structures sampled during the dynamics.

However, for MD studies we need to select one of the 26 NMR conformers as the initial structure. Then we assume that over the time scale of the MD, the trajectory would sample the various configurations sampled in the experiment. We would like to validate our structures by

comparing the predicted NOE pair wise distances over our simulation trajectory (see table 6) to experiment. However the NMR paper did not publish the actual NOE distances their structures were fitted against. Thus for a figure of merit, we assumed that the average values from the 26 reported provide a good measure of the data.

We selected conformer 1 as the most representative of the ensemble: because it had

- the lowest average coordinate root mean square deviation (CRMS) deviation to the other 25 conformers (1.57Å),
- the second lowest CRMS to the starting structure before minimization (0.065 Å)
- an acute arm A – arm C stacking angle (ψ) of 53^{o}, which is closest to the experimentally reported average of $53^{o} \pm 11^{o}$ (see table S1 of supplementary material).

The minimized NMR Conformer 1 was used as the starting point for both the TWJ1_Na and MG_A simulations. For MG_B, each of the three arms were created independently using standard B-DNA geometry by the Namot2[28] nucleic acid builder program and arranged to have A/B stacking with a stacking angle of 90^{o} to arm C (figure 4e).

Each of the TWJ1 starting structures was inserted in a orthorhombic water-box (with initial cell dimensions of 70Å x 66Å x 74Å along the x, y, and z axes, respectively) of pre-equilibrated water molecules, described using the TIP3[29] FF with the LEAP module of AMBER8[30]. This procedure eliminated waters within 1Å of the DNA structure.

To achieve charge neutrality (within the periodic box), we inserted counter ions. Since TWJ1 has a net charge of -35, we replaced 35 water molecules with either 35 Na^{+} for TWJ1_Na or 17 Mg^{2+} ions and 1 Na^{+} ion for MG_A and MG_B respectively. In each case we used the

APBS[31-32] Poisson Boltzmann implicit solvation module to calculate the electrostatic potential map of the starting structure. Then for each type of counter ion, we neutralized the system by placing the counter-ions one water layer away from the DNA at the location with the highest electrostatic potential (in each case replacing the overlapping water molecule). In all three cases, we adjusted the system to have 8409 water molecules.

b) Force Field and MD Simulations Protocol

The molecular dynamics (MD) simulations were performed using the SANDER MD code[30] with the all atom 1999[33] force field with the PARMSBC0 modifications, which has been shown to give good structural and dynamical properties of nucleic acids on timescales as long as 2μs[34]. The short-range non-bonded interactions (van der Waals and Coulombic) were computed with a real space cutoff of 10Å. The long range electrostatics were evaluated with the particle mesh ewald (PME) method[35-36] with an interpolation order of 4 and an tolerance of 10^{-4}. The MD was performed using integration time step of 2fs and the neighbor list rebuilt every 10 steps.

For each of the three simulations described below, we first minimized the total energy of the system as described in our previous studies[3-5], followed by slow heating of the system from 0K to 300K. This was followed by 5 ns MD equilibration at constant pressure (1 atmosphere) and temperature (NPT)[37-40] using an Anderson thermostat with a temperature-coupling constant of 100.0 femtosecond, and an Berendsen barostat with a 2.0 ps damping constant. This NPT MD allowed the periodic cell to equilibrate at an average pressure of 1 atm. (these calculations kept the cell angles at 90º, orthorhombic). We found that equilibration in explicit water took nearly 5 ns to equilibrate, which is consistent with other studies[41]. From the

last 500 ps of MD we calculated a snapshot of the system whose volume is closest to the average volume of the cell during the MD. The MD was continued from this "best volume" structure for an additional 100 ns of MD using the constant volume, constant temperature (NVT) ensemble (only 50 ns in the case of MG_B). The trajectory snapshots (atomic positions and corresponding velocities) were saved every 10 ps for structural analysis.

c) Coordinate and Stacking Angle analysis

From the NMR, it was estimated that for the average stacking angle of the quasi-helix formed by the A/B arms on the helix formed by arm C is $53^{\underline{o}} \pm 11^{\underline{o}}$, determined by the 12 NOEs at the branch region. A similar analysis can be performed for each snapshot of TWJ1 MD: the dihedral (stacking) angle between these two intersecting planes is equal to the dot product of their two normal unit vectors (figure 1a):

$$\emptyset_{12} = cos^{-1}(\overrightarrow{\boldsymbol{n_1}} . \overrightarrow{\boldsymbol{n_2}}) = cos^{-1}\left(\frac{a_1a_2 + b_1b_2 + c_1c_2}{\sqrt{a_1^2 + b_1^2 + c_1^2}\sqrt{a_2^2 + b_2^2 + c_2^2}}\right) \quad (1)$$

for normal vectors $\overrightarrow{n_i} = a_i x + b_i y + c_i z$

The center-of-mass (COM) of the base-pairs comprising each arm was calculated for each snapshot of the MD simulation.

- Arm_A was taken to be bases 1(5'C) – 7(C) and 30 (G) – 37(3'G),
- Arm_B: bases 8(C) – 11(C) and 14(G) – 17(G) and
- Arm_C: 18(C) – 21(C) and 24(G) – 27(G)

as depicted in figure 1b. Considering each unique arm separately, a least-squares plane[42] through the COM of the base-pairs was then obtained, and the stacking between each pair of arms calculated (Fig 1.a):

- ϕ is the angle between arms C and A

- ψ is the angle between arms C and B and
- θ is the angle between arms A and B.

The average angle and relevant range of angles for a particular conformer was obtained by performing a non-linear fit of the distribution to a Gaussian function (figure 2).

Combined with the above mentioned arm stacking distributions, the stacking preference of TWJ1 (table 1) can then be obtained by the cross correlation function of the three unique dihedral stacking angles. The time dependent dihedral angle cross-correlation function $C_{l,m/n}(t)$ is defined as the cross correlation of angle *m* with angle *n* during dynamics:

$$C_{l,m/n}(t) = \langle P_l(t) \cdot P_l(0) \rangle = \frac{1}{t_{max}} \int P_l(t + t') \cdot P_l(t') dt' \quad \text{(2a)}$$

where $P_l(x)$ is the l^{th} order Legendre polynomial of objective function *x*. In this analysis, we use the 1st order polynomial:

$$P_l(x) = P_1(x) = x \quad \textit{(2b)}$$

Given a cross correlation functions $C_{m/n}(t)$, the characteristic cross correlation time $t_{m/n}$ can be calculated by

$$t_{m/n} = \int_0^\infty C_{m/n}(t) dt \quad \text{(2c)}$$

The cross correlation functions for each of the three systems are shown in figure 3 and the cross correlation times in table 2.

d) Structural Analysis of TWJ1 conformers

The criteria for selecting snapshots from the MD that are representative of a particular conformer are as follows: the trajectory was scanned for snapshots with all stacking angles within 1 standard deviation of the three average angles determined above. The “equilibrated” MD structures for each simulation was taken as the snapshot with stacking angles closest to the

average(s) of these distributions (figure 4), and the CRMS of TWJ1 along the entire trajectory was calculated as using the equilibrated MD structure as a reference. Of particular interest was the CRMS of TWJ1_Na along the 100 ns MD, using both the average Na_aP and Na_P structures the references (figure 5a).

Table 3 presents a comparison of the CRMS of the following five systems: the NMR ensemble, Na_P, Na_aP, MG1 and MG2. Here, the 10 best structures for each conformer were selected based on the following rules:

1) the structure has a stacking angle within half a standard deviation of the average for the set

2) each structure does not occur within 1ns of another in the set during MD and

3) the 10 lowest energy structures from the first two rules are selected.

These rules ensure that each snapshot was sufficiently independent, yet representative of the TWJ1 conformer.

TWJ1 structures can be thought of as being partitioned into two independent helices: co-axially stacked arms (helix 1) linked through the branch site to the other arm (helix 2). We calculated the helical twist, rise and roll parameters of each helix using the CURVES 5.1[43-45] software package (tables 4 and S2 of supplementary materials).

The structural size and rigidity of the various TWJ1 structures was estimated in terms of the mass-weighted root mean squared distance of all the atoms, relative to their COM (R_g) during dynamics. This analysis was done for both the TWJ1_Na and MG_A simulations (figure 6a). Additionally, a measure of the TWJ1 compactness is obtained by calculating the Solvent Accessible Surface Area, (SASA, Lee-Richards[46]). Here, a solvent molecule of probe radius 1.4Å

(representing a water molecule) is "rolled" around the solute (TWJ1) surface[47], using Van der Waals radii from the AMBER03[33]. The results of this calculation are shown in figure 6b.

Finally, the DNA-water and the DNA-ion radial distribution functions (RDFs) were calculated (figure 7), along with the ion-water RDF (figure S1 of supplementary materials). We considered the DNA surface to be represented by the PO_4^- groups and the water molecule by the position of the oxygen atom. The coordination number $\langle N_{ij}(r)\rangle$ represents the numbers of atoms of type *i* within a distance r of atoms of type j (and vice versa), and is related to the pair distribution function g_{ij} by:

$$\langle N_{ij}(r)\rangle = \rho_{ij}\int_0^r g(r)4\pi r^2 dr \qquad (3)$$

Table 5 presents the results of the RDF, SASA, and coordination number calculations for TWJ1_Na and MG_A. A cutoff of 10Å was used in calculating the RDFs. We designate the solvation shells as the regions on the RDF where the slope vanishes (valley to valley) and the coordination numbers are computed with respect to the first solvation shell only.

e) Comparison of NOE distances with experiment

The percent of the experimentally observed NOEs that are satisfied (within 0.5Å of the experimentally measured NOEs) as a function of simulation time was computed for both Na_aP and Na_P of TWJ1_Na (figure S2 of the supplementary materials). Using the experimental assignments[24], we calculated the NOE distances for the experimental NMR ensemble (they were not explicitly presented in the paper), and compared each of them to each snapshot of Na_aP, Na_P and MG_A during dynamics (table S3 of supplementary materials).

The anti-parallel Na_aP structures from our MD simulation satisfy roughly the experimentally observed NOEs; accounting on an average 55% of the experimentally observed

NOE distances (table 6). In a similar analysis 94% of the NOEs are satisfied by the minimized NMR structures. On the other hand, the parallel conformer Na_P, which is the equilibrium structure of TWJ1_Na at our simulation temperature of 300K, has a correlation to the NMR NOEs that is only 37% on average since it's A/C stacking angle is exactly opposite. A similar analysis of MG_A reveals that only 48% of the experimentally NOE's are accounted for by this MD structure.

f) Dynamical Properties of water molecules on the DNA surface

The dynamical properties of water molecules on the surface of biological molecules may play an important role in recognition and thermodynamics. However, the inhomogeneity of the surface makes it difficult to characterize experimentally. Nevertheless, techniques such as dielectric relaxation have been used to probe some aspects of water molecules on proteins[48-51] and in colloidal systems[52-53]. To compare with such data, we partition our analysis of the TWJ1 system into several groups (figure 8):

- DNA
- Ions
- Water molecules in the 1st solvation shell
- Water in the 2nd solvation shell
- Bulk water molecules

The self diffusion constant (*D*) for each of the five groups listed above was calculated during dynamics by computing the slope of the mean squared displacement according to Einstein's relation:

$$\langle r^2 \rangle = 6Dt \qquad (4)$$

In the case of the water molecules, the assignment to a particular shell was based on the DNA-water RDFs (figure 7a). A molecule was only taken as residing in a particular shell if it is consistently measured to be within range for at least 1 ps (the experimental bulk diffusion constant of water is 0.22Å^2/ps). This helps remove much of the noise in the data associated with transient molecules. The self diffusion constants are presented in table S4 of the supplementary materials.

g) Free Energy analysis – the 2PT-FE-MD method

The 2PT-FE-MD method uses the following protocol to obtain free energies from the MD. After equilibrating the system for 5 ns we use standard MD, with an integration time step of 1 fs, for an additional 20 ps of MD in which the velocities and structures are save every 4 fs (must be shorter than the fastest vibrational levels which have periods of ~10 fs for a 3000 cm-1 vibration

The components of the enthalpy per water molecule in the various hydration shells are presented in table 8. The water hydration shell entropy at 300K results is presented in table 8.

We now describe the computational procedure followed to obtain the thermodynamics of TWJ1 systems using the 2PT method. For Na_aP, Na_P, and MG_A, the input coordinates and velocities of the 10 best structures were used as a starting point for the entropy calculations. An additional ensemble, denoted Na_T, was considered as occurring during TWJ1_Na dynamics midway along the transition path from Na_aP to Na_P (37 – 42 ns). A set of 10 structures was taken 100ps apart, from 39 – 40 ns and is meant to represent the metastable intermediate in the transition. All four sets of structures were then run for 10ps of NVT dynamics at a temperature of 300K for temperature equilibration, with the SHAKE[54] constraints and a

timestep of 1fs. This was followed by a further 20ps NVT run, with the atomic coordinates and velocities saved every 4fs. Lin et al[55] showed that that a 20ps trajectory is sufficient to obtain an accurate entropy.

The results of the thermodynamics analysis are presented in tables 8 and 9. The translational and rotation components of the entropy for the different waters in Na_aP, Na_P and MG_A structure are shown in figure 9. A comparison of the system thermodynamics of the four ensemble structures under investigation at the three different temperatures is presented in table 9. The major findings are:

- Each water molecule in the first solvation shell has lower entropy (more constrained) by $T\Delta S$ ~ 0.4 kcal/mol than the bulk-like water molecules in the system, with the rotational and translational entropy components both decreasing equally. This is due to the smaller fluctuations in the orientations and positions due to the presence of the DNA. The per molecule enthalpic contributions decrease as the distance from the DNA surface increases.
- We see an anomalous *increase* in the entropy of the shell 2 water molecules by $T\Delta S$ = 1.9 kcal/mol and 1.4 kcal/mol compared to the shell 1 and bulk water molecules respectively. This results in the 2nd hydration shell water molecules having a higher per atom free energy than the 1st hydration shell (by 0.5 kcal/mol) and the bulk water (by 0.3 kcal/mol).
- At 300K, Na_aP has higher enthalpy (by 85 kcal/mol) than Na_P, but a lower overall entropy (by 108 kcal/mol) and a lower relative free energy of 23 kcal/mol. However, at 285K, Na_P is only 76 kcal/mol more favored entropically, while being disfavored

enthalpically by 92 kcal/mol , leading to Na_aP having a 16 kcal/mol more favorable free energy than to Na_P.

- The Na_P -> Na_aP barrier (as measured from the free energy of the Na_T conformer) is 54.5 kcal/mol at 300K and decreases to 9 kcal/mol at 285K. The reverse Na_aP -> Na_P process has a barrier of 31.5 kcal/mol at 300K and 26.0 kcal/mol at 285K.
- MG_A is has a much more favorable enthalpy than either Na_P (by 4860 kcal/mol) or Na_aP (by 4775 kcal/mol) at 300K, but less favorable entropy (by $T\Delta S$ = 266 and 158 kcal/mol respectively) and an overall lower free energy (by 4594 and 4617 kcal/mol) respectively.

Summarizing these findings, we find that the Na_aP structure (which is closer to the experimental NMR Conformer 1) is favored at 285K, while the Na_P conformer is favored at higher temperatures. We note that the transition temperature from our MD simulations would be forcefield and thermostat dependent, and thus may not correspond exactly to the real temperature. The differentials and general trends in the thermodynamics would therefore be the more useful measure.

To this end, we find a significant 54 kcal/mol barrier for the Na_P -> Na_aP transition, suggesting that such a transition is not likely at our simulation temperature of 300K. There is a lower barrier for both the forwards and reverse processes at 285K, suggesting that while the NMR-like conformer would be dominant at this temperature, a significant percentage of the population would be Na_P like. Thus although the force field is slightly biased toward Na_P, the MD indicates that the TWJ1 system will make a transition from the Na_aP (observed experimentally at 300K) to the Na_P structure at higher temperatures.

Finally, for MG_A the negatively charged counterions effectively shield the repulsive DNA backbone interactions leading to a dramatic in decrease in enthalpy, making the otherwise enthalpically unfavored A/C conformation favored. The divalent cation also affects the entropy of the water molecules in the 1^{st} and 2^{nd} hydration shells, lowering them by 0.5 cal/mol/K on average.

Thus this study shows that the stacking preference of TWJ1 is determined by two opposing forces: the DNA backbone repulsion which favors more open structures, and the entropy of the water molecules, particularly in the first solvation shell, which favors more compact DNA structures. We also find that the enthalpic and entropic components of the free energy are both temperature dependent, but with opposite slopes. From this, we see there are two regimes during which TWJ1_Na adopts conformers with opposite stacking angles, based on the subtle interplay between these two components. We note that it is only through rigorous considerations of entropic effects, that these two conformers can be rationalized. Applying the 2PT method to these molecular systems allows one to probe the free energy surface with unprecedented resolution.

II. Results and Discussion

a) Helical stacking angle and Atom Coordinate analysis

Table 1: Stacking angles of three TWJ1 MD simulations. The two distinct conformers of TWJ1_Na are reported: the anti-parallel Na_aP and the parallel Na_P, both of which exhibit A/B stacking.

Structure	Helical Stacking Angle (degrees)			Population(%)[a]	Best Snapshot[b]	Stacking Preference
	Φ	Ψ	Θ			
TWJ1_Na						
Na_aP	63.6 ± 12.3	66.0 ± 22.8	115.2 ± 8.8	11.8	206	A/B (anti-parallel)
Na_P	142.6 ± 10.9	44.5 ± 18.6	24.0 ± 9.7	45.6	7892	A/B (parallel)
MG_A	30.9 ± 13.0	80.9 ± 10.7	84.0 ± 13.9	41.3	4084	A/C
MG_B	36.5 ± 8.2	110.2 ± 12.1	84.3 ± 11.3	30.2	2038	A/C

[a]For each stacking angle, the snapshots with angles within 1 standard deviation of the average angle (as determined from a fit of the distribution to a Gaussian function) was taken as representative of the ensemble. Only conformers satisfying all three stacking angle ranges are retained. The average and standard deviation of each stacking angle was calculated from the final set of structures for each conformer.
[b]The best structure is the snapshot during MD with stacking angles closest to the average calculated (figure 8). Snapshots of the system were recorded every 10ps, we would multiply the snapshot number by 10 to obtain the corresponding simulation time.

The three components of the helical stacking angle (as a function of simulation time) for TWJ1_Na are shown in figure 2. We see that over the 100ns of MD, TWJ1_Na adopts two distinct conformations: Na_aP, with an acute ϕ stacking angle (arms A and C) of $63.6^{\underline{o}}$ and Na_P with an obtuse ϕ angle of $142.6^{\underline{o}}$ (table 1). The stacking of arms B and C (θ) is opposite ϕ angle – $115.2^{\underline{o}}$ for Na_aP and $24.0^{\underline{o}}$ for Na_P respectively, which is evidence that arms A and B move as a rigid helix, pivoting on arm C (A/B stacking).

One might expect the stacking angle of arms A and B (ψ) to be close to zero, however our calculation show that there is some kink in the helix: ψ = $44.5^{\underline{o}}$ and $66.0^{\underline{o}}$ for Na_aP and Na_P respectively. Additionally, the standard deviations in the ψ angles ($18.6^{\underline{o}}$ and $22.8^{\underline{o}}$ respectively) are about $10^{\underline{o}}$ larger than the other two stacking angles, underscoring that the A/B helix exhibits large fluctuations during dynamics. Finally, the acute ϕ stacking angle of Na_aP is

within range of the experimentally determined stacking angle ($53^{o} \pm 11^{o}$); the starting NMR structure does indeed exist in the anti-parallel A/B stacked conformation.

Unlike the unimodal Gaussian-like distributions of MG_A/MG_B, the distribution functions for TWJ1_Na more varied, displaying bimodal (in the case of the ϕ dihedral angle) and even trimodal (θ dihedral angle) distributions.

As shown in table 2, the ϕ/θ cross correlation time (as determined from equation 2) for TWJ1_Na is +7.0, as opposed to the 2 other cross correlation times which have negative values, indicating no correlation (-3.2 for ϕ/ψ and -5.3 for ψ/θ). These cross correlation functions (figure 3) give further evidence that in TWJ1_Na, arms A and B can be thought of as forming a helix, separate from arm C (the A/B stacking previously observed). Compare this to the cross correlation functions for MG_A: here, ϕ/ψ has the only positive correlation time of 11.9 ns. We conclude that unlike in the case of TWJ1_Na, MG_A can be thought of as a helix with arms A and C and another helix with arm B, exhibiting A/C stacking. The stacking preferences of TWJ1 in different salt conditions are consistent with experimental observations in RNA systems, and the agreement to our simulations is a major result of this paper.

Scanning the trajectory and selecting snapshots within 1 standard deviation of the

Table 2: Characteristic stacking-angles cross-correlation times (ns) during dynamics

Table 3: Matrix CRMS comparison of average structures of all three TWJ1 simulations. Note the Na_aP has the lowest CRMS to the starting NMR structure, which is to be expected as it has a similar stacking angle and the same stacking preference as the NMR structure.

	NMR	Na_aP	Na_P	MG_A	MG_B
NMR	1.42± 0.72	3.57 ± 0.16	10.52 ± 0.20	6.48 ± 0.20	6.52 ± 0.74
Na_aP	3.40 ± 0.43	1.50±0.63	11.43 ± 0.22	5.87 ± 0.21	7.28 ± 0.57
Na_P	10.40 ± 0.41	11.22 ± 0.44	1.22±0.47	11.98 ± 0.12	11.92 ± 0.58
MG_A	6.75 ± 0.50	6.42 ± 0.35	12.26 ± 0.20	1.15±0.42	2.61 ± 0.50
MG_B	6.39 ± 0.64	7.59 ± 0.61	12.15 ± 0.26	2.47 ± 0.33	1.42±0.71

average stacking angles, we can extract times in the trajectory where each of the conformers of TWJ1_Na appears. We find that Na_aP conformer is dominant from 12 – 35 ns and Na_P conformer is dominant from 42 – 100 ns. There is a transition from Na_aP to Na_P from 37 to 42 ns. Population analysis reveals that TWJ1_Na spends 11.8% of the time as Na_aP before transitioning to Na_P Na_P (table 1). This observation is reinforced by the CRMS calculations using both the equilibrated Na_aP and Na_P structures as references (figure 5a). Here we see that TWJ1 make transitions between both conformers: starting with Na_aP, there is a conversion to Na_P after between 37 and 42 ns of dynamics. We did not observe a back conversion to the Na_aP structure in the timescale of our simulation (100ns). Also important to note is that Na_aP is structurally similar to the anti-parallel starting NMR structure (average 3.40 – 3.57Å CRMS) whereas the parallel Na_P conformer has a large CRMS (10.40 – 10.52Å) to the starting NMR structure (table 3).

For MG_A, we see an initial transition from the A/B stacked NMR starting structure to the A/C stacked MG_A structure (around 40 ns), after which there is rapid convergence to the MG_A solution structure (figure 5b). The CRMS of MG_A after 50ns of dynamics is 1.89Å. The solutions structure of MG_A diverges from the starting NMR structure (average 6.5 – 6.75Å CRMS after 40ns) and is different from either Na_aP (5.9 – 6.4Å) or Na_P (12.0 – 12.26Å) (figure 5c). This can be compared to the MG_B simulation. Here, the solution structure of MG_B (figure 4e) is observed after 30 ns of dynamics, with a calculated CRMS of 6.7Å to the starting MG_B structure, and converges to within 2.5Å CRMS of the MG_A solution structure (figure 5d). The comparison of the CRMS of the top 10 structures representing each TWJ1 conformer is given in table 3. Here, we see that these structures have low fluctuations within each set (along the

diagonals), the CRMS between sets varies from low (for the structurally similar Na_aP and NMR structures: 3.4Å) to vast (Na_P and MG_A: 12.26Å).

b) Structural comparison of the two conformers of TWJ1_Na

The helical parameters were analyzed for each of the two helical domains for Na_aP and Na_P. Both conformers show significant deviations from a standard BDNA geometry, although they were closer to the measured values of the NMR ensemble (table 4). The twist angles for the individual helices were approximately 26^{o}, about 10^{o} smaller than that of standard BDNA. In comparison the 26 NMR structures lead to 32^{o} which is only 4^{o} smaller than experiment.

Table 4: Comparison of helical parameters of the NMR ensemble, MG_A and TWJ1_Na during dynamics. Both the NMR ensemble and TWJ1_Na exhibits A/B stacking: helix 1 is defined as arms A and B, while helix 2 is defined to be arm C. MG_A exhibits A/C stacking: helix 1 is defined to be arms A and C while helix 2 is arm B.

	Helix 1		**Helix 2**	
	Rise (Å)	**Twist (deg)**	**Rise (Å)**	**Twist (deg)**
B-DNA	3.4	36		
Exp/NMR	3.7±0.3	33.8±4.5	4.4±0.7	30.0±6.3
Na_P	3.7±0.4	27.2±6.5	3.3±0.5	26.6±9.8
Na_aP	3.5±0.6	25.1±7.4	3.3±0.4	26.6±10.9
MG_A	3.3±0.4	30.4±2.1	3.0±0.5	27.9±8.1

We find that this difference is dominated by two factors: unraveling of the terminal base-pair of helix 1 (table S2 of supplementary materials) and the previously mentioned fluctuations in unpaired thymines capping helix 2. The twist angle for terminal bases were 14.2^{o} and 12.5^{o} for Na_P and Na_aP respectively, contrasted with the 36.9^{o} for the NMR ensemble. Ignoring the terminal base value, the average helical twist of the TWJ1_Na structures becomes 31.5^{o}.The unraveling of the terminal base-pair in a DNA double helix occurs commonly in MD simulations, because the terminal base has only half the normal favorable interactions. A similar trend is observed for helix 2 (terminal twist angles of 9^{o} in Na_P and 8^{o} Na_aP), however

in this case the helices are capped with unpaired thymines. These thymines are much more flexible than the rest of the DNA during the MD simulation, since they are not constrained by the hydrogen bonding network of the other bases so that twist angle 25^{o} less than standard values may be realistic.

In order to quantify the range of motions of the two sets of unpaired thymines that cap helix 2 (T12/T13 and T22/T23), as well as the thymines at the branch site (T28 and T29), we performed two different analyses. For each set of unpaired thymines, dihedral angle (Φ) between the COM of the thymine-phosphate-phosphate-thymine was plotted against the dihedral angle (Ψ) between the least squares plane through the COM of each thymine (figure S3 of supplementary materials). The Φ angle measures the relative translational position of the COM of one thymine to the other. Here $\Phi = 90^{o}$ indicates that one is directly above the other. In contrast, the Ψ angle is a measure of the relative rotational orientations about the COM. Here $\Psi = 0^{o}$ indicates they are parallel while $\Psi = 90^{o}$ indicates that they are perpendicular.

For thymines 12 & 13, we see that they sample all of the possible configurational space, with near uniform sampling from 45^{o} to 135^{o}. Thus the thymines 12 & 13 can be thought of as freely rotating about their COM. This is qualitatively similar to the distribution of the other end capping thymines 22 & 23, although thymines 22 & 23 are more sample less of their configurational space and preferentially adopts a configuration in which their planes are perpendicular.

The sampling profile of the thymines (28 and 29) at the branched region (Fig S3c of supplementary materials) is different from the thymines capping the ends. Here, we see two distinct domains, as outlined by the red circles. In each of these two domains, the thymines

rotate freely with respect to each other while simultaneously rotating about their COM. We interpret this to mean that the thymines at the branched region move in a correlated fashion, attempting to maximize the favorable stacking interactions with each other. This is consistent with experimental[24] observations, where these thymines have been determined to exist as a type 1 hairpin fold. We should note that the NOEs associated with these thymines and arm B were used to determine the experimental stacking of helix 1 on helix 2. Our observation of two modes in the orientations of these thymines relative to each other is further evidence of the existence of the Na_aP and Na_P conformers.

The helical rise in helix 1 for both conformers (3.6Å for helix 1 and 3.3Å for helix 2) are close to, though larger than, the experimentally accepted value of 3.4Å for B-DNA. We find that this increased rise in the conformers is due primarily to the flexible end of the DNA responsible for the under twisting above. Large fluctuations in the rise indicate a flexible helix. The terminal base-pair is so flexible that it occasionally (25% of the time) becomes completely unpaired during dynamics.

Compared to the NMR ensemble helical rise (3.7Å for helix 1 and 4.4Å for helix 2), we find significant difference in helix 2 that is more difficult to explain. The average rise for both conformers is more than 1Ǻ less than that extracted from the NMR analysis, meaning that TWJ1_Na has a significantly shorter helix 2 and then NMR ensembles. A possible explanation for this compression is that the unpaired, thymines capping the ends are attempting to form favorably van der Waals contacts by base stacking on the adjacent base-pairs during dynamics. This would lead to an overall reduced rise, as each base-pair in the helix attempts to optimize its base-stacking distance.

c) Structural Parameters of TWJ1 in Mg^{2+} (MG_A)

The various helical stacking angle distributions for MG_A from the last 50ns of the 100ns run is shown in figure 2. Here we see only the 1 conformer with $\Phi=30.9^{o} \pm 13.0^{o}$, $\Psi=80.9^{o} \pm 10.7^{o}$ and $\theta=84.0^{o} \pm 13.9^{o}$ for MG_A. This data, coupled with the arm-arm cross correlation functions indicates that MG_A exhibits A/C stacking, in contrast to the A/B stacking seen in TWJ1_Na. Indeed, the CRMS profile (figure 5b), we see MG_A diverging from the starting structure almost immediately (black) and converging to on average 1.15Å of the solution structure (red) after 47ns of MD. Based on these results, the structures from the 50 – 100 ns MD window were used when calculating the various properties for MG_A.

Na_P is not topologically similar to MG_A, with average CRMS deviations of 11.9Å. Na_aP is more structurally similar, with CRMS deviations of 5.9Å. These differences are primarily due to the structure being more compact in the presence of Mg^{2+} and the fact that MG_A has a different topological fold than either TWJ1_Na conformers. The subtleties of the TWJ1 structure in these different solvent environments underscores the fact that TWJ1 is a dynamic structure, whose structural properties are not be captured by a single RMSD.

As a measure of the size we have computed the radius of gyration (R_g) of the 3-way

---

Table 5: The RDF and coordination number analysis of TWJ1_Na and MG_A during dynamics. The solvation shell is taken as the minima – minima distance, and the coordination number as the normalized number of molecules within the first shell.

	SASA (Å^2)	Rgyr (Å)	DNA – water			DNA – ion			Ion – water	
			1st solv. shell (Å)	2nd solv. shell (Å)	# in 1st shell	1st solv. shell (Å)	2nd solv. shell (Å)	# in 1st shell	1st solv. shell (Å)	Coord.#
NMR	7246±48	18.1±0.2								
Na_aP	6901±98	17.5±0.3	4.63	6.84	441	3.76	6.75	3.0/35	3.33	6.27
Na_P	6747±64	16.3±0.5	4.60	6.85	425	3.73	6.79	4.9/35	3.30	6.21
MG_A	6588±65	14.9±0.2	4.59	6.82	420	3.60	6.33	13.0/17	2.46	5.06

junction during the MD simulation (figure 6a). We find that MG_A has on average 1.3Å and 2.5Å smaller R_g than Na_P and Na_aP respectively (table 5). TWJ1 is therefore more compact in the presence of Mg^{2+} compared to Na^+, an observation supported by the SASA measurements: MG_A has 160$Å^2$ and 320$Å^2$ less accessible surface than Na_P and Na_aP respectively (figure 6b). In addition to being more compact, the MG_A is less flexible, with fluctuations of 0.2Å in R_g compared to 0.5Å for Na_P and 0.3 for Na_aP.

An additional important difference between MG_A and the TWJ1_Na structures is the ion-DNA coordination (figure 7b), with the normalized occupancy of MG_A being an order of magnitude larger than TWJ1_Na. We find that, on average, 13 of the 17 Mg^{2+} ions are located within the first solvation shell of 2.5Å of the DNA for MG_A, whereas only 3 Na^+ ions are within the first solvation shell (3.3Å) for TWJ1_Na.

Additionally, we find a difference in the ion-water coordination numbers for the two ions: 5 for Mg^{2+} and approximately 6 for the Na^+. While the Na^+ coordination number is consistent with other reported results, the Mg^{2+} coordination number of 5 is less than the 6 waters in an octahedral arrangement reported by other authors[56-57], with the value for similar doubly charged cations in between 7 and 10[58]. An interpretation of this is that the Mg^{2+}ions partially de-solvates and bind to the DNA backbone, losing 1 water molecule from its solvation shell.

The structural parameters of MG_A provide further evidence that the DNA is more compact and rigid than either of the conformers extracted from the simulations in presence of Na^+. The average helical rise in helix 1 of MG_A is 3.3Å while that of helix 2 is 3.0Å, shorter by 2.1Å and 2.4Å respectively from that of Na_P and Na_aP. We interpret the decreased helical

rise of MG_A to be a consequence of the Mg^{2+} ions condensing on the DNA backbone, reducing the backbone repulsion.

Additionally, the twist angles for MG_A are 30.4^{o} for helix 1 and 27.9^{o} for helix 2, which is 4^{o} and 1.5^{o} larger for helix 1 and helix 2 of TWJ1_Na respectively. Combining the helical rise and twist results, we see that MG_A is a more tightly wound helix than TWJ1_Na. The magnitudes of the fluctuations in the helical parameters (2.1^{o} and 8.1^{o} for helices of MG_A versus 7.4^{o} and 10.9^{o} for TWJ1_Na) also indicate that the magnesium conformer is more structurally rigid than either of the sodium conformers, supporting our previous SASA and radius of gyration results. Studies on similar irregular BDNA systems[5] have shown that the structures in magnesium are more rigid than in sodium.

d) NOE analysis of conformers of TWJ1

Experimentally, the overall structural features of TWJ1 were obtained directly from 46 unambiguous NOEs, of which 12 were used to determine the stacking of helix 1 (arms A & B) on helix 2 (arm C) [24]. Only NOEs between bases C7, C8, G17, C18, G27, G28, G29, and G30 were presented. We calculated these 46 distances from the 26 experimental conformers[27]. From our measurements, 19 of the 46 NOEs assignments led to surprisingly large distances (> 5Å - pair "C8_H6 <-> G27_2H2" is 7.91Å and pair "C8_H6 G27_2H2" is 7.49Å – see table S3 of supplementary materials), which we assume must have arisen after final optimization of the structures after identifying

Table 6: Comparison of NOE distances of the NMR ensemble, Na_aP, Na_P and MG_A during dynamics. Here, < .. > represents the statistical average.

	NMR	Na_aP	Na_P	MG_A
<% NOEs within 0.5Å>	93	55	37	48
<deviation (Å)>	0.87	1.19	1.38	0.71
Max deviation (Å)	1.2	4.11	5.22	3.49
# structures	26	628	2002	2002

them. We optimized each of the 26 structures with 100 steps of steepest decent minimization to remove any biases of the force field used in the experimental optimization. We then recalculated the helical stacking angles for each. Though all the structures exhibit A/B stacking, there are large fluctuations in the stacking angles: $\Psi = 62.4 \pm 45.5^{o}$, $\Phi = 71.7 \pm 39.5^{o}$ and $\theta = 84.5 \pm 38.8^{o}$. Indeed, our analysis finds that 16% of the NMR structures would be classified as Na_P instead of Na_aP.

For Na_aP, we find that on average, 55% of the experimentally observed NOEs are measured to be within 0.5Å of the NMR distances (table 6). This may not seem a large percentage, since we calculate an average of 93% for the 26 NMR structures. It is important to note, however, that our MD ensembles contain no constraints to satisfy the NMR NOEs and are averaged over a more complete set of structure (2002 for Na_P/MG_A and 628 for Na_aP) than the relatively small set (26) of NMR structures, for which both distance and dihedral restraints were employed in their optimization.

The adequacy of a tolerance of 0.5Å is also a source of debate: this was the reported confidence level of the measured NOE distances; however we calculated deviations as large as 1.3Å in the NMR structures. If the tolerance is relaxed to the measured average standard deviation measured from the NMR structures (0.87Å), we find that 63% of the Na_aP distances satisfy the experimental NOE distances.

Finally, from the analysis of the individual NOE distances (table S3 of supplementary materials), we see that of the 12 NOEs used to represent characterize the helical stacking angle, Na_aP satisfied 9 of the 12 distances. Indeed, the source of largest discrepancies (> 1Å) between the two sets of structures all involved distances between hydrogens residues C7 and

G28 and the unpaired thymines (T28 and T29) at the branch site. These unpaired thymines are expected to be flexible during dynamics so deviations from idealized structures are expected, since the forcefield used has been optimized to reproduce the properties of double stranded DNA.

A comparison between the NOE statistics of Na_aP, Na_P, MG_A and the NMR structure is presented in figure S2 of the supplementary materials.

e) Thermodynamic comparison of TWJ1_Na and MG_A

Table 7: Comparison of Enthalpy per water (kcal/mol) for Na_aP, Na_P, Na_T and MG_A conformers at 300K. These enthalpies have not been corrected from zero-point energy effects. The enthalpy decreases as a function of distance from the DNA surface, due to the favorable electrostatic interaction with DNA backbone.

	Na_aP			Na_P			MG_A			Na_T			Water box[a]		
	Total	**VDW**	**Coul**	**Total**	**VDW**	**Coul**	**Total**	**VDW**	**Coul**	**Total**	**VDW**	**Coul**	**Total**	**VDW**	**Coul**
shell1	-10.54	1.30	-11.84	-10.51	1.31	-11.82	-11.87	1.55	-13.42	-10.46	1.23	-11.69	-9.98	1.59	-11.58
shell2	-10.16	1.41	-11.57	-10.14	1.42	-11.55	-10.95	1.56	-12.51	-10.27	1.43	-11.70			
bulk	-9.90	1.46	-11.35	-9.89	1.46	-11.35	-9.92	1.47	-11.39	-9.89	1.46	-11.36			

[a]Water box values obtained from 5 1ns independent simulations of 800 tip3p molecules at 300K

From the DNA-water RDFs (figure 7a) we calculate the number of water molecules in the first solvation shell. Consistent with the SASA, we find that Na_aP has the larger number of water molecules in the first shell (441) than Na_P (425). A comparison of the enthalpy of the various water molecules (figure 8) is presented in table 7. Here we see that there is a systematic decrease in the enthalpy per molecule, going from shell 1 to bulk, across all the conformers. When looking at the individual energetic components (Coulomb and van der Waals) of the water molecules, it's evident that the columbic interactions are indeed decreasing, with shell 1 and shell 2 waters having 0.5kcal/mol and 0.2kcal/mol better columbic

interactions than the bulk. This effect we attribute to the large charge density (due to the PO_4^- groups) on the DNA surface, leading to a favorable point – dipole interaction. Also evident is an increase (1.5 kcal/mol) in the electrostatic interactions in the case of MG_A compared to the TWJ1_Na systems. This effect is the consequence of having the divalent cation condensed on the DNA backbone: the strong water – ion enthalpic interactions increases the average electrostatics of the first shell water molecules.

When considering the per-water entropies (as shown in table 8), we see an unexpected increase in the entropy of the waters in the second hydration shell, relative to the bulk. Analysis of the contributions to this entropy (figure 9a-f) reveals that the translational entropy of shell 2 water is increased and red shifted compared to the bulk-like waters. We postulate that this apparent anomaly can be explained by noting that the first hydration shell is between 0.5 (TWJ1_Na) to 1.0 (MG_A) cal/mol/K more constrained than the bulk-like waters. The second hydration layer can then be thought of as being frustrated layer, interfacing two incompatible water structures: the entropically favorable bulk water layer and the enthalpically favored 1^{st} hydration shell. This is analogous to that of water molecules at the ice-water interface. We believe that this is the first time such an analysis has been done on biological systems.

The increase in the rotational entropy as one proceeds away from the DNA surface can be rationalized by the dielectric effect: molecules will align their dipoles to maximize their electrostatic interaction with an external field (the DNA surface in our system). We would therefore expect the water molecules at the surface to be both rotationally and translations hindered. This has the effect of lowering the diffusion constant (table S4 of supplementary materials) of the first solvation shell waters, as well as lowering the translational and rotational

entropies. In the case of MG_A, the translational and rotational entropies of the water in the first hydration shell are reduced relative to the TWJ1_Na conformers (by 0.6 and 0.07 cal/mol/K respectively). Here we assume that the decreased mobility of the Mg^{2+} ions, condensed on the

Table 8: Components of the per water entropy (cal/mol/K) at 300K. We see an anomalous enhancement of the entropy of the second shell waters due to increased diffusivity of these molecules compared to the bulk-like waters. The experimental bulk entropy of water at 293K[1] is 16.73 cal/mol/K.

	Na_aP			Na_P			Na_T			MG_A			Water box[a]		
	S_{trans}	S_{rot}	S_{tot}	S_{trans}	S_{rot}	S_{tot}	S_{trans}	S_{rot}	S_{tot}	S_{trans}	S_{rot}	S_{tot}	S_{trans}	S_{rot}	S_{tot}
shell1	13.73	2.40	16.13	13.69	2.41	16.10	13.66	2.41	16.07	13.15	2.33	15.48	12.16	2.28	14.44
shell2	15.46	2.56	18.02	15.41	2.57	17.98	15.30	2.57	17.88	15.04	2.49	17.52			
bulk	13.90	2.60	16.49	13.93	2.61	16.54	13.88	2.60	16.48	13.91	2.60	16.51			

DNA backbone, relative to the Na^{+} ions in solvent introduces more order into the system.

Turning to the entire system thermodynamics, we find that at 300K, Na_aP is more favorable than Na_P by 85 kcal/mol. Included in this is a net increase in the DNA enthalpy of 13kcal/mol and an almost 175 kcal/mol increase in the shell 1 water enthalpy (table S5 of supplementary materials), due to the 20 additional water molecules in the first solvation shell. We know that water molecules in the first solvation shell have a more favorable enthalpic interaction compared to the bulk, hence this result is consistent with the per – molecule enthalpic analysis.

We find that Na_P has a larger entropy than Na_aP by 0.36 kcal/mol/K, which when considering a temperature of 300K translates into a 108 kcal/mol increase in the entropy. Again, the normalized water entropies are instructive: as noted before, the water molecules in the first solvation shell of the TWJ1_Na structures are less entropically favored than the bulk. Hence Na_aP pays a larger entropic penalty than Na_P, primarily based on the water molecules

in the first solvation shell, ultimately leading to Na_P being 23 kcal/mol more favorable than Na_aP at 300K.

Table 9: Free energy analysis of TWJ1 structures at 3 different temperatures. For TWJ1_Na, the Na_aP conformer is energetically favored at 285K, whereas the Na_P conformer becomes more favorable as the temperature is increased. The barrier for the Na_P <-> Na_aP transition is calculated using the Na_T conformer as the transition state. The MG_A conformer is the most favorable of all the TWJ1 structure, owing to a much more favorable enthalpic energy.

	[a]Helmholtz Free Energy(A) (kcal/mol)		[b]Enthalpy(U) (kcal/mol)		Entropy(S) (cal/mol/K)		[c]ΔA_{Na_aP} (kcal/mol)	ΔU_{Na_aP} (kcal/mol)	ΔS_{Na_aP} (cal/mol/K)
	Avg	±	Avg	±	Avg	±			
285K									
MG_A	-133820.47	44.61	-95582.65	53.44	134167.76	249.51	-4589.88	-4734.04	-505.82
Na_aP	-129230.58	46.76	-90848.61	57.28	134673.58	230.31	0.00	0.00	0.00
Na_P	-129214.42	36.27	-90756.58	33.72	134939.80	267.60	16.16	92.04	266.22
Na_T	-129204.56	34.69	-90846.24	43.33	134590.59	225.98	26.03	2.37	-82.99
300K									
MG_A	-136337.49	45.63	-94030.33	46.55	141023.89	166.20	-4617.11	-4774.95	-526.14
Na_aP	-131720.39	39.60	-89255.38	38.25	141550.03	302.63	0.00	0.00	0.00
Na_P	-131743.39	39.02	-89170.25	39.34	141910.46	202.41	-23.00	85.12	360.43
Na_T	-131688.94	42.53	-89247.57	49.76	141471.23	221.36	31.45	7.81	-78.80
315K									
MG_A	-138969.94	192.07	-92458.99	54.16	147653.82	198.47	-4601.84	-4741.49	-443.34
Na_aP	-134368.11	309.79	-87717.50	91.86	148097.16	311.16	0.00	0.00	0.00
Na_P	-134412.90	114.90	-87582.26	40.39	148668.70	232.49	-44.79	135.24	571.54
Na_T	-134338.32	191.45	-87650.49	47.35	148215.33	160.85	29.79	67.01	118.17

[a]A = U – TS

[b]U is the internal energy, corrected for all zero point energy effects

[c]All the reference energy are taken using the Na_aP conformer as the reference

At 285K, we find that the enthalpy is the dominant component in determining the relative free energy of Na_aP compared to Na_P. Here, Na_aP is 92 kcal/mol enthalpically favored, but only 75 kcal/mol disfavored entropically: the net result is that Na_aP has a 16 kcal/mol favorable free energy, relative to Na_P. Finally, at 315K, we find that Na_P is further stabilized wrt Na_aP (44 kcal/mol in free energy). We conclude therefore that increases in the

simulation temperature above 285K increases the influence of the entropy in determining the relative free energy of the TWJ1_Na conformers. The more entropically favored Na_P conformer becomes dominant at elevated temperatures while the NRM-like Na_aP conformer is favored at 285K and below. We propose that the NMR studies be performed at $\pm 15^{o}$ to confirm the existence of our Na_P TWJ1 conformer.

When considering the MG_A conformer, at 300K we find it has a more favorable enthalpy than either of the TWJ1_Na conformer (4860 and 4775 kcal/mol for Na_aP and Na_P respectively). Entropically, the more compact, less mobile DNA molecule (table S4 of supplementary materials) is lower in entropy by 158 and 266 kcal/mol respectively, leading MG_A to have a significantly larger Helmholtz free energy (> 4500 kcal/mol) than either of the TWJ1_Na conformers.

Finally, we calculated the free energy of the Na_T intermediate in order to ascertain the barrier for the Na_aP <-> Na_P transition. At 300K, we find that the Na_T intermediate is more enthalpically favorable than Na_P (77 kcal/mol) and slightly less enthalpically favorable than Na_aP (7.8 kcal/mol). Overall however, based on free energy, it is less favorable than either conformer (with a Na_P -> Na_aP barrier of 54.5 kcal/mol and 31.5 kcal/mol for the reverse process) due to decreased entropy (131.8 and 23.6 kcal/mol relative to Na_P and Na_aP respectively). These free energy barriers are reduced to 9 and 26 kcal/mol respectively at 285K, and increased to 74.6 and 29.8 kcal/mol at 315K. Na_T consistently has nearly the greatest number of water molecules in the first hydration shell (table S5 – S7 of supplementary materials), and a reduced DNA backbone repulsion due to a more open structure (calculated SASA 7054$Å^2$).

We have shown that the helical stacking preference of TWJ1 can be rationalized on the basis of the overall free energy of the system: for TWJ1_Na, a conformation with a 90^{o} helical stacking angle (Na_T) minimizes the backbone-backbone repulsion but increases the number of water molecules in the 1^{st} and 2^{nd} solvation shells, which in turn decreases the entropy. Conformers with a smaller helical stacking angle minimize the number of water molecules in the solvation shells, but increase the backbone repulsion. The situation is somewhat different in the presence of the divalent counter-ion Mg^{2+}: the magnesium ions condense along the backbone, thereby shielding the backbone repulsion. However due to the high charge, these ions bind more water molecules in the 1^{st} solvation shell and also increase the entropy of these water molecules, relative to the sodium systems, by provide a surface of alternating charges. Therefore MG_A adopts a conformation with the lowest helical stacking angle, as this conformation requires the least water molecules in the 1^{st} solvation shell, yet it is as well solvated as the much larger TWJ1_Na structures.

f) MD of alternate initial configuration of TWJ1 in Mg^{2+} (MG_B)

Both the TWJ1_Na and MG_A systems studied above started with the NMR (Conformer 1) structure. A natural question one might ask would be whether simulations could be used to make *a-priori* predictions concerning the structure and thermodynamics of these nucleic acid systems. If this were possible, it would make the simulations a far more powerful tool to examine systems prior to experiment.

We thus constructed an alternate conformation of MG_A, MG_B, using the Namot2[28] nucleic acid builder program, making no use of our simulations or the NMR experiment structures. The system was constructed using Mg^{2+} instead of Na^{+} ions, since MG_A shows one

dominant conformer at equilibrium and we wanted a comparison to a well defined state. The starting structure (figure 4e) has a 6.2Å heavy atom CRMS to the starting NMR structure of MG_A. We then carried out 50 ns of NVT dynamics on MG_B using the same procedure as described for the previous two TWJ1 simulations.

Figure 5d shows the CRMS during dynamics, using the best NMR structure and the equilibrated MG_A structure as reference. Here we see that MG_B convergences to within 1.96Å of the MG_A structure after 27 ns and diverge to 5.7Å of the NMR structure (after starting 2.0Å away). Using the last snapshot as the equilibrium structure (figure 4f), we see that MG_B converges to within 1.55Å of this equilibrated structure after 30ns. The top 10 best structures of MG_A and MG_B are within 2.47Å and 2.61Å average CRMS deviation from each other, respectively. For comparison the fluctuations of MG_A with itself are 1.15Å while that of MG_B with itself are 1.42Å (table 3).

Previously, it was found that the helical stacking angle was a better measure of structural similarity than CRMS, although it should be noted that a CRMS of < 2 Å suggests that the structures are topologically identical. Table 1 lists the helical stacking angles of the top 10 MG_B structures. We see that stacking angles are similar to that of MG_A: $\Phi = 36.5^{o} \pm 8.2^{o}$, $\Psi = 110.2^{o} \pm 12.1^{o}$, $\Theta = 84.3^{o} \pm 11.3^{o}$, compared to $30.9^{o} \pm 13.0^{o}$, $80.9^{o} \pm 10.7^{o}$ and $84.0^{o} \pm 13.9^{o}$ for MG_A respectively. Of note is that the magnitude of the fluctuations in for MG_B are also similar to that of MG_A, giving confidence that we have indeed found a converged structure.

III. Concluding Remarks

We demonstrate here the use of 2PT-FE-MD for application to a practical problem, the conformations of the DNA triplex, TWJ1. We developed a scheme for solvating the DNA systems

that leads to rapid equilibration of the ions on the 10ns timescale. We then used long term MD (100 ns) to investigate the equilibrated structures for our TWJ, starting from the best NMR structure. In the presence of the Na^+ ion, we found two distinct conformers: Na_aP and Na_P, which we calculate as being in thermodynamic equilibrium. Na_aP has an average helical stacking angle of $\Phi=63.6^o$ and we posit is representative of the deposited NMR ensemble. Analysis of the 26 NMR conformers shows that 6 had stacking angle within range of the Na_P conformer.

Population analysis reveals that at the simulation temperature of 300K, 1180 (12%) of the snapshots in the TWJ1_Na trajectory can be assigned conformer Na_aP and 4560 (46 %) assigned conformer Na_P, based on the three unique stacking angle. Although we did not choose structures representative of the different conformers based on CRMS, we observe that the structures chosen are indeed similar to each other, with average CRMS differences ranging from a high of 1.42Å (NMR structures) to a low of 1.15Å (MG_A structures).

Similarly, the comparison of the CRMS between sets of conformers following the stacking angle analysis. The topologically similar NMR and Na_aP structures are within 3.5Å CRMS of each other on average. Again, this would be an appropriate time to underscore the difference between these two structures. Na_aP is an ensemble of structures obtained from unconstrained MD, starting from the best representative NMR structure. The NMR ensemble is the energy minimized structures, obtained from constrained NMR refinement to fit the available NOE data. The equivalence of these sets of structures was by no means assured.

Energy comparisons between these systems are tenuous due to the large fluctuations arising from the water-water interactions, fluctuations that usually dwarf the small energetic

differences. Additionally, it has previously been computation intractable to obtain reliable entropic effects for such large systems. We show here that the 2PT statistical mechanics scheme [59] to extract the entropy and zero-point energy corrections directly from the dynamics provides a practical solution to this problem.

These results show that Na_aP has a lower enthalpy due to reduced electrostatic backbone repulsion, but Na_P have a lower entropic penalty since it has less water molecules near the surface of the DNA, where they are more constrained than in the bulk. Na_aP is more thermodynamically favored at 300K by 23 kcal/mol, but unfavored at 285K by 16 kcal/mol. This presents a well defined theoretical prediction for experimental validation, and the authors urge our experimental colleagues to perform such an investigation.

We also investigated the effects of replacing the Na^+ with the divalent Mg^{2+} cation on the TWJ1 structure. Most experiments on DNA nanostructures use MgCl rather than NaCl because it leads to more stable structures with fewer fluctuations. Indeed our previous simulations on PX and JX junctions [3,23] showed just such results. We observe dramatic differences for TWJ1 with Mg^{2+} ions, most prominent of which was the observation that TWJ1 adopt a fold that is consistent with A/C stacking in Mg^{2+}, while adopting A/B stacking in Na^+. In MG_A, the Mg^{2+} ions condensed along the DNA backbone, effectively screening the helical backbone repulsions. This leads to just one conformer observed for MG_A, with a Φ stacking angle of 31^o (cf 64^o for Na_aP of Na_aP and 143^o for Na_P). Structural analysis revealed that the helices of MG_A are closer to regular B-DNA geometries that either of the sodium conformers. We also found that the MG_A structurally more rigid, with a radius of gyration of 14.9±0.2Å

compared to 16.3±0.5Å, 17.5±0.3 and 18.1±0.2 for Na_P, Na_aP and the NMR ensembles respectively.

The magnesium ion has a pronounced effect when considering the thermodynamics. Unlike in the sodium simulations, where approximately 90% of the ions are in solution, 13 of the 17 magnesium ions are coordinated with the DNA surface. This increases the enthalpic interactions as discussed previously. It also affects the interactions of the water molecules in the first solvation shell, increasing the electrostatic energy of these atoms. This leads MG_A system that is the most compact, yet who overall free energy is lower that would be expected from the sodium simulations.

An important question is whether modest level MD simulations (50 nanosecond) can provide accurate thermodynamics while traversing the vast energy landscape sufficiently to find the equilibrium structure. To determine how well we could predict these structures without use of experimental data, we used the Namot program to constructed a starting structure with a Φ stacking angle of 90^{o} (as opposed to 53^{o} for the NMR structure) making no use of the NMR data. We found that after 27 ns of MD, this MG_B simulation had converged to within 1.5Å of the equilibrated MG_A structure, with an Φ stacking angle of 37^{o}. This indicates that theoretical structures can be used to estimate equilibrium geometries of these nanoscale structures and that a modest time scale of dynamics is capable of finding the global minima.

It was our aim in the paper to provide the first complete thermodynamic argument for the folding of a TWJ, using MD simulations and 2PT analysis. We find that for TWJ1_Na, there is a tradeoff between the entropy which favors the more compact Na_P structure, and the enthalpy which favors the more open Na_aP structure. We calculate that the temperature

determines which of these conformers is favored, and the relative populations of each conformer at equilibrium. The essential physics in these systems can be captured by observing the nature of the 1st shell water molecules.

IV. Acknowledgements

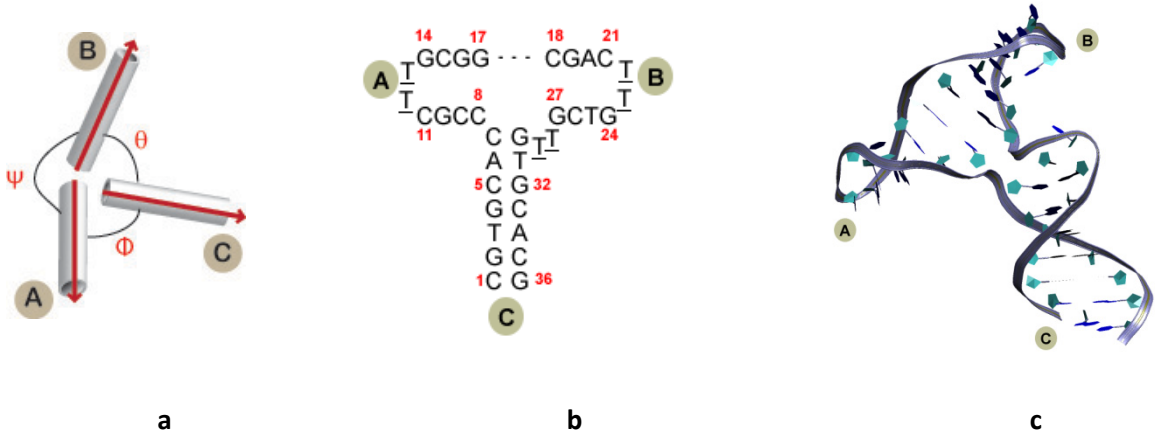
We thank Prof. Ned Seeman, Prof. Eric Winfree, Prof. Shang Tai Lin, and Dr. Mario Blanco for useful discussions. Also, we thank Gene Carter for helping adapt the Namot2 program to our needs. TAP thanks the DOE CSGF for a graduate fellowship. Partial support for this work was from the National Science Foundation, Grants CTS-0608889 and CTS- 0548774). The computational facilities were provided by ARO-DURIP and ONR-DURIP.

(1) Cox, J. D. W., D.D.; Medvedev, V.A. *CODATA Key Values for Thermodynamics*; Hemisphere Publishing Corp.: New York, 1984.
(2) Eimer, W.; Pecora, R. *The Journal of Chemical Physics* **1991**, *94*, 2324.
(3) Marcus, Y. *Chemical Reviews* **1988**, *88*, 1475.
(4) Kitts, P. A.; Nash, H. A. *Nature* **1987**, *329*, 346.
(5) Maiti, P. K.; Pascal, T. A.; Vaidehi, N.; Goddard, W. A. *NUCLEIC ACIDS RES* **2004**, *32*, 6047.
(6) Maiti, P. K.; Pascal, T. A.; Vaidehi, N.; Heo, J.; Goddard, W. A. *Biophys. J.* **2006**, *90*, 1463.
(7) Maiti, P. K. P., Tod A.; Vaidehi, Nagarajan; Goddard, William A. *J. Nanosci. Nanotechnol.* **2007**, *7*, 1712.
(8) Holliday, R. *Genetical Research* **1964**, *5*, 282.
(9) Kallenbach, N. R.; Ma, R. I.; Seeman, N. C. *Nature* **1983**, *305*, 829.
(10) Nunesduby, S. E.; Matsumoto, L.; Landy, A. *Cell* **1987**, *50*, 779.
(11) Altona, C.; Pikkemaat, J. A.; Overmars, F. J. J. *Curr. Opin. Struct. Biol.* **1996**, *6*, 305.
(12) Lilley, D. M. J.; Clegg, R. M. *Quarterly Reviews of Biophysics* **1993**, *26*, 131.
(13) Noller, H. F. *Annual Review of Biochemistry* **1991**, *60*, 191.
(14) Rettberg, C. C.; Prere, M. F.; Gesteland, R. F.; Atkins, J. F.; Fayet, O. *J. Mol. Biol.* **1999**, *286*, 1365.
(15) Seeman, N. C. *DNA and Cell Biology* **1991**, *10*, 475.
(16) Shen, Z. Y.; Yan, H.; Wang, T.; Seeman, N. C. *Journal of the American Chemical Society* **2004**, *126*, 1666.
(17) Maiti, P. K.; Pascal, T. A.; Vaidehi, N.; Goddard, W. A. *Nucleic Acids Research* **2004**, *32*, 6047.
(18) Warner, R. C.; Fishel, R. A.; Wheeler, F. C. *Cold Spring Harbor Symposia on Quantitative Biology* **1978**, *43*, 957.
(19) Duckett, D. R.; Murchie, A. I. H.; Diekmann, S.; Vonkitzing, E.; Kemper, B.; Lilley, D. M. J. *Cell* **1988**, *55*, 79.
(20) McKinney, S. A.; Declais, A. C.; Lilley, D. M. J.; Ha, T. *Nature Structural Biology* **2003**, *10*, 93.
(21) Yu, J.; Ha, T. J.; Schulten, K. *Nucleic Acids Res.* **2004**, *32*, 6683.
(22) Altona, C. *J. Mol. Biol.* **1996**, *263*, 568.
(23) Welch, J. B.; Walter, F.; Lilley, D. M. J. *J. Mol. Biol.* **1995**, *251*, 507.
(24) Muhuri, S.; Mimura, K.; Miyoshi, D.; Sugimoto, N. *Journal of the American Chemical Society* **2009**, *131*, 9268.
(25) Yan, H.; Zhang, X. P.; Shen, Z. Y.; Seeman, N. C. *Nature* **2002**, *415*, 62.
(26) van Buuren, B. N. M.; Overmars, F. J. J.; Ippel, J. H.; Altona, C.; Wijmenga, S. S. *J. Mol. Biol.* **2000**, *304*, 371.
(27) Ouporov, I. V.; Leontis, N. B. *Biophys. J.* **1995**, *68*, 266.
(28) Overmars, F. J. J.; Pikkemaat, J. A.; vandenElst, H.; vanBoom, J. H.; Altona, C. *J. Mol. Biol.* **1996**, *255*, 702.
(29) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535.
(30) Tung, C. S.; Carter, E. S. *CABIOS* **1994**, *10:*, 427.

(31) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.
(32) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668.
(33) Holst, M. J.; Saied, F. *J. Comput. Chem.* **1995**, *16*, 337.
(34) Holst, M.; Saied, F. *J. Comput. Chem.* **1993**, *14*, 105.
(35) Cheatham, T. E.; Cieplak, P.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845.
(36) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817.
(37) Cheatham, T. E.; Miller, J. L.; Fox, T.; Darden, T. A.; Kollman, P. A. *Journal of the American Chemical Society* **1995**, *117*, 4193.
(38) Darden, T.; York, D.; Pedersen, L. *Journal of Chemical Physics* **1993**, *98*, 10089.
(39) Hoover, W. G. *Physical Review A* **1986**, *34*, 2499.
(40) Hoover, W. G. *Physical Review A* **1985**, *31*, 1695.
(41) Nose, S. *J. Chem. Phys.* **1984**, *81*, 511.
(42) Nose, S. *Molecular Physics* **1984**, *52*, 255.
(43) Ponomarev, S. Y.; Thayer, K. M.; Beveridge, D. L. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 14771.
(44) Schomaker, V.; Waser, J.; Marsh, R. E.; Bergman, G. *Acta Crystallographica* **1959**, *12*, 600.
(45) Lavery, R.; Sklenar, H. *J. Biomol. Struct. Dyn.* **1989**, *6*, 655.
(46) Ravishanker, G.; Swaminathan, S.; Beveridge, D. L.; Lavery, R.; Sklenar, H. *J. Biomol. Struct. Dyn.* **1989**, *6*, 669.
(47) Lavery, R.; Sklenar, H. *J. Biomol. Struct. Dyn.* **1989**, *6*, 655.
(48) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379.
(49) Shrake, A.; Rupley, J. A. *J. Mol. Biol.* **1973**, *79*, 351.
(50) Boresch, S.; Hochtl, P.; Steinhauser, O. *J. Phys. Chem. B* **2000**, *104*, 8743.
(51) Mashimo, S.; Kuwabara, S.; Yagihara, S.; Higasi, K. *J. Phys. Chem.* **1987**, *91*, 6337.
(52) Nandi, N.; Bhattacharyya, K.; Bagchi, B. *Chemical Reviews* **2000**, *100*, 2013.
(53) Balasubramanian, S.; Bandyopadhyay, S.; Pal, S.; Bagchi, B. *Curr. Sci.* **2003**, *85*, 1571.
(54) Baar, C.; Buchner, R.; Kunz, W. *J. Phys. Chem. B* **2001**, *105*, 2906.
(55) Pal, S.; Balasubramanian, S.; Bagchi, B. *J. Chem. Phys.* **2004**, *120*, 1912.
(56) Lin, S. T.; Jang, S. S.; Cagin, T.; Goddard, W. A. *J. Phys. Chem. B* **2004**, *108*, 10041.
(57) Lin, S. T.; Blanco, M.; Goddard, W. A. *J. Chem. Phys.* **2003**, *119*, 11792.
(58) Lin, S. T.; Maiti, P. K.; Goddard, W. A. *J. Phys. Chem. B* **2005**, *109*, 8663.
(59) Li, Y. Y.; Lin, S. T.; Goddard, W. A. *Journal of the American Chemical Society* **2004**, *126*, 1872.
(60) Jang, S. S.; Lin, S. T.; Maiti, P. K.; Blanco, M.; Goddard, W. A.; Shuler, P.; Tang, Y. C. *J. Phys. Chem. B* **2004**, *108*, 12130.

(61) Jana, B.; Pal, S.; Maiti, P. K.; Lin, S. T.; Hynes, J. T.; Bagchi, B. *J. Phys. Chem. B* **2006**, *110*, 19611.
(62) Lim, K. T.; Brunett, S.; Iotov, M.; McClurg, R. B.; Vaidehi, N.; Dasgupta, S.; Taylor, S.; Goddard, W. A. *J. Comput. Chem.* **1997**, *18*, 501.
(63) Belhadj, M.; Alper, H. E.; Levy, R. M. *Chem. Phys. Lett.* **1991**, *179*, 13.
(64) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
(65) Lightstone, F. C.; Schwegler, E.; Hood, R. Q.; Gygi, F.; Galli, G. *Chem. Phys. Lett.* **2001**, *343*, 549.
(66) Ohtaki, H.; Radnai, T. *Chemical Reviews* **1993**, *93*, 1157.
(67) Naor, M. M.; Nostrand, K. V.; Dellago, C. *Chem. Phys. Lett.* **2003**, *369*, 159.

VI. Figures



a **b** **c**

Figure 1 *(a)* Schematic of DNA TWJ1 showing each of the three arms (A, B and C in the shaded circles) and the 3 characteristic orientation angles (Φ, ψ and θ) used to determine the helical stacking preference. A/B stacking corresponds to $\psi \leq 60^{o}$, and Φ, $\theta \approx 90^{o}$, whereas A/C stacking corresponds to $\Phi \leq 60^{o}$, and ψ, $\theta \approx 90^{o}$. Note that there is a third stacking configuration ($\theta \leq 60^{o}$, and ψ, $\Phi \approx 90^{o}$) that is termed B/C stacking. *(b)* TWJ1 nucleotide sequence with unique arms indicated as before (capital letter in shaded circles). Base sequence numbers are indicated with red numbers next to the nucleotide. The unpaired thymines at branch point and capping arms A and B are underlined. As drawn, this structure corresponds to A/B stacked T-shaped conformer *(c)* Ribbon structure of the A/B stacked starting structure (conformer 1) from the NMR ensemble. This conformer (used as the starting structure for the TWJ1_Na and MG_A MD simulations) was calculated to have had the second lowest average CRMS to the other conformers. Further, the stacking angle of unique arms A and C is 53^{0}, closest to the experimentally determined average of $53^{o} \pm 11^{o}$.

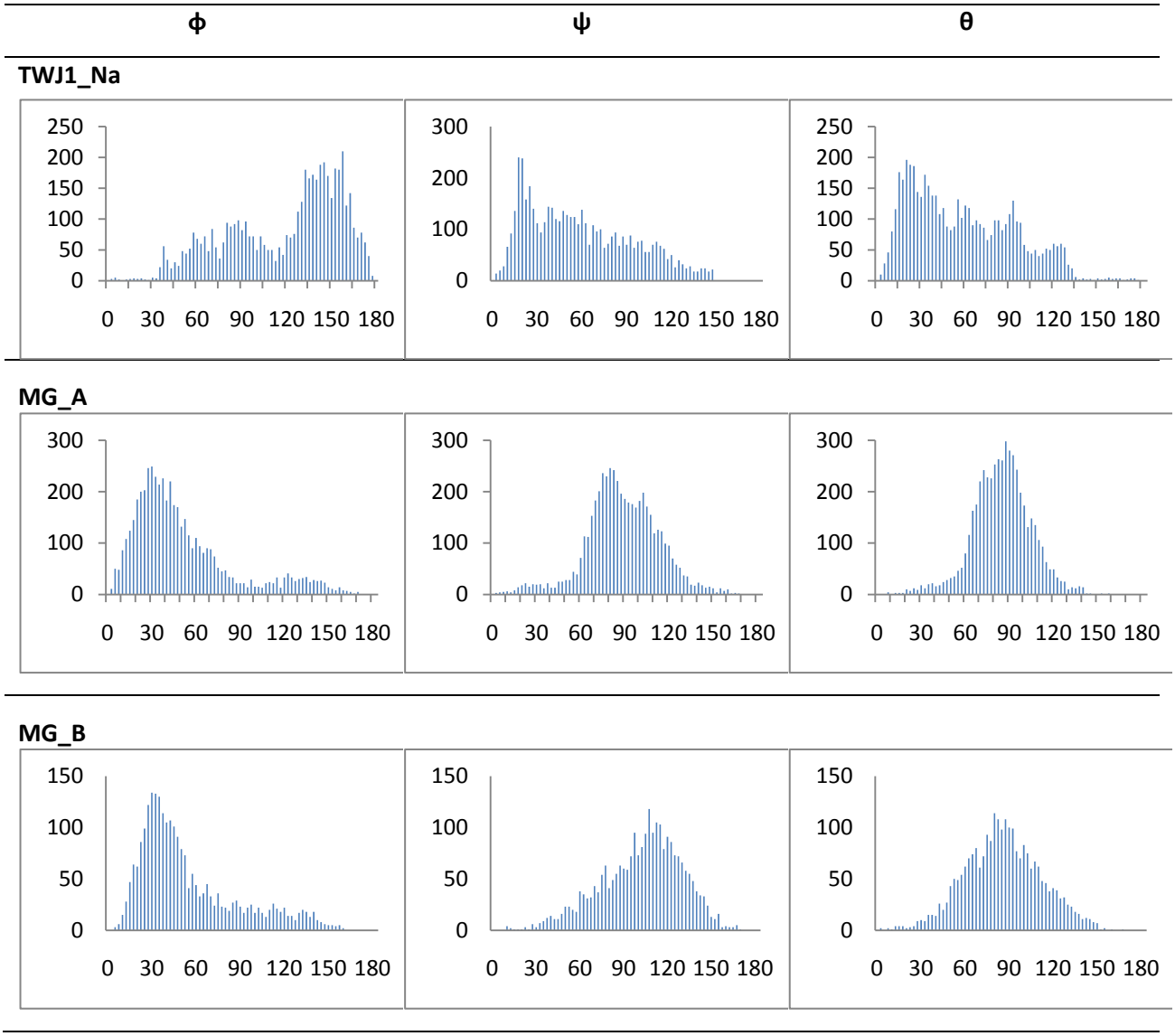


Figure 2 From left to right: distribution of ϕ, ψ, θ angles (degrees) for various TWJ1 simulations during 100 ns NVT MD simulation (25ns for MG_B). Note the multi-modal distribution in the ϕ and θ angles of TWJ1_Na, representing the two conformers: Na_aP (ϕ ≈ 60^o, θ ≈120^o) and Na_P (ϕ ≈ 150^o, θ ≈ 25^o), with both conformers showing A/B stacking (ψ ≈ 20^o ± 7.3^o). On the other hand, MG_A exhibits A/C stacking (ψ/ θ ≈ 90^o) with Gaussian like distributions for all three dihedral angles. The relative flexibility of TWJ1_Na (compared to the structurally more rigid MG_A) is reinforced by comparing the width of the various distributions. MG_B converged to the average MD structure of MG_A within 20ns of NVT simulations, and was therefore only simulated for 50ns. Analysis of the dihedral angle show similar distributions to that of MG_A, as expected, with half the frequency.

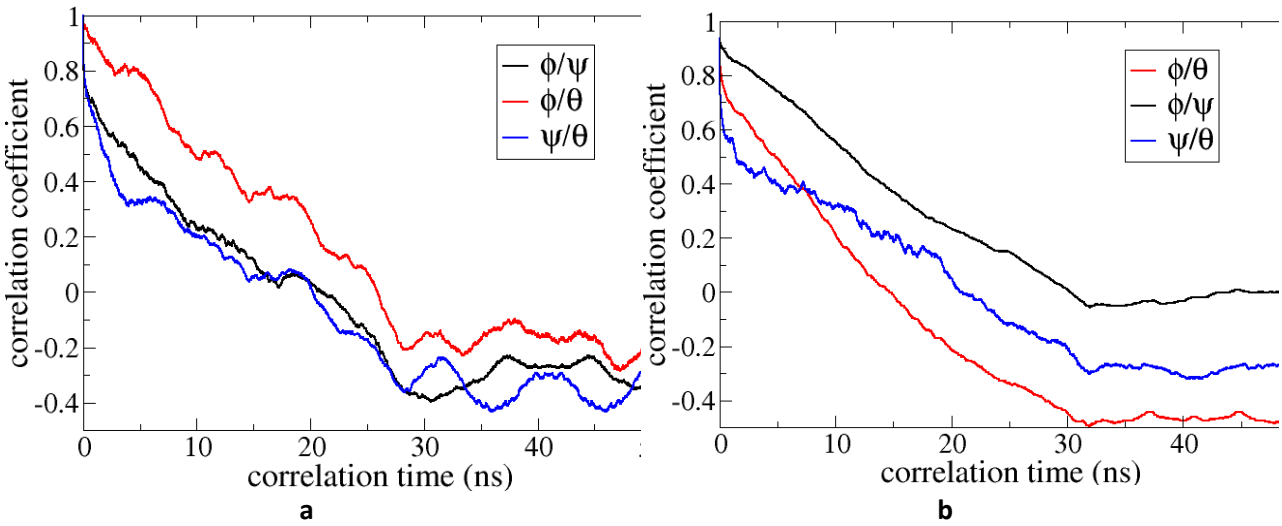


Figure 3 Plot of cross correlation function of each set of the three unique dihedral angles characterizing the stacking preference of TWJ1. Cross correlations of ϕ/θ (red) , ϕ/ψ (black) and ψ/θ (blue) are plotted as a function of correlation time (ns). (a) TWJ1_Na dihedral angle cross-correlation. The ϕ/θ angles move in a highly correlated fashion. This indicates an A/B helix exploring its conformation spaced as a rigid body stacked on arm C (A/B stacking). (b) The dihedral angle cross-correlation of MG_A. Here, unlike TWJ1_Na, the ϕ/ψ angles are highly correlated, indicative of the underlying A/C stacking configuration of MG_A.

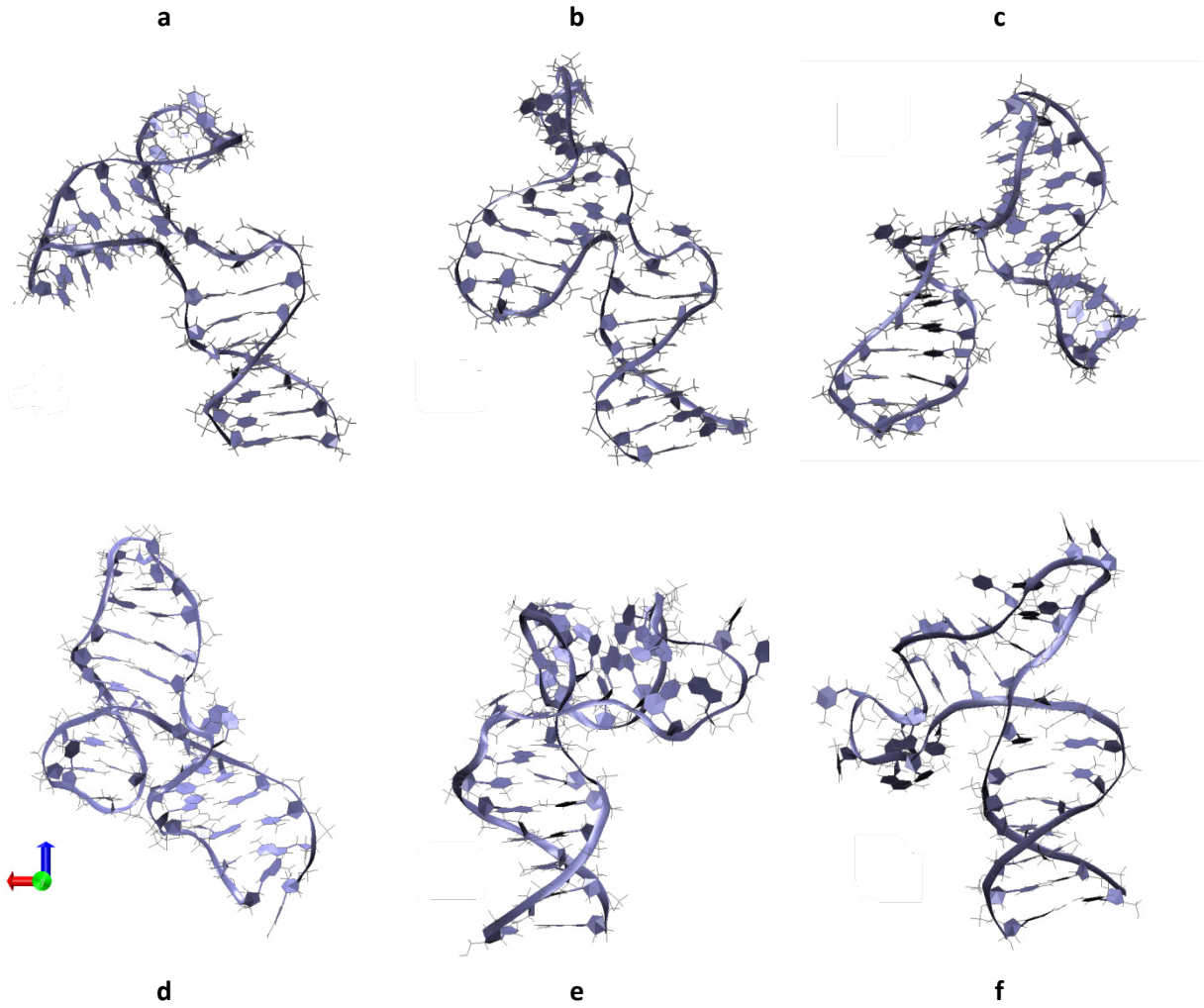


Figure 4 Best MD structure for of TWJ1_Na (a – c), MG_A (d) and MG_B (e and f) (a) NMR/starting structure. Each snapshot is rotated so as to minimize the total CRMSD (excluding the unpaired thymines), thereby illustrating the differences in the stacking of helix 2 on helix 1. Conformer 1 of the 26 NMR ensemble was taken as representative and used as starting structure for TWJ1_Na and MG_A simulations. A/B stacking is observed, with an acute A/C Φ stacking angle of 53°. (b) Anti-parallel conformer Na_aP. It is structurally similar to NMR structure with acute Φ stacking angle of 64°. (c) Parallel conformer Na_P. It also shows A/B stacking with obtuse acute A/C Φ stacking angle of 143°. During dynamics, TWJ1_Na adopts the Na_aP conformation for the first 25ns. There is then a transition to Na_aP between 27 and 32 ns. Na_P is assumed to be the true solution structure (at 300K), as TWJ1_Na stays in this conformation for the remaining 60ns of dynamics. We later show that the experimentally similar Na_aP structure would be more favorable at 285K. (d) MG_A solution structure. This structure is more structurally compact than any of the TWJ1_Na structures and has a different topology: it exhibits A/C stacking. (e) Starting structure of MG_B. MG_B is the only simulated TWJ1 structure that was built from scratch, using standard B-DNA geometries. (f) MD Solution structure of MG_B, showing convergence to the solution structure of MG_A (d) after 25ns.

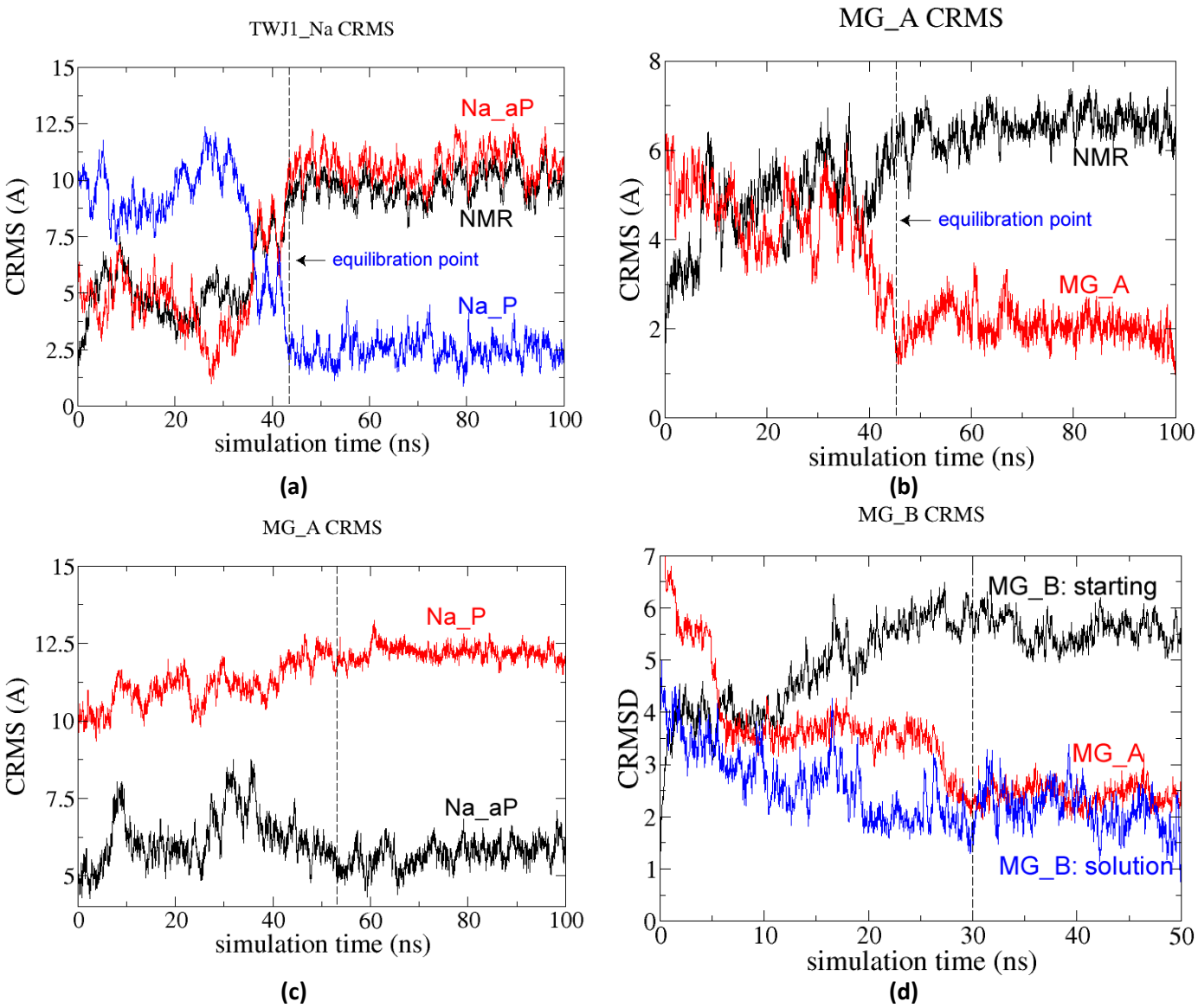


Figure 5 Cartesian root mean-squared deviation (CRMS) of (a) TWJ1_Na during dynamics, using the three best average md structures as reference: NMR/starting structure (black), Na_aP (red), Na_P (blue). The NMR/starting structure and Na_aP profiles are highly correlated during dynamics, further indicating the similarities of both structures. The transition from Na_aP to Na_P occurs between 37 and 42 ns, after which time, TWJ1_Na remains in the Na_P conformation (b) MG_A during dynamics, using the average md structure (red) and the starting NMR structure (black) as references. MG_A converges to it solution structure after 45ns of simulation, which is 6Å away from the starting NMR structure (c) Comparison of the CRMSD of MG_A during dynamics using the TWJ1_Na solution conformers as reference: Na_aP (black) and Na_P (red). MG_A is topologically different from either TWJ1_Na conformers, exhibiting B/C stacking. It is however more similar to Na_aP (6.42Å) than Na_P (12.26Å). (d) CRMS of MG_B during dynamics using the starting (Namot2 built) structure (black), the MD solution structure (blue) and the MD solution structure of MG_A (red) as references. Note that MG_B converges to with 2.5 Å of the global MG_A structure after 30ns of dynamics, after starting 7Å away.

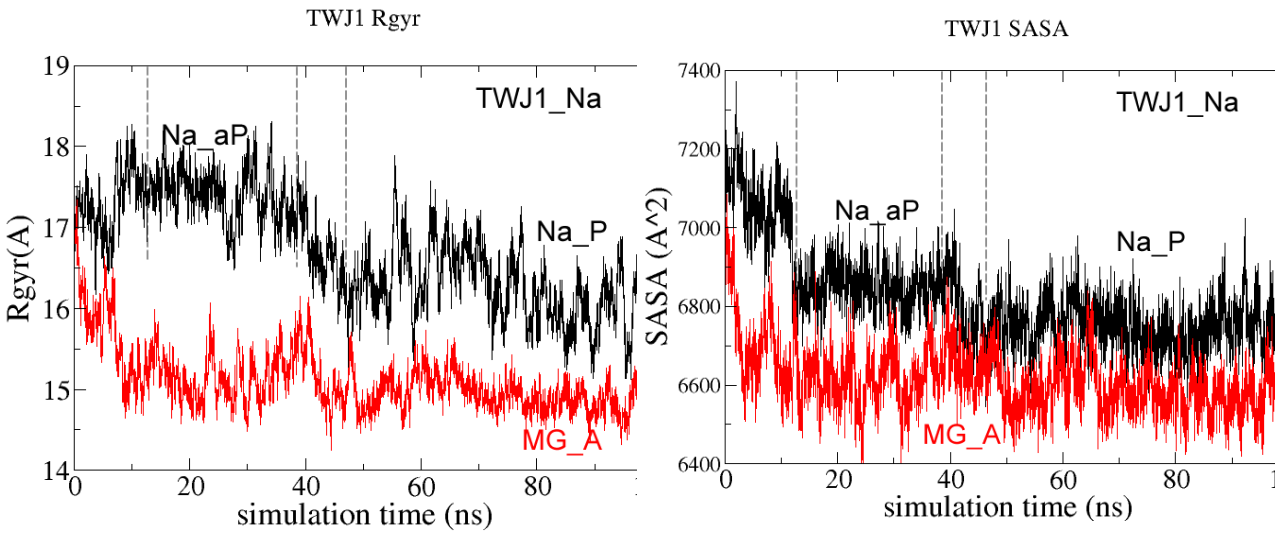


Figure 6 (a) Comparison of the radius of gyration of TWJ1_Na (black) and MG_A (red) during dynamics. After equilibrations, MG_A has a radius of gyration of 14.9Å, whereas Na_aP (12 – 35 ns) is 17.5Å and Na_P (42 – 100ns) is 16.3Å. (b) Comparison of the SASA of TWJ1 during dynamics. Na_P has a smaller SASA (6747 $Å^2$) than Na_aP (6901 $Å^2$), indicating a more compact structure. This is consistent with the radius of gyration measurements, and the fact that Na_P has a smaller A/C stacking angle than Na_aP. MG_A is the most compact (6588 $Å^2$) of the TWJ1 structure, which when combined with having the smallest radius of gyration indicates that MG_A is the most structurally rigid of all the TWJ1 conformers.

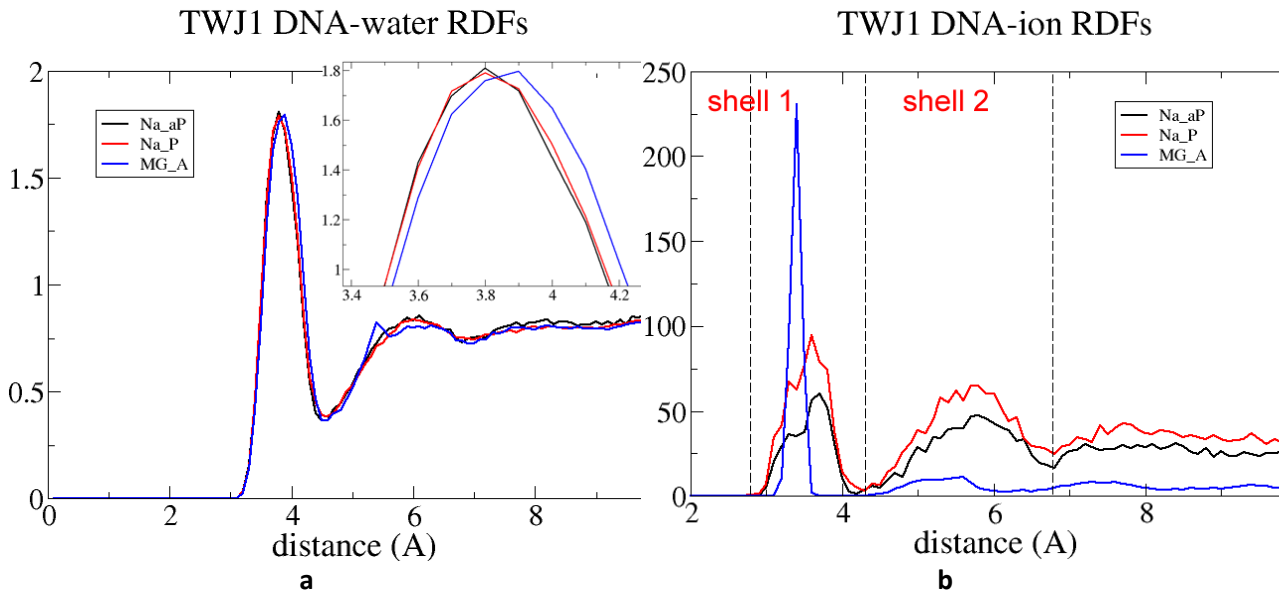


Figure 7 RDFs of Na_aP (black), Na_P (red) and MG_A (blue) during dynamics. (a) PO_4^- — water. A magnification of the first hydration shell is shown in the upper left corner. TWJ1_Na has slightly more water in the first solvation shell than the more compact MG_A. Two distinct hydration shells are found, the 1st from 2.5 – 4.6Å and the 2nd from 4.6 – 6.5Å (b) PO_4^-—ion RDFs using the different counterions. The RDFs for the TWJ1_Na structures have been multiplied x10 for visualization purposes. Note that the double charged counter-ion Mg^{2+} is strongly coordinated to the DNA backbone (13 of the 17 ions in the 1st shell), whereas TWJ1_Na has as many Na^+ ions in the 1st hydration shell (3 and 5 for Na_aP and Na_P respectively) as in the second, with approximately 90% of the ions at least 1 solvation shell away from the DNA.

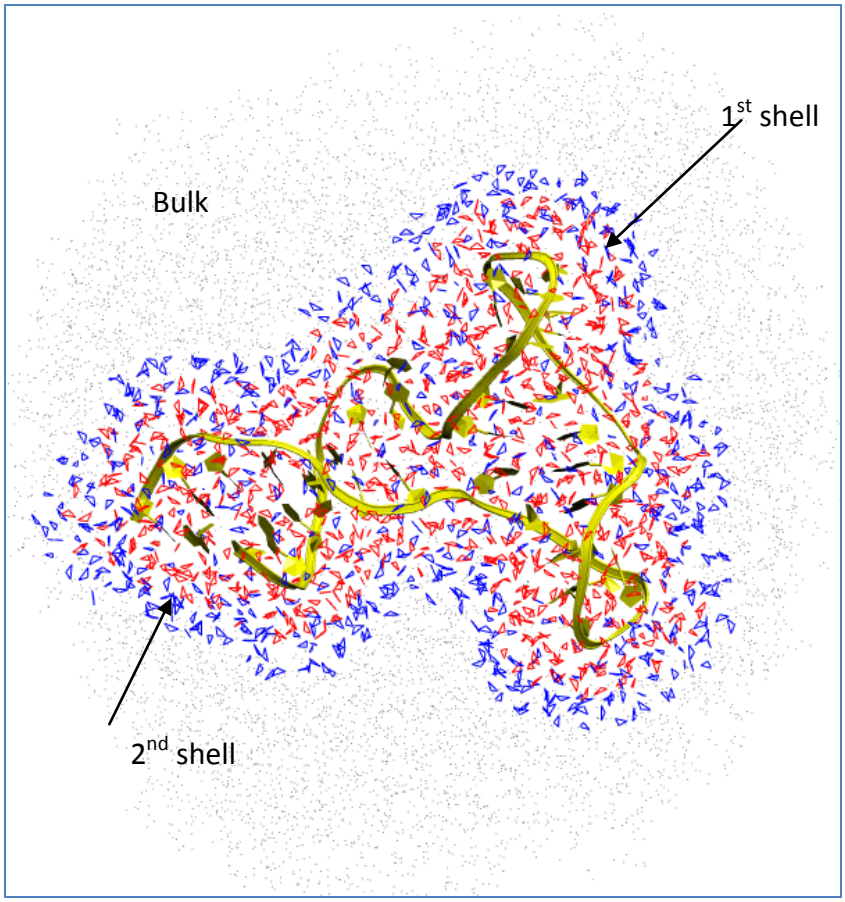


Figure 8 Two-dimensional projection of water in different groups around the starting NMR TWJ1 structure, used in the entropy calculation. Waters in the solvation shells are obtained from observing the water-phosphate RDF. Here all waters within 4.6Å of the phosphate backbone (249 molecules) are considered in shell 1, whereas all waters within 6.8Å of the phosphate group (380 molecules) are considered in shell 2. Water molecules centered at distances greater than 7.0Å from the TWJ1 backbone (7780 molecules) are considered as bulk water. The unit cell is shown by the enclosing blue box.

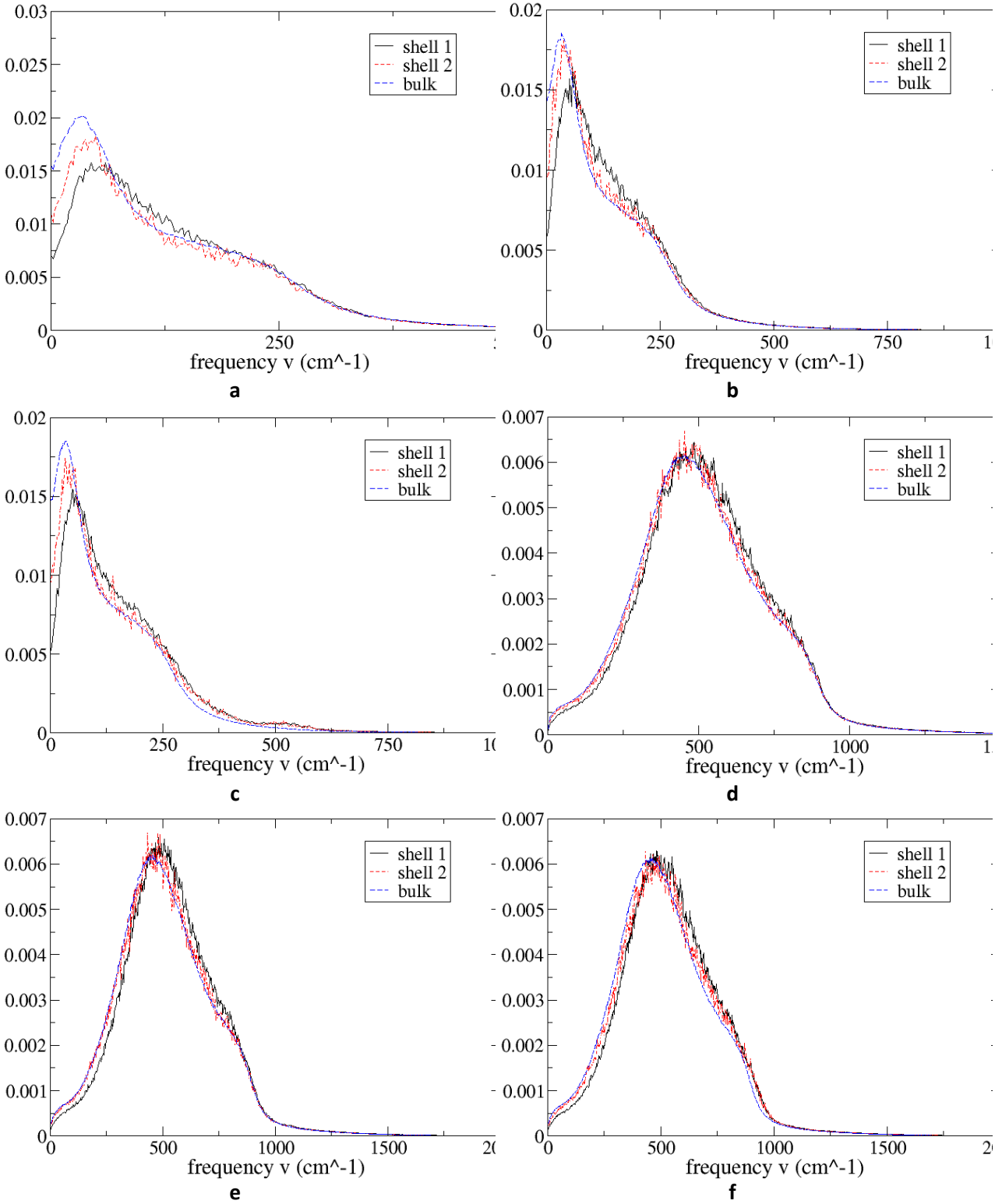


Figure 9 Translational (a – c) and rotational (d - f) components of the density of state for various types of water molecules in the Na_aP (a,d), Na_P (b,e) and MG_A (c,f) ensembles respectively. These calculations are done at 300K. We see an increase in the translational entropy of the waters in the 2nd hydration shell compared to bulk, whereas the waters in the 1st hydrations shell are more constrained,

both translationally and rotationally. Shown are the contributions obtained from decomposition of the total atomic velocities (equations 11 – 13) for water molecules in the 1st solvation shell (solid black line), 2nd solvation shell (red dashed line) and bulk waters (dashed blue line). The vibrational contributions to the velocity are all zero, since the SHAKE algorithm was used to constrain the bonds and angles of the water molecules to their equilibrium values.

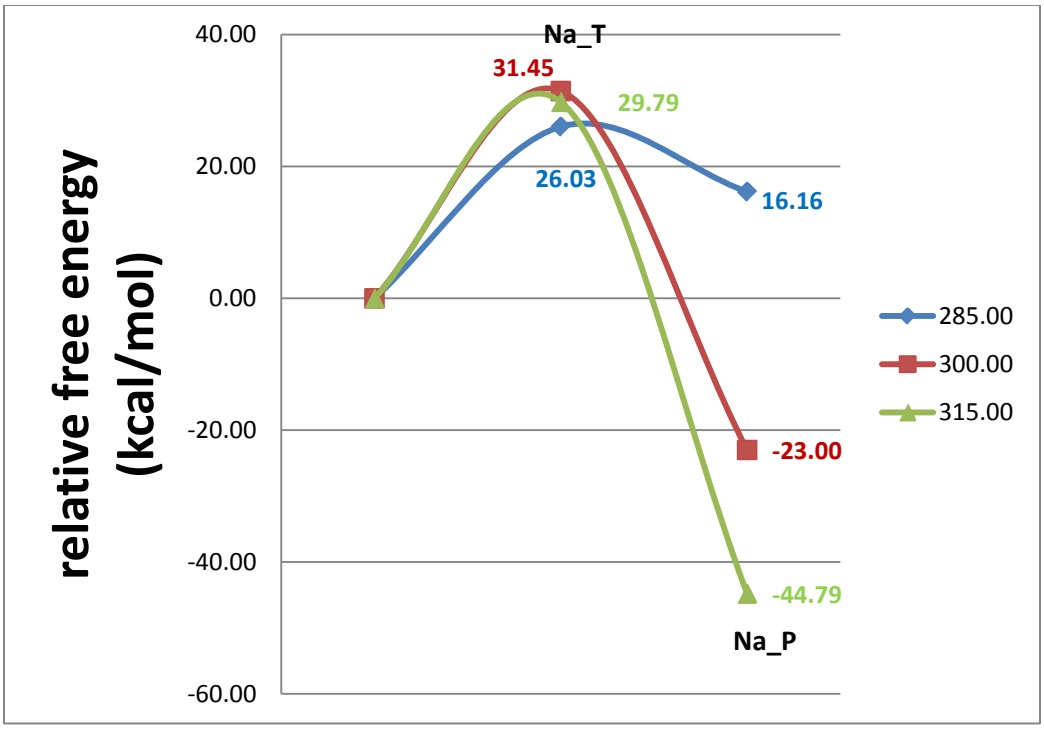


Figure 10: Helmholtz free energy (kcal/mol) comparison of Na_T and Na_P at 285K (blue), 300K (red) and 315K (green), relative to Na_aP. The statistics were averaged over the 10 best structures for each conformer. See figure S6 of the supplementary materials for the enthalpic and entropic contributions. Note that Na_P is less favorable than Na_aP at 285K

Na

V. References

(1) Cox, J. D. W., D.D.; Medvedev, V.A. *CODATA Key Values for Thermodynamics*; Hemisphere Publishing Corp.: New York, 1984.

(2) Kitts, P. A.; Nash, H. A. *Nature* **1987**, *329*, 346.

(3) Maiti, P. K.; Pascal, T. A.; Vaidehi, N.; Goddard, W. A. *NUCLEIC ACIDS RES* **2004**, *32*, 6047.

(4) Maiti, P. K.; Pascal, T. A.; Vaidehi, N.; Heo, J.; Goddard, W. A. *Biophys. J.* **2006**, *90*, 1463.

(5) Maiti, P. K. P., Tod A.; Vaidehi, Nagarajan; Goddard, William A. *J. Nanosci. Nanotechnol.* **2007**, *7*, 1712.

(6) Holliday, R. *Genetical Research* **1964**, *5*, 282.

(7) Kallenbach, N. R.; Ma, R. I.; Seeman, N. C. *Nature* **1983**, *305*, 829.

(8) Nunesduby, S. E.; Matsumoto, L.; Landy, A. *Cell* **1987**, *50*, 779.

(9) Altona, C.; Pikkemaat, J. A.; Overmars, F. J. J. *Curr. Opin. Struct. Biol.* **1996**, *6*, 305.

(10) Lilley, D. M. J.; Clegg, R. M. *Quarterly Reviews of Biophysics* **1993**, *26*, 131.

(11) Noller, H. F. *Annual Review of Biochemistry* **1991**, *60*, 191.

(12) Rettberg, C. C.; Prere, M. F.; Gesteland, R. F.; Atkins, J. F.; Fayet, O. *J. Mol. Biol.* **1999**, *286*, 1365.

(13) Seeman, N. C. *DNA and Cell Biology* **1991**, *10*, 475.

(14) Shen, Z. Y.; Yan, H.; Wang, T.; Seeman, N. C. *Journal of the American Chemical Society* **2004**, *126*, 1666.

(15) Maiti, P. K.; Pascal, T. A.; Vaidehi, N.; Goddard, W. A. *Nucleic Acids Research* **2004**, *32*, 6047.

(16) Warner, R. C.; Fishel, R. A.; Wheeler, F. C. *Cold Spring Harbor Symposia on Quantitative Biology* **1978**, *43*, 957.

(17) Duckett, D. R.; Murchie, A. I. H.; Diekmann, S.; Vonkitzing, E.; Kemper, B.; Lilley, D. M. J. *Cell* **1988**, *55*, 79.

(18) McKinney, S. A.; Declais, A. C.; Lilley, D. M. J.; Ha, T. *Nature Structural Biology* **2003**, *10*, 93.

(19) Yu, J.; Ha, T. J.; Schulten, K. *Nucleic Acids Res.* **2004**, *32*, 6683.

(20) Altona, C. *J. Mol. Biol.* **1996**, *263*, 568.

(21) Welch, J. B.; Walter, F.; Lilley, D. M. J. *J. Mol. Biol.* **1995**, *251*, 507.

(22) Muhuri, S.; Mimura, K.; Miyoshi, D.; Sugimoto, N. *Journal of the American Chemical Society* **2009**, *131*, 9268.

(23) Yan, H.; Zhang, X. P.; Shen, Z. Y.; Seeman, N. C. *Nature* **2002**, *415*, 62.

(24) van Buuren, B. N. M.; Overmars, F. J. J.; Ippel, J. H.; Altona, C.; Wijmenga, S. S. *J. Mol. Biol.* **2000**, *304*, 371.

(25) Ouporov, I. V.; Leontis, N. B. *Biophys. J.* **1995**, *68*, 266.

(26) Overmars, F. J. J.; Pikkemaat, J. A.; vandenElst, H.; vanBoom, J. H.; Altona, C. *J. Mol. Biol.* **1996**, *255*, 702.

(27) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535.

(28) Tung, C. S.; Carter, E. S. *CABIOS* **1994**, *10:*, 427.

(29) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.

(30) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668.

(31) Holst, M. J.; Saied, F. *J. Comput. Chem.* **1995**, *16*, 337.

(32) Holst, M.; Saied, F. *J. Comput. Chem.* **1993**, *14*, 105.

(33) Cheatham, T. E.; Cieplak, P.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845.

(34) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817.

(35) Cheatham, T. E.; Miller, J. L.; Fox, T.; Darden, T. A.; Kollman, P. A. *Journal of the American Chemical Society* **1995**, *117*, 4193.

(36) Darden, T.; York, D.; Pedersen, L. *Journal of Chemical Physics* **1993**, *98*, 10089.

(37) Hoover, W. G. *Physical Review A* **1986**, *34*, 2499.

(38) Hoover, W. G. *Physical Review A* **1985**, *31*, 1695.

(39) Nose, S. *J. Chem. Phys.* **1984**, *81*, 511.

(40) Nose, S. *Molecular Physics* **1984**, *52*, 255.

(41) Ponomarev, S. Y.; Thayer, K. M.; Beveridge, D. L. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 14771.

(42) Schomaker, V.; Waser, J.; Marsh, R. E.; Bergman, G. *Acta Crystallographica* **1959**, *12*, 600.

(43) Lavery, R.; Sklenar, H. *J. Biomol. Struct. Dyn.* **1989**, *6*, 655.

(44) Ravishanker, G.; Swaminathan, S.; Beveridge, D. L.; Lavery, R.; Sklenar, H. *J. Biomol. Struct. Dyn.* **1989**, *6*, 669.

(45) Lavery, R.; Sklenar, H. *J. Biomol. Struct. Dyn.* **1989**, *6*, 655.

(46) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379.

(47) Shrake, A.; Rupley, J. A. *J. Mol. Biol.* **1973**, *79*, 351.

(48) Boresch, S.; Hochtl, P.; Steinhauser, O. *J. Phys. Chem. B* **2000**, *104*, 8743.

(49) Mashimo, S.; Kuwabara, S.; Yagihara, S.; Higasi, K. *J. Phys. Chem.* **1987**, *91*, 6337.

(50) Nandi, N.; Bhattacharyya, K.; Bagchi, B. *Chemical Reviews* **2000**, *100*, 2013.

(51) Balasubramanian, S.; Bandyopadhyay, S.; Pal, S.; Bagchi, B. *Curr. Sci.* **2003**, *85*, 1571.

(52) Baar, C.; Buchner, R.; Kunz, W. *J. Phys. Chem. B* **2001**, *105*, 2906.

(53) Pal, S.; Balasubramanian, S.; Bagchi, B. *J. Chem. Phys.* **2004**, *120*, 1912.

(54) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.

(55) Lin, S. T.; Blanco, M.; Goddard, W. A. *J. Chem. Phys.* **2003**, *119*, 11792.

(56) Lightstone, F. C.; Schwegler, E.; Hood, R. Q.; Gygi, F.; Galli, G. *Chem. Phys. Lett.* **2001**, *343*, 549.

(57) Ohtaki, H.; Radnai, T. *Chemical Reviews* **1993**, *93*, 1157.

(58) Naor, M. M.; Nostrand, K. V.; Dellago, C. *Chem. Phys. Lett.* **2003**, *369*, 159.

(59) Lin, S. T.; Jang, S. S.; Cagin, T.; Goddard, W. A. *J. Phys. Chem. B* **2004**, *108*, 10041.

(60) Eimer, W.; Pecora, R. *The Journal of Chemical Physics* **1991**, *94*, 2324.
(61) Marcus, Y. *Chemical Reviews* **1988**, *88*, 1475.

SI. Supplementary Tables

Table S1: Statistical analysis of the 26 NMR conformers deposited as representative of TWJ1 structure. Conformer 1 is selected on the basis of having the smallest average CRMS deviation to the other 25 conformers and the ψ angle (53°) closest to the experimentally reported 53° ± 11°. Note that the structure differences between the minimized and starting structures (as measured by the CRMS in column 2) is negligible (0.12Å average), suggesting that each of the NMR structures are in a local minima as determined by our force field.

	Starting Structure		Minimized Structure		CRMS (Å) to other conformers				Stacking Angles (degrees)		
Conformer	**[a]CRMS(Å)**	**Enthalpy (kcal/mol)**	**Enthalpy (kcal/mol)**	**ΔH (kcal/mol)**	**Avg.**	**±**	**Min**	**Max**	**Ψ**	**ϕ**	**θ**
1	0.07	-4900.5	-5387.5	-487.0	1.6	0.8	1.3	4.8	53.4	63.0	36.4
2	0.11	-4831.7	-5468.8	-637.1	2.2	0.6	0.8	3.9	19.3	152.6	160.0
3	0.07	-4749.0	-5309.7	-560.8	1.6	0.7	0.5	3.4	33.8	47.1	32.9
4	0.1	-4905.9	-5531.1	-625.2	3.0	0.9	1.3	5.2	34.7	53.3	27.5
5	0.13	-4975.5	-5613.0	-637.5	1.6	0.7	0.5	3.3	28.3	48.0	35.1
6	0.08	-4795.0	-5376.8	-581.9	1.8	0.7	1.0	3.9	58.2	63.1	90.5
7	0.14	-4448.4	-5532.4	-1084.0	3.5	0.6	2.5	4.9	52.1	17.4	69.3
8	0.15	-4684.7	-5319.5	-634.8	2.5	0.8	1.3	4.8	16.6	118.1	108.4
9	0.05	-5061.9	-5351.9	-290.0	2.1	0.6	1.3	3.4	165.1	78.4	88.9
10	0.13	-4764.7	-5303.1	-538.4	1.9	0.6	0.9	3.1	46.8	49.2	92.0
11	0.1	-4738.5	-5263.9	-525.5	1.9	0.7	1.0	3.9	37.4	61.8	49.1
12	0.13	-5104.2	-5631.4	-527.3	2.1	0.7	1.2	3.8	24.4	67.7	53.9
13	0.12	-4829.3	-5394.4	-565.1	2.4	0.6	1.1	4.4	120.3	39.9	123.0
14	0.08	-4854.0	-5394.3	-540.3	1.7	0.7	0.8	3.7	144.5	42.7	132.3
15	0.14	-5085.5	-5673.8	-588.2	1.9	0.5	1.2	3.1	46.6	38.5	65.8
16	0.13	-4972.2	-5479.5	-507.3	2.7	0.7	1.6	4.5	63.2	34.0	86.0
17	0.12	-4958.9	-5504.9	-545.9	1.9	0.7	0.8	3.6	132.9	52.5	99.6
18	0.11	-4836.6	-5370.0	-533.5	2.0	0.7	0.8	3.8	44.4	46.6	36.5
19	0.14	-4801.8	-5362.4	-560.6	1.7	0.7	0.8	3.5	30.4	46.7	52.7
20	0.14	-4995.4	-5449.3	-454.0	1.9	0.7	1.2	3.9	24.7	65.4	45.0
21	0.14	-4686.7	-5330.9	-644.2	1.9	0.7	1.0	3.9	138.8	70.2	125.5
22	0.11	-5035.8	-5535.6	-499.8	1.9	0.6	0.9	3.2	41.5	128.1	102.2
23	0.13	-5234.0	-5812.9	-578.9	1.6	0.6	0.9	3.4	138.7	52.0	107.3
24	0.14	-4711.9	-5303.7	-591.8	3.3	0.8	2.0	5.2	59.4	159.8	117.8
25	0.13	-4934.8	-5508.7	-573.9	1.9	0.6	0.8	3.7	39.8	132.0	105.9
26	0.11	-5008.8	-5582.4	-573.6	1.8	0.7	0.8	3.4	26.8	135.5	153.8

[a]CRMS computed between starting NMR and minimized structures

Table S2: List of Twist angle and helical rise per base-pair for TWJ1_Na (figure 5b,c)

	NA_P				Na_aP				NMR[a]			
	Rise	**±**	**Twist**	**±**	**Rise**	**±**	**Twist**	**±**	**Rise**	**±**	**Twist**	**±**
B DNA	3.4		36									
Helix 1												
C-1/G-2	4.2	2.9	14.2	7.0	4.5	2.5	12.5	9.1	3.5	0.1	36.9	0.3
G-2/T-3	3.9	0.9	31.0	7.8	3.4	0.8	27.5	6.7	3.7	0.1	33.4	0.3
T-3/G-4	3.3	1.0	29.8	9.9	2.8	0.8	24.3	8.5	3.5	0.1	37.2	0.5
G-4/C-5	3.9	0.9	30.6	6.2	3.9	0.6	31.4	5.3	3.6	0.1	32.5	0.5
C-5/A-6	3.6	1.0	30.7	6.6	3.6	0.7	32.9	5.8	4.2	0.2	37.4	0.8
A-6/C-7	3.3	0.7	26.8	6.3	3.1	0.5	22.1	4.7	3.6	0.2	25.6	1.7
AVG	**3.7**	**0.4**	**27.2**	**6.5**	**3.5**	**0.6**	**25.1**	**7.4**	**3.7**	**0.3**	**33.8**	**4.5**
Helix 2												
G-1/T-2	3.2	0.3	8.8	11.5	3.0	0.4	7.8	11.0	3.6	0.1	22.2	0.7
T-2/C-3	3.1	0.6	21.5	8.2	3.2	0.4	29.1	9.1	4.6	0.2	30.2	1.0
C-3/G-4	3.6	0.5	39.0	5.3	3.8	0.5	39.8	6.1	4.6	0.2	36.3	1.3
G-4/C-5	3.2	0.4	32.3	6.6	3.5	0.6	23.5	8.9	4.6	0.3	27.7	3.6
C-5/C-6	4.2	0.5	32.5	4.9	3.9	0.5	37.9	5.4	4.9	0.2	39.1	2.0
C-6/G-7	2.7	0.5	28.8	8.2	2.8	0.4	27.8	8.0	5.3	0.4	31.5	1.2
G-7/C-8	3.3	0.4	23.2	6.3	3.2	0.3	20.1	6.2	3.5	0.1	23.0	0.6
AVG	**3.3**	**0.5**	**26.6**	**9.8**	**3.3**	**0.4**	**26.6**	**10.9**	**4.4**	**0.7**	**30.0**	**6.3**

[a]These values are obtained from the minimized NMR structures.

Table S3: Comparison of NOE distances of NMR (26 structures), Na_aP (628 structures), Na_P (2002 structures) and MG_A (2002 structures). The NMR structures were obtained from the minimized ensemble structures, Na_aP from snapshots 5 – 25ns of TWJ1_Na MD, Na_P from snapshots 60 – 100 ns of TWJ1_Na MD and MG_A from 60 – 100 ns of TWJ1_MG_A MD.

	NMR		Na_aP			Na_P			MG_A		
	AVG	±	AVG	±	Diff	AVG	±	Diff	AVG	±	Diff
C18_2H2* G17_2H2*	5.96	0.32	5.51	0.35	0.45	5.57	0.31	0.39	5.69	0.36	0.27
C18_H5 G17_2H2*	3.47	0.28	3.75	0.55	0.28	3.81	0.53	0.34	3.62	0.62	0.15
C18_H5 G17_H1*	4.89	0.17	4.87	0.46	0.02	4.93	0.43	0.04	4.90	0.48	0.01
C18_H5 G17_H2*	2.55	0.12	3.05	0.53	0.50	3.40	0.56	0.85	3.06	0.51	0.51
C18_H5 G17_H3*	4.32	0.16	5.28	0.60	0.96	5.45	0.62	1.13	4.97	0.58	0.65
C18_H5 G17_H8	4.09	0.38	4.02	0.39	0.07	4.11	0.40	0.02	4.37	0.42	0.28
C18_H6 G17_2H2*	2.62	0.17	2.45	0.33	0.17	2.45	0.33	0.17	2.52	0.37	0.10
C18_H6 G17_H1*	4.66	0.27	4.30	0.53	0.36	4.04	0.50	0.62	4.28	0.54	0.38
C18_H6 G17_H2*	2.96	0.25	3.01	0.58	0.05	3.10	0.59	0.14	3.08	0.60	0.12
C18_H6 G17_H3*	4.15	0.27	4.56	0.40	0.41	4.66	0.41	0.51	4.49	0.39	0.34
C7_2H2* G27_H3*	3.97	0.13	4.96	0.60	0.99	3.69	0.37	0.28	3.01	0.38	0.96
C7_2H2* T28_H6	3.68	0.12	6.38	1.02	2.70	3.70	0.38	0.02	4.10	0.52	0.42
C7_2H2* T28_H6	3.68	0.12	6.38	1.02	2.70	3.70	0.38	0.02	4.10	0.52	0.42
C7_H1* T28_H4*	4.63	0.06	4.22	0.53	0.41	4.18	0.45	0.45	3.14	0.34	1.49
C7_H1* T28_H6	4.52	0.13	7.16	0.98	2.64	4.57	0.39	0.05	4.48	0.48	0.04
C7_H2* G27_H3*	5.01	0.18	5.40	0.87	0.39	5.21	0.45	0.20	3.95	0.43	1.06
C7_H2* T28_H6	4.41	0.10	6.56	1.18	2.15	2.47	0.31	1.94	5.47	0.57	1.06
C7_H2* T28_H6	4.41	0.10	6.56	1.18	2.15	2.47	0.31	1.94	5.47	0.57	1.06
C7_H3* C8_H5	6.20	0.14	6.42	0.75	0.22	9.91	0.64	3.71	2.71	0.30	3.49
C7_H3* T28_H6	6.30	0.07	8.55	1.16	2.25	4.36	0.41	1.94	6.72	0.49	0.42
C7_H3* T28_H6	6.30	0.07	8.55	1.16	2.25	4.36	0.41	1.94	6.72	0.49	0.42
C7_H5 T28_H6	5.94	0.07	6.19	1.17	0.25	4.89	0.51	1.05	4.93	1.14	1.01
C7_H6 T28_H6	5.55	0.07	6.80	1.21	1.25	4.31	0.44	1.24	4.74	0.98	0.81
C8_H5 G27_2H2	7.91	0.57	7.62	0.48	0.29	6.36	0.59	1.55	6.66	0.48	1.25
C8_H5 G27_H1*	5.81	0.39	4.97	0.57	0.84	4.02	0.74	1.79	4.63	0.42	1.18
C8_H5 G27_H8	4.28	0.31	3.92	0.44	0.36	4.10	0.47	0.18	3.99	0.36	0.29
C8_H6 G27_2H2	7.49	0.51	6.71	0.50	0.78	5.34	0.58	2.15	5.77	0.45	1.72
C8_H6 G27_H1*	5.24	0.39	4.77	0.39	0.47	3.09	0.44	2.15	3.66	0.37	1.58
G27_2H2* T28_3H5M	5.97	0.49	9.29	1.20	3.32	9.17	1.12	3.20	6.77	0.96	0.80
G27_2H2* T28_H6	5.61	0.25	7.47	0.59	1.86	7.46	0.52	1.85	5.41	0.48	0.20
G27_H2* T28_3H5M	5.40	0.51	9.19	1.28	3.79	10.10	1.03	4.70	6.00	0.99	0.60
G27_H2* T28_H6	5.71	0.27	7.77	0.64	2.06	8.11	0.49	2.40	5.26	0.45	0.45
G27_H3* T28_3H5M	3.46	0.49	7.57	1.02	4.11	8.42	0.94	4.96	4.15	0.93	0.69
G27_H3* T28_H6	3.57	0.30	6.08	0.49	2.51	6.19	0.48	2.62	3.24	0.39	0.33
G27_H4* T28_3H5M	5.25	0.54	8.34	0.99	3.09	7.63	0.97	2.38	5.13	0.88	0.12
G27_H4* T28_H6	3.96	0.56	6.05	0.82	2.09	6.26	0.42	2.30	5.24	0.35	1.28
G27_H8 T28_3H5M	6.32	0.51	9.26	1.25	2.94	11.54	1.10	5.22	7.35	1.00	1.03
G30_H1* T28_H4*	5.61	0.25	6.98	0.58	1.37	6.33	0.50	0.72	5.92	0.51	0.31
T28_1H2* T29_H4*	5.65	0.09	6.08	0.51	0.43	6.09	0.60	0.44	4.78	0.51	0.87
T28_1H2* T29_H6	3.22	0.06	3.20	0.40	0.02	3.15	0.52	0.07	2.55	0.35	0.67
T28_2H2* T29_H6	2.42	0.10	2.53	0.44	0.11	2.61	0.46	0.19	3.70	0.44	1.28
T28_H1* T29_H4*	4.05	0.10	4.48	0.65	0.43	4.97	0.52	0.92	5.82	0.35	1.77
T28_H1* T29_H6	4.51	0.14	4.18	0.45	0.33	4.36	0.47	0.15	5.06	0.35	0.55
T28_H3* T29_3H5M	3.38	0.12	5.51	0.93	2.13	5.28	0.95	1.90	4.11	0.90	0.73
T28_H3* T29_H6	4.02	0.12	4.44	0.37	0.42	4.64	0.43	0.62	3.93	0.40	0.09
T28_H4* T29_3H5M	6.08	0.07	6.35	0.82	0.27	7.14	0.82	1.06	6.80	0.84	0.72
T28_H4* T29_H4*	5.74	0.12	5.54	0.34	0.20	6.58	0.32	0.84	6.37	0.30	0.63
T28_H4* T29_H6	5.73	0.05	6.24	0.27	0.51	6.14	0.31	0.41	5.95	0.30	0.22
T28_H6 T29_3H5M	5.03	0.06	5.24	0.86	0.21	4.36	0.89	0.67	4.65	0.87	0.38

Table S4: Self diffusion constants for Na_aP, Na_P and MG_A conformers during dynamics. The values are calculated during the dynamics when the system is in equilibrium: 10 – 20 ns for Na_aP and 90 – 100 ns for Na_P and MG_A. Experimental results from reference[60-61]

	Self Diffusion Constant (x 10^{-5} cm^2 s^{-1})			
	Na_aP	**Na_P**	**MG_A**	**Exp**
DNA	0.057	0.024	0.008	0.015-0.011
Ions	0.46	0.42	0.04	2.36
Water (Shell1)	1.37	1.42	2.14	
Water (Shell2)	2.38	2.52	2.49	
Water (Bulk)	3.40	3.39	3.37	

Table S5: Contributions to free energy (kcal/mol) of TWJ1 conformers during dynamics at 285K. The 10 best structures for each conformer were split into 5 groups and the results averaged over 10ps over NVT dynamics.

[a]Group	A (kcal/mol)		U (kcal/mol)		S (cal/mol/K)		# molecules	
	Avg.	±	Avg.	±	Avg.	±	Avg.	±
MG_A								
DNA	-5733.42	15.22	-4978.53	13.35	2212.05	15.30	1.00	0.00
IONS	-129.53	0.77	-96.33	0.27	114.60	8.19	18.00	0.00
Shell1	-6413.51	140.79	-4623.04	123.81	6261.54	165.74	430.50	9.11
Shell2	-2450.18	159.73	-1706.46	139.95	2631.12	191.75	158.70	11.27
Bulk	-119093.82	176.51	-84178.29	169.02	122948.46	408.33	7819.80	16.22
Na_aP								
DNA	-5473.86	20.93	-4722.60	13.77	2234.21	17.90	1.00	0.00
IONS	-293.60	1.30	-181.99	0.46	405.28	8.36	35.00	0.00
Shell1	-6466.61	142.88	-4553.23	93.62	6746.05	133.74	444.60	7.36
Shell2	-2379.10	189.75	-1620.80	122.27	2699.51	194.82	158.10	10.36
Bulk	-114760.38	227.10	-80054.10	167.69	122722.91	190.68	7806.30	10.96
Na_P								
DNA	-5466.58	13.71	-4713.53	10.33	2242.38	26.02	1.00	0.00
IONS	-291.65	1.18	-181.65	0.29	399.12	10.50	35.00	0.00
Shell1	-6293.57	239.46	-4431.54	156.31	6564.91	209.06	433.56	13.32
Shell2	-2436.83	181.50	-1664.29	118.05	2748.34	174.12	162.67	9.97
Bulk	-114666.87	179.93	-79970.58	109.99	122696.11	243.47	7812.78	8.16
Na_T								
DNA	-5493.17	16.05	-4738.67	11.15	2244.42	20.88	1.00	0.00
IONS	-291.80	0.85	-182.47	0.36	396.47	3.80	35.00	0.00
Shell1	-6531.20	233.22	-4607.51	154.08	6778.21	181.95	448.70	12.74
Shell2	-2512.40	153.95	-1716.71	100.89	2830.34	140.80	167.00	8.43
Bulk	-114774.51	194.87	-80135.41	131.73	122451.95	218.90	7793.30	13.25

[a]shell1, shell2 and bulk waters as determined by the DNA-water RDFs..

Table S6: Contributions to free energy (kcal/mol) of TWJ1 conformers during dynamics at 300K.

Group	A (kcal/mol)		U (kcal/mol)		S (cal/mol/K)		# molecules	
	Avg.	±	Avg.	±	Avg.	±	Avg.	±
MG_A								
DNA	-5725.94	16.70	-4896.19	11.66	2333.31	20.16	1.00	0.00
IONS	-131.78	0.90	-94.76	0.22	121.64	6.95	18.00	0.00
Shell1	-6401.32	201.54	-4444.12	131.18	6510.66	207.05	420.70	10.80
Shell2	-2311.99	216.61	-1548.60	140.64	2565.64	206.42	146.40	11.57
Bulk	-121766.46	263.48	-83046.66	188.50	129492.65	175.05	7841.90	10.41
Na_aP								
DNA	-5461.82	27.03	-4623.74	16.23	2355.77	21.18	1.00	0.00
IONS	-302.67	0.87	-178.13	0.54	427.92	7.71	35.00	0.00
Shell1	-6560.18	159.14	-4427.39	101.43	7117.78	172.35	441.40	8.38
Shell2	-2530.47	199.42	-1650.66	129.31	2963.09	197.44	164.40	10.98
Bulk	-116865.25	223.31	-78375.46	177.13	128685.46	265.49	7803.20	11.16
Na_P								
DNA	-5462.07	20.01	-4619.53	13.69	2358.74	19.93	1.00	0.00
IONS	-299.96	1.24	-177.95	0.44	418.94	9.65	35.00	0.00
Shell1	-6311.05	238.94	-4259.59	156.45	6844.06	233.59	425.11	12.95
Shell2	-2537.65	211.31	-1656.17	136.12	2968.23	203.90	165.11	11.69
Bulk	-117132.65	137.65	-78457.01	99.15	129320.50	274.32	7818.78	15.48
Na_T								
DNA	-5469.71	17.58	-4624.66	11.96	2381.66	22.61	1.00	0.00
IONS	-300.88	1.22	-178.12	0.39	421.43	6.96	35.00	0.00
Shell1	-6425.17	91.73	-4340.51	59.78	6955.49	103.63	432.80	4.83
Shell2	-2534.12	157.21	-1657.19	101.69	2951.64	147.97	165.10	8.73
Bulk	-116959.06	241.30	-78447.10	166.61	128761.00	287.54	7811.10	9.45

Table S7: Contributions to free energy (kcal/mol) of TWJ1 conformers during dynamics at 315K.

Group	A (kcal/mol)		U (kcal/mol)		S (cal/mol/K)		# molecules	
	Avg.	±	Avg.	±		Avg.	±	Avg.
MG_A								
DNA	-5720.43	26.03	-4810.85	16.84	2455.46	26.70	1.00	0.00
IONS	-134.33	0.55	-93.14	0.29	129.14	4.82	18.00	0.00
Shell1	-6477.54	228.05	-4324.80	150.27	6827.56	196.28	416.30	11.93
Shell2	-2367.70	299.07	-1522.91	192.94	2704.39	285.34	146.40	15.91
Bulk	-124269.94	140.03	-81707.28	97.01	135537.27	426.97	7846.30	21.69
Na_aP								
DNA	-5463.59	20.56	-4542.43	12.55	2490.14	20.05	1.00	0.00
IONS	-313.62	1.56	-174.98	0.40	453.34	12.44	35.00	0.00
Shell1	-6603.78	218.38	-4272.96	137.85	7416.40	220.21	433.40	11.60
Shell2	-2559.88	164.15	-1599.50	104.83	3080.72	173.99	162.10	9.19
Bulk	-119427.23	149.39	-77127.64	119.36	134656.56	342.84	7813.50	12.48
Na_P								
DNA	-5456.87	18.27	-4535.44	10.50	2476.98	19.30	1.00	0.00
IONS	-311.70	1.95	-174.71	0.35	447.88	13.16	35.00	0.00
Shell1	-6469.17	244.24	-4186.42	156.97	7260.00	254.77	425.33	13.49
Shell2	-2576.48	132.79	-1608.18	84.99	3106.25	136.93	163.22	7.39
Bulk	-119598.68	133.76	-77077.50	100.91	135377.60	202.46	7820.44	14.18
Na_T								
DNA	-5461.70	17.64	-4538.55	11.70	2493.89	30.28	1.00	0.00
IONS	-311.30	1.46	-174.84	0.41	445.86	11.52	35.00	0.00
Shell1	-6562.86	223.27	-4248.61	142.85	7361.80	236.44	431.30	12.23
Shell2	-2622.27	141.76	-1637.82	90.84	3157.99	152.33	166.10	7.92
Bulk	-119380.18	200.62	-77050.67	142.66	134755.79	320.60	7811.60	11.93

SII. Supplementary Figures

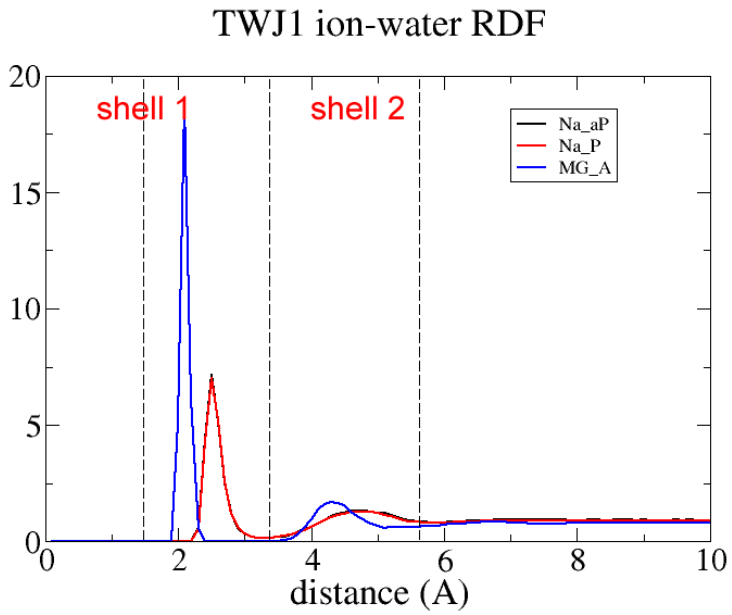


Figure S1: Ion-water RDFs of Na_aP (black), Na_P (red) and MG_A (blue). The water molecules are tightly coordinated to the Mg^{2+} (max at 2.1Å) ion with a coordination number of 5. The Na^{+} ions has a coordination number of 6.25, with the water molecules more loosely coordinated (max at 2.5Å)

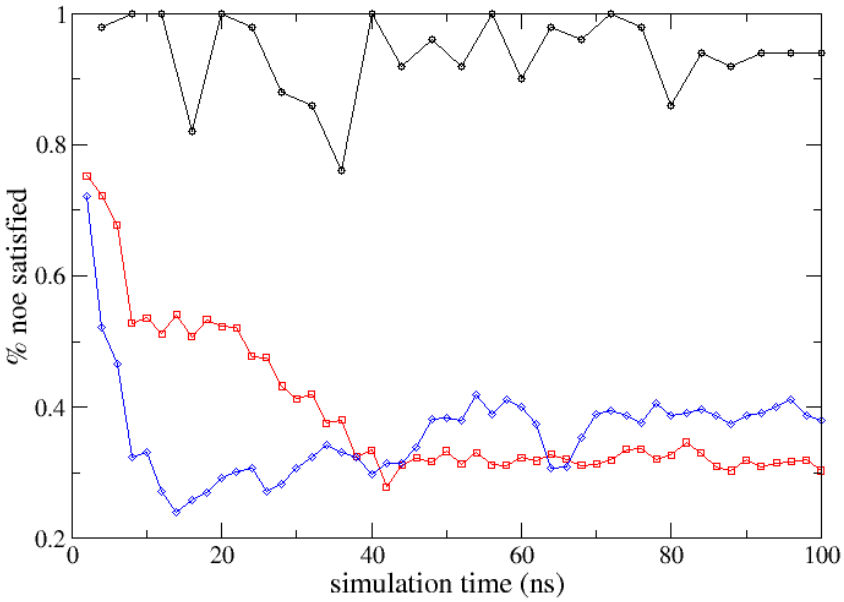
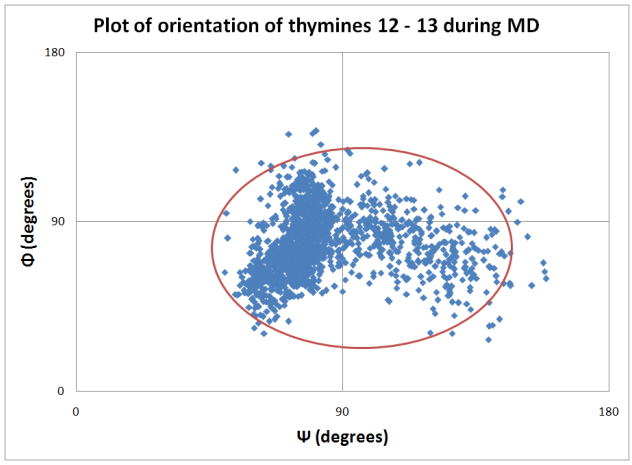
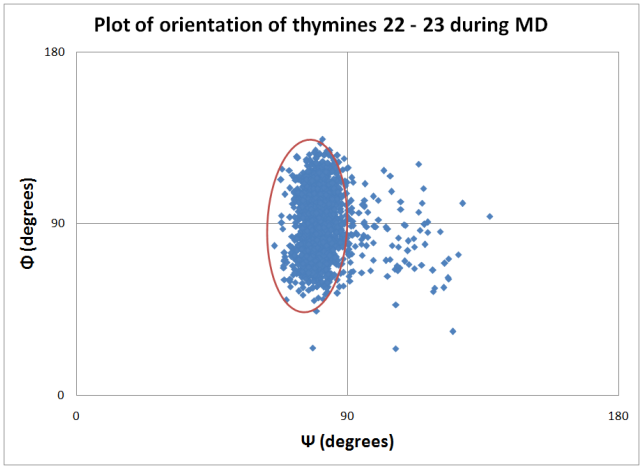


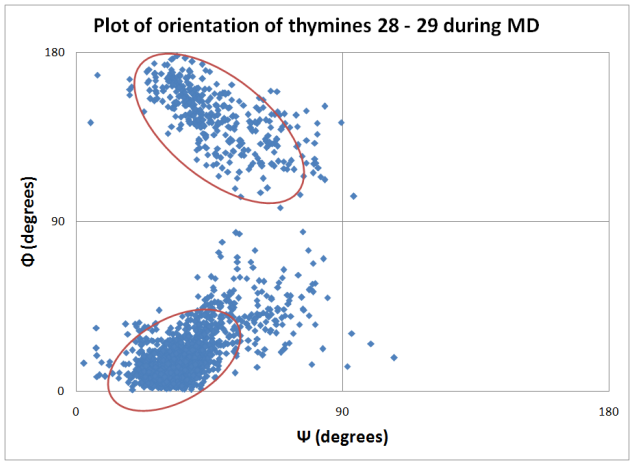
Figure S2: Percent of experimental NMR NOEs satisfied during TWJ1_Na MD (red) and MG_A MD (blue). The 26 NMR structures (black) are added as a reference purposes. The Na_aP conformer is taken as the snapshots from 5 – 25 ns of TWJ1_Na MD, while Na_P and MG_A conformers are taken from snapshots 60 – 100 ns of TWJ1_Na and TWJ1_MG_A MD, respectively.



a



b



c

Figure S3: Φ, Ø plot of the configurational orientations of the thymines capping helix 2 and the thymines at the branched region for TWJ1 in Na^+. Data is of all snapshots represented by Conformer P and Conformer T. The red circles indicate the range with the largest density.

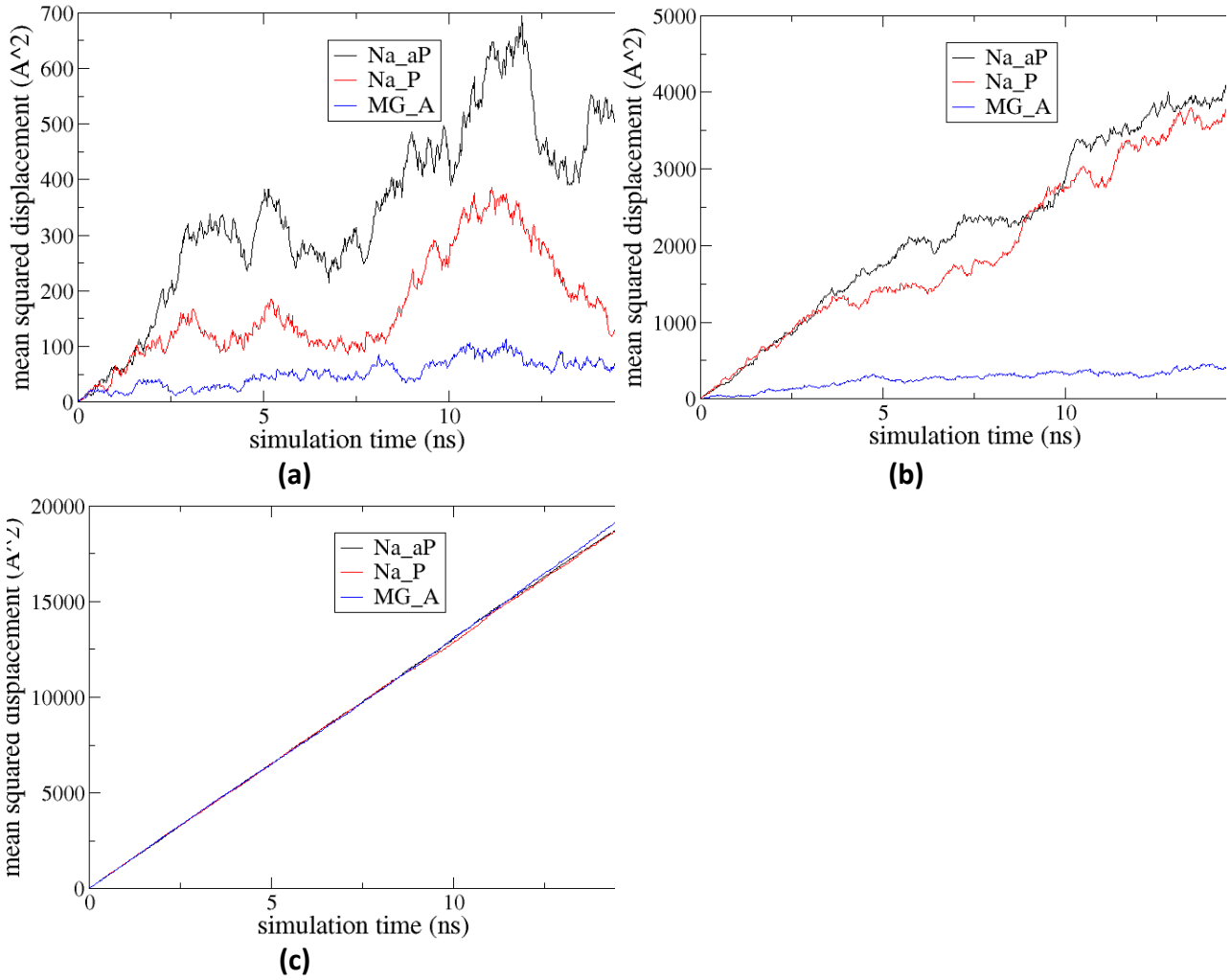


Figure S4: a) DNA b) Ion and c) bulk water diffusion profiles for Na_aP (black) Na_P (red) and MG_A (blue) during dynamics. For each conformer, the diffusion constants are calculated over a 15 ns window during dynamics: Na_aP: 5 – 20 ns of TWJ1_Na MD, Na_P: 85 – 100 ns of TWJ1_Na MD, and MG_A: 85 – 100 ns of TWJ1_Na MD.

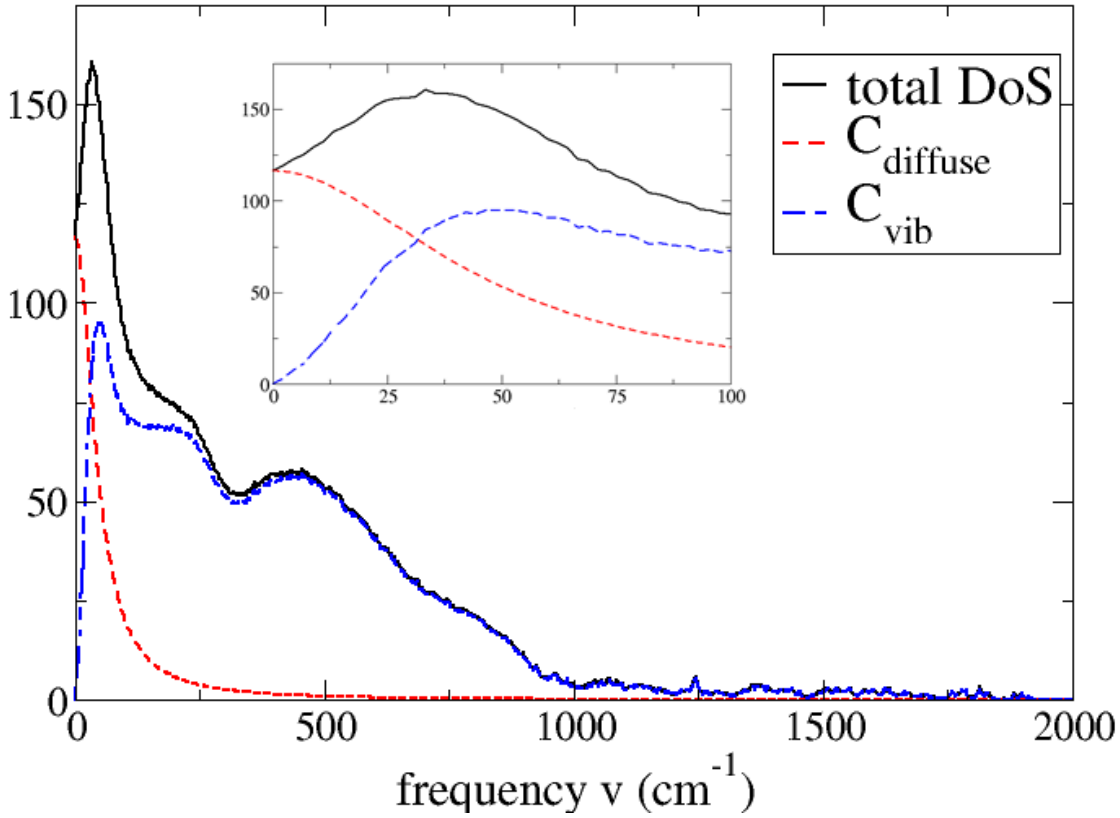
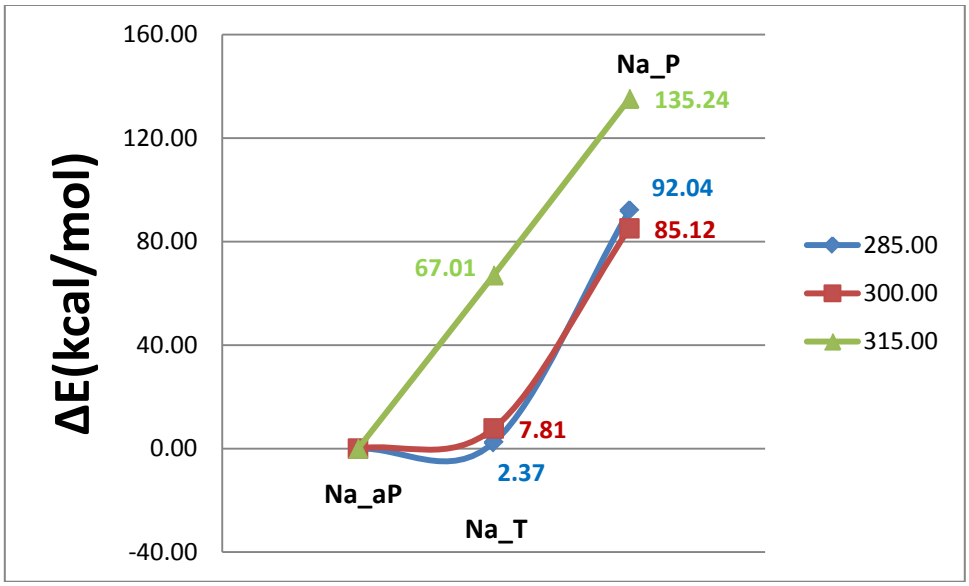
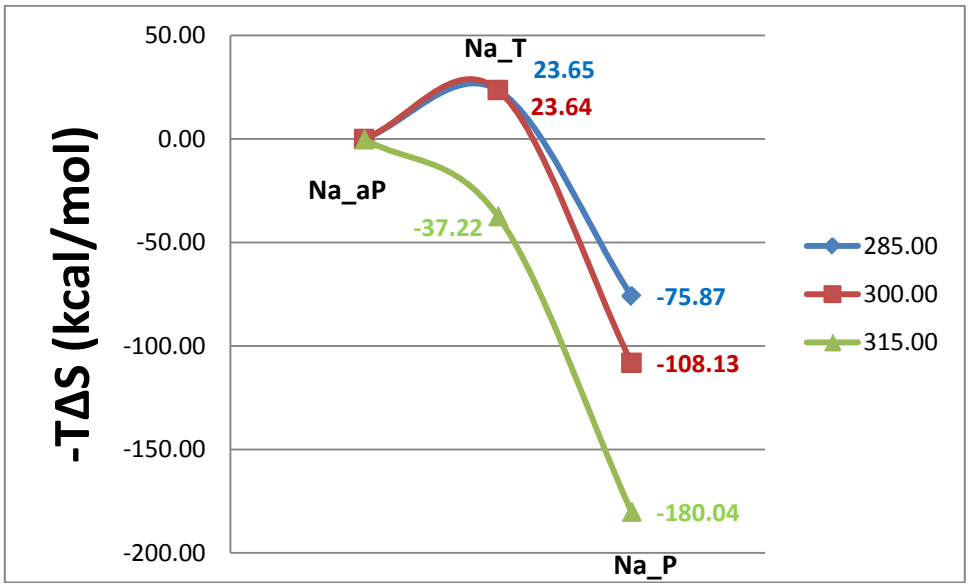


Figure S5: Representative density of state (DoS) for TWJ1. The DoS is the Fourier Transform of the velocity auto-correlation function (VAC). The two components of the total DoS (the diffusive contribution $C_{diffuse}$ modeled as a hard sphere and the vibrational contribution C_{vib} modeled as a Debye solid) are also shown. Note that the hard sphere contribution is the only contribution at the zero frequency mode and exponentially decays to zero. This is emphasized by considering the low frequency modes as shown in the inset.



a



b

Figure S6: Components of Free Energy for Na_aP, Na_T and Na_P conformers of TWJ1_Na: a) Zero point corrected enthalpy b) –TS (kcal/mol).

Chapter 2. Interaction between OmpA and Ecgp-96: Role in Bacterial Meningitis

Neonatal bacterial meningitis is the most serious infection of the central nervous system (CNS) resulting in significant neurological sequelae in half of the survivors, such as hearing loss, convulsive disorders, abnormal speech patterns, cortical blindness and mental retardation[1-4]. The incidence of bacterial meningitis in infants is about 5 cases per 100,000 live births per year in developed countries[5] but may be 10 times higher in underdeveloped countries. *Escherichia coli* K1 is a predominant pathogen causing neonatal meningitis and septicemia[6]. The mortality and morbidity due to *E. coli* K1 has remained significant for last few decades despite the use of effective antimicrobial therapy. This poor outcome is due to the lack of therapeutic strategies for preventing this deadly disease. Previous studies in the newborn rat model of hemotogenous meningitis demonstrated that a high degree of bacteremia is required for the onset of meningitis. The survival and entry of *E. coli* K1 into human brain microvascular endothelial cells (HBMEC), which comprises the blood-brain barrier, has been shown to require the expression of outer membrane protein A (OmpA) in *E. coli* K1. In addition, OmpA expression is known to be important to entry of monocytes, macrophages, and dendritic cells, indicating the important role of this protein in the pathogenesis of neonatal meningitis by *E. coli* K1[7-10].

OmpA is a 35 kDa membrane protein of *E. coli* K1, containing eight transmembrane domains and four extracellular loops[11]. The N-terminal 177 amino acids form the transmembrane region of the protein, while the remainder of the C-terminal portion is buried in the periplasmic area of the bacterial cell wall. OmpA interacts with the Ec-gp96 homologue of the gp96 receptor on the HBMEC for invasion of HBMEC by *E. coli* K1[12-13]. Gp96 (also referred to as GRP94) is expressed ubiquitously, but its expression is significantly lower in human umbilical

vein endothelial cells and other non-brain endothelial cells compared to HBMEC[12]. Gp96 contains an N-terminal signal sequence and a C-terminal KDEL sequence for ER retention, making it an unusual member of the Hsp90 protein family. Furthermore, OmpA interacts initially with GlcNAc1-4GlcNAc (chitobiose) epitopes of Ec-gp96 followed by the peptide portion of the receptor[12]. Ec-gp96 contains two N-glycosylation sites at the N-terminal region whose GlcNc1-4GlcNAc moieties might interact with OmpA during the invasion process.

In 2003 we reported computer simulation studies of chitobiose epitopes interacting with OmpA, predicting that the sugar moieties interact with OmpA in two different regions[14] (referred to here as L and B). However, these earlier calculations did not consider the lipid membrane bilayer and ignored the solvent (estimating solvation effects using an implicit model, based on continuum electrostatic theory, with no explicit water molecules). Moreover the predicted structures were not subjected to molecular dynamics (MD) to ensure proper sampling of various protein-ligand conformations. In addition, no attempt was made to account for entropy or quantum (zero-point energy) effects.

The current studies remove most of these restrictions while utilizing dramatically improved ligand docking procedures, which are expected to provide much more accurate computational predictions. Here we compare the new predictions with mutation and binding experiments aimed at assessing the accuracy of the predictions. Thus we used the new GenDock docking technique (aimed at a very complete sampling of docking poses) to predict the binding site of chitobiose to OmpA and 9 mutants selected to test the accuracy of the predicted structures. We found the same two L and B binding regions as before. Then we then carried out molecular dynamics (MD) simulations including explicit membrane, explicit solvent,

and counterions, independently for each of the two binding regions for each of the 10 ligand/ protein complexes. Here the simulation cell consisted of the ligand/OmpA complex embedded and centered in a 1-palmitoyl-2-oleoyl-phosphatidylcholine (POPC) membrane (a total of 45303 atoms for WT).

Most important, the results reported here use new approach to extract directly from the MD simulations the entropies and quantum effects (zero-point energies) required to calculate the free energy of binding. Entropy is generally believed to play a critical factor in protein folding[15-16], in binding of various proteins to DNA,[17] and in the hydrophobic effect.[18] However, there has been no practical but accurate approach to calculate the entropies of such complex systems as ligands interacting with proteins in a membrane with explicit solvent. Here we use our recently developed two-phase thermodynamics (2PT)[19] method to extract the entropies, quantum effects (zero-point energies), and free energies directly from the velocity autocorrelation function derived from short (20 ps) MD simulations.

The 2PT method was previously validated by calculating the entropy of water in different domains of PAMAM dendrimers[20-21], determining various phases of dendrimer liquid crystals[22], and calculating the relative stability of various aggregates[23]. Recently, Jana et al[24] used 2PT to compute the entropies of water molecules in both grooves of DNA, showing that they were significantly lower than that in bulk water. However the current work is the first attempt to obtain free energies of binding for ligand/protein complexes.

Our goal in the paper is to determine whether all these improvements in the methodologies affect the two site (L and B) model for chitobiose/OmpA binding and whether the computed binding free energies correlate sufficient well to experimental invasion efficiency

that the computations could be used to identify new lead therapeutics for neonatal meningitis. To validate this 2PT-FE approach, we predicted the free energies for the interaction of 9 mutated OmpA proteins with Ec-gp96 and compare to new experiments reported here of invasion inhibition involving exactly the same mutations. We find an excellent correlation of the predicted binding free energy with the experimental invasion efficacy, suggesting that *in-silico* design could be used to select from a large number of ligand variations the most promising ligands for experimental tests, a procedure that should accelerate the development of improved therapeutics.

I. Results

a) Conversion of small stretches (three to four) of amino acids to alanines in loops 1 and 4 of OmpA significantly reduce the invasion of E. coli K1 in HBMEC.

Previous studies using synthetic peptides representing portions of loops 1 and 2 of OmpA showed that these loops might be involved in *E. coli* K1 interaction of HBMEC for invasion[7]. Therefore, we carried out 9 sets of mutation experiments in each of which two to four amino acids were mutated to alanines in various loops of OmpA thought from our previous computations to be possible candidates for participating as extracellular loops in the invasion process. We aimed to develop a series of mutations in all four loops, but some mutations produced lethal consequences to the bacteria. The mutations generated in various OmpA loops are shown in **Figure 1** using various primer pairs (Table 1, supplementary materials (SM)). The growth characteristics of these mutants were all similar to that of wild type (WT) *E. coli* K1. In addition, these mutants all express similar amounts of OmpA as examined by Western blotting with an anti-OmpA antibody (Figure S1 SM).

b) The stability of the mutated OmpA structures

The AMBER99SB protein forcefield has been shown to provide a good description of secondary structure elements for a large set of experimental structures[25]. An important question in such mutation experiments is whether our selected mutations would alter the OmpA fold (thereby destroying protein function). We tested this computationally, by performing 10ns of MD on the WT OmpA and on all nine mutants, without chitobiose bound. We characterized the effect of the mutations by analyzing the simulations to determine the changes in three characteristics:

1. the Cartesian Root Mean Square (CRMS) change in the structure
2. the Solvent Accessible Surface Area (SASA), a measure of the overall size exposed to solvent and
3. the Radius of gyration (R_{gyr}), a measure of the overall size of the protein (table S2 of SM).

We find that none of the 9 sets of mutations led to a significant change in the structure, with the changes from the starting structure converging after 2ns. We found an average CRMS change of 1.6±0.4Å. The largest deviation was for the 1b mutants (2.19Å) and the smallest for the 2b mutant (1.16Å).

Further, we find that the R_{gyr} of the mutants are similar to that of the WT (18.4±0.2Å), with Rgyr ranging from 18.4Å for 1b, 1c and 4c, to 18.9Å for 4a. In addition we find that the SASA ranges from 1% larger (4b) to 6% larger (2b) than the calculated WT value of 11196.9±103.1Å^2, (the NMR structure has SASA = 11043.1Å^2) (Table S2 of SM).

That the simulations find only small changes from WT gives us confidence in the accuracy of our methods and in the integrity of the mutated structures in mimicking the WT OmpA.

Figure 2a shows a schematic of the WT OmpA with chitobiose bound to both the L and B sites. We observed convergence to the equilibrated structure after 3.2 ns (Figure S3 of SM). Most of the changes from the starting structure occurred in the loop region (CRMS = 3.30Å), with the CRMS of the B (barrel) binding site region < 0.5Å, while the CRMS of the L (loop) binding site regions are > 2.0Å. In particular loops 1, 2 and 3 (region L) fluctuate more than loop 4 (region B) during the MD.

c) Chitobiose binding in loop region L differs across mutants.

Our previous studies reported two important binding sites for the interaction of chitobiose with OmpA[14]: one site being in the 1-2-3 loop regions (region L for loop) and the other closer to the barrel of the OmpA and loop 4 (region B for barrel) (**Figure 2a**). Despite major improvements in our ligand docking procedure[26-27], we obtained binding sites quite similar to the previous predictions (Figure S2 of SM), with the binding to region B consistently stronger than to region L (table 1), as predicted previously.

Indeed the MD studies lead to differences of only 0.82±0.04Å CRMS for the chitobiose bound to region L and 0.79±0.02Å CRMS for region B (compared to the starting positions). However the new predicted binding sites lead to vastly improved interaction energies of chitobiose with the protein residues (c.f. -85.6 kcal/mol vs -15 kcal/mol for regions L and -148.8 vs -20 kcal/mol for region B, table 1).

The binding is dominated by electrostatic energies (based on the RESP[28-29] scheme in this study, but on the Mulliken[30] scheme in our earlier study). The long range nature of the electrostatic interactions can lead to interaction energies that are highly sensitive to structural details remote from the binding regions. Thus, in our studies of ligands binding to G-Protein Coupled Receptors (GPCR), we showed that special approaches to neutralize the charged residues lead to dramatically better comparisons with experimental Kd values[31]. The studies reported here use the more traditional approaches but lead to very consistent interaction energies.

We find that chitobiose in site L makes strong hydrogen bonds (HB) with Asn26 (-2.5 kcal/mol) and Asn27/28 (-4.0 kcal/mol), as reported previously.[14] In addition we also find HBs to Arg61 (-3.2 kcal/mol) and a somewhat weaker HB to Asn70 (-1.0 kcal/mol). Thus we consider these four interactions to be critical for binding.

As expected the loss of the HB for the Asn28 → Ala mutation in the 1a mutant (N28A-G29A-P30A) dramatically decreases the interaction enthalpy for the L site (by 83%) but had no effect on the binding in site B. The interaction energy of the 1b mutant (T31A-H32A-E33A) decreases by 84% (table 2), which we explain as due primarily to the loss of the favorable interaction with Glu33 (-7.67 kcal/mol). These two cases with the largest predicted reduction in binding are observed experimentally to prevent invasion. For both mutants we find that the chitobiose became unbound during the 10ns of MD. Thus the calculated binding energy (< 20% of the WT) represents a lower bound on chitobiose interaction energy required for function.

The predicted decreases in binding for the 1c (Y19A-H20A-D21A), 2b(P63A-Y64A-K65A) and 3a(N110A-V111A-Y112A) mutants site L also correlate with reduced invasion efficiencies

(by 62%, 73% and 59% respectively). The 2a(G66A-S67A-V68A) mutant has a binding energy close to the WT (86 %), due to the strong, primarily electrostatic interaction with Glu69 (-21.9 kcal/mol). On the other hand for the loop 4 mutants, we expected no significant changes in the binding energies for region L, which indeed we find (changes of 6%).

For binding of chitobiose to region B, we find four important HB in the wild-type structure: Met101, Ser121, Thr145 and Thr153. Thus as expected the 1a-c, 2a, 2b and 3a mutants all have binding energies similar to WT (from 86% to 105%), while the loop 4 mutants all show substantially decreased binding [mutant 4a(I148A-G149A-D150A) -36%, 4b(H152A-T153A-I154A) -26%, and 4c(P158-D159-N160) -38%] (Table S3 of SM). Of the four HB found in the wild-type structure, only Ser121 is found in all the 4a-c mutants. The remaining three are entirely missing in the 4c mutant, and the chitobiose forms a weak hydrogen bond with Thr124 (~ 1.5 kcal/mol) in the 4a and 4b mutants.

d) The Free energy of chitobiose binding

For WT OmpA and the 9 OmpA mutants we calculated the MD of the OmpA/chitobiose complex for 10ns, of which the first 3 ns led to equilibration. Then, we carried out the 2PT analysis for a 20 ps segment every 2ns, to obtain the free energy at that time. These 5 independent calculations are then averaged to estimate the statistical fluctuations.

For the WT MD, the fluctuations over the entire 10ns of dynamics are 0.21% (free energy, A), 0.11% (enthalpy, H), and 0.12% (entropy, S). For the 1a mutant, we obtain fluctuations of 1.12% in A, 0.59% in H, and 1.34% in S. These values are tabulated in Table S4 of SM. We find that the fluctuations are much smaller than the differences between the various mutants justifying the comparison with experiment.

The free energy as a function of time is illustrated in **Figure 3** for two extremes cases:

- WT OmpA/chitobiose complex (region L chitobiose bound during the entire dynamics) (**Figure 3a**)
- the 1b mutant (region L chitobiose becoming unbound after 6ns) (**Figure 3b**).

We observe a reorientation of the chitobiose unit in region L of the 1b mutant after 4ns of MD. This reorientation allows the chitobiose unit to become more solvent exposed, reducing the free energy by 1.3%. The enthalpy is reduced (by 1.0%) due to the weak interactions between the chitobiose and OmpA and the entropy is increased by 0.4%, since the chitobiose unit is more mobile. The major contributor to the increased entropy is the chitobiose unit, and to a lesser extent the atoms in region L. By 6ns (immediately before the chitobiose unit is ejected), the free energy is further reduced, and is 2.6% higher (more positive) than the value at 2ns. Here we see reductions in both the enthalpy (1.1%) and entropy (1.0%).

The free energy increases after the chitobiose becomes unbound, ultimately ending 0.15% lower than the value at 2ns. Here, the enthalpy is reduced by 0.5% presumably since the chitobiose is not longer interacting favorably with the protein. There is however an increase in the total system entropy of 1.6% due again to the unbound, mobile chitobiose unit. Of note is that the 0.5% difference in the enthalpy of the bound and unbound states is lower than the calculated binding energy (gas-phase) of the chitobiose in region L (1.3% of the total enthalpy). This indicates that while the enthalpy of the unbound state becomes less attractive as the chitobiose becomes unbound, the magnitude of this change is mitigated by the increased interactions between the hydrophilic chitobiose unit and nearby water molecules (the solvation effect). The 1a mutant shows a similar profile (reduction in free energy of 1.2% after 4ns, further increase to 1.5% at 6ns and a final increase in the free energy of 0.04 % after 10ns). The

profile of these two mutants is unique from the rest, with the entropy dramatically switching signs during the transition. This indicates that the 2PT-FE-MD method is sensitive to structural changes in the system under consideration.

Summarizing, we observe that the system acts to minimize its free energy, decreasing a component of its free energy for a larger compensatory increase in the other. The calculated changes in the free energy are smaller than would be expected by the changes in the calculated enthalpy only, providing direct computational evidence for the entropy-enthalpy compensation[32-34] effect.

All mutants have a reduced enthalpy compared to the WT, with the largest difference seen in the 1b mutant (3.1%) and the smallest difference in the 2a mutant (0.4%) (table S5 of SM). Interestingly, the correspondence in the entropy is not as straightforward. We would expect the entropy to increase (be more favorable) for each of the mutants relative to the WT, as the mutation of bulky group to alanines would increase the available space (and mobility) of the chitobiose unit. This is indeed the case for 2a (0.32%), 4b (0.77%) and 4c (0.57%) mutants, however the entropy of the remaining mutants is lower than the WT, from 1b (-2.35%) to 4a (-0.63%). In the case of the 1a and 1b mutants, we obtained statistics prior to the chitobiose unit becoming unbound.

e) Comparison between theory/experiment and theoretical prediction of *E. coli* invasion.

Invasion assays using the OmpA mutant strains revealed that

- mutations 1a and 1b completely abrogated the invasion of *E. coli* K1 in HBMEC, (the measured value for 1a of 1.5% is zero within experimental uncertainty)

- mutations 1c and 2b reduced the invasion by 60% (Table 3).
- mutations 2a and 3a had very little effect on the invasion.
- loop 4 mutations reduced the invasion by 60-80%.

These results suggest that regions 1a and 1b are very critical while region 2b and loop 4 also contribute significantly to the interaction. These experimental results correlate well (92%) with our predicted free energy changes (Figure 4a and table S5 of SM). Ignoring entropy (considering just enthalpies) the correlation drops to 79%.

We find that the entropic contributions are important for accurately predicting the invasion of the 1b, 4b, 1c, 2a and 3a mutants. Using only the enthalpies, the 1b mutant would be predicted to equally effective in inhibiting invasion (6.3%) as the 4b mutant (4.6%), while experimentally we know that the 1b mutant completely blocks invasion while the 4b mutant has a 20% invasion rate. Further, using only enthalpy, the 1c mutant (77%) would be predicted to be as effective as the 2a mutant (73%), while the experimental results are 43% and 82% respectively. These discrepancies are eliminated by including the entropy to predict free energies: 0%, 22%, 60% and 83% for the 1b, 4b, 1c and 2a mutants respectively.

The MD simulation for the 1a and 1b mutants led to a loss of binding to the OmpA and indeed the experiments indicated no invasion for these systems. This is consistent with the idea that there is a binding threshold below which there is no invasion. Thus in fitting the correlation discussed above we do not require that the binding energy correlated linearly with invasion: the 1a and 1b mutants are not considered in the fit. This threshold is predicted to be 94% of the WT free energy and 96% of the enthalpy. We suggest that these limits can provide an energy criterion for rapid screening to find small molecules that may prevent invasion.

Using the correlation with binding free energy allows a prediction of the experimental invasion for new mutants to ±4% on average. The high correlation between the experimental and predicted invasions gives us confidence to make a prediction of the effects of a mutation (2c) not yet determined experimentally. For the 2c mutant, our predicted binding energy leads to a predicted invasion rate of 23.4 ± 1.7% (problems in carrying out the experiments prevented us from measuring this case).

Although the current study considers binding of only the chitobiose ligand, we would hope that the minimum threshold in binding for invasion might also apply to other ligands, providing an energy criterion for rapid screening to find lead molecules most likely to prevent invasion.

f) OmpA interacts with Ec-gp96 by cooperative binding – the LbD model.

Our experiments showed that *E. coli* invasion can be blocked by the 1a and 1b selective mutations in the region L chitobiose binding site of OmpA. Computationally, we find that these two mutants lead to unbinding of chitobiose during dynamics, in contrast to all other region L mutants (1c, 2a, 2b, 1c and 3a). Mutations in region B (loop 4) reduce binding by at most 65 – 85%.

Since region B is not as solvent exposed as region L, we propose a new model for the interaction of OmpA on *E. coli* and HBMEC protein Ec-gp96, an etiological factor in the pathogenesis of bacterial meningitis in neonates. In this Loop-bi-Dentate (LbD) model, the solvent exposed region L binding site on OmpA acts as the primary recognition element for the sugar moieties present in the glycosylation sites on Ec-gp96. After recognition, there is a subsequent binding event to region B, with binding to both regions being necessary for

invasion. Analysis of the Ec-gp96 sequence (there is no crystal structure) reveals two possible glycosylation sites (residues 142 – 145: NASD and 251 – 254: NDSQ) with geometric dimensions consistent with the OmpA loop structure.

II. Discussion

OmpA has been generally described as a structural protein with porin activity, however studies from the Prasadarao lab have shown that OmpA binds to HBMEC via Ec-gp96 during *E. coli* invasion. We investigated these interactions computationally with all atom MD simulations, including explicit treatment of the membrane and solvent, leading to three significant conclusions that support the experimental data.

First, we show that the initial configuration of OmpA incorporated into the membrane environment is sufficient to generate a model in which the binding of chitobiose moieties correlates with experimental invasion efficiency. The 92% correlation of the calculated binding free energy of the OmpA/chitobiose complex with the experimentally observed invasion rate mediated by OmpA and nine of its mutants supports the hypothesis that *E. coli* meningitis is facilitated by OmpA interacting with the glycosylation sites on Ec-gp96.

It might have been expected that large mutations to the OmpA structure would result in misfolded structures, destroying protein function. However, we showed computationally that these specific three to four residue mutations to the loop regions maintain the structure fidelity of OmpA, despite the dramatic alteration in the system thermodynamics. This result is particularly dramatic in the 1a and 1b mutants, where we find similar structural fluctuations as in the WT, but huge changes in the calculated free energy with chitobiose becoming unbound after 6 ns of MD. This explains the experimentally observed loss of invasion.

Interactions in the buried region B are predicted to be much stronger than the solvent exposed region L. Strong HB interactions are removed in region B by the 4b and 4c mutants, but the chitobiose remains bound during dynamics. Both the calculated free energy and the experimentally observed invasion rates are reduced, suggesting that the strength of binding of chitobiose in region B is important to invasion.

Second, we propose the LbD model that explains the mechanism of *E. coli* invasion in terms of our theoretical findings. The LbD model assumes that the first step is for the loop region L on OmpA to recognize and bind to a glycosylation site on Ec-gp96. This is followed by cooperative binding of loop region B on OmpA to the second glycosylation site. The experimental result that mutants 1a and 1b of loop 1 prevent invasion is consistent with our previous studies in which synthetic peptides representing the loop 1 regions significantly blocked invasion[7]. Since a very few amino acids in OmpA are critical for interacting with Ec-gp96 to invade HBMEC, we suggest that targeting these areas for inhibition by small molecules might lead to therapeutics that prevent neonatal meningitis.

Finally, we find that the 2PT Free Energy (2PT-FE) method for quantum corrected MD leads to an excellent correlation with invasion efficiency, substantially better than using standard MD energies. This correlation indicates that the OmpA – chitobiose interactions are critical components in the interactions necessary for invasion. This study represents, to our knowledge, the first complete theoretical characterization of protein-protein interactions that relates directly to a transition in disease dynamics.

Since a very few amino acids in OmpA are critical for interacting with Ec-gp96 to invade HBMEC, targeting those areas for inhibition by small molecules would be a potential therapeutic strategy to preventing the neonatal meningitis.

III. Materials and Methods

a) Cell culture and invasion assays

Human brain capillaries were isolated from small fragments of cerebral cortex, which were obtained from surgical resections of 4- to 7-year-old children with seizure disorders at Children's Hospital Los Angeles. Microvascular endothelial cells (HBMEC) were isolated from these capillaries and cultured as described previously[7]. HBMEC were maintained at 37 °C in a humidified atmosphere of 5% CO_2 in a medium containing M-199/Ham F-12 (1:1 vol/vol) supplemented with 10% fetal bovine serum, sodium pyruvate, and 2 mM glutamine and then cultivated in a cell culture incubator at 5% CO_2 and 37°C. HBMEC were used between 12 and 16 passages for all experiments. For invasion assays, HBMEC grown in 24-well cell culture plates to 95% confluence were infected with 10^7 cfu of *E. coli* strains in experimental medium (1:1 mixture of Ham F-12 and M-199 containing 5% heat inactivated fetal bovine serum) and incubated for 90 min at 37°C in an atmosphere containing 5% CO_2. The monolayers were washed three times with RPMI 1640 medium followed by addition of gentamicin (100 μg/ml) and further incubated for 1 h at 37 °C. The cells were then washed three times with RPMI 1640 and lysed with 0.5% of Triton X-100. The released bacteria were diluted with saline and enumerated by plating on blood agar. Results were expressed as an average of four independent determinations ± S.D. of the mean.

b) Generation of OmpA mutants in *E. coli*

Primer pairs (Table 1 of M) containing the desired mutations were synthesized and cartridge-purified. PCR for the plasmid mutation was carried out using *PfuTurbo* DNA polymerase and pKE325 as the template, which is a 4-kb plasmid containing the complete *E. coli ompA* gene. The reaction mix was heated at 96^{o}C for 90 sec before the enzyme was added, and followed by 18 cycles at 95^{o}C for 30 sec, 55^{o}C for 1 min, 68^{o}C for 5 min, and a 5 min extension at the end. A faint band was visualized on agarose gel if the amplification succeeded. DNA was then digested with *Dpn*I and transformed into *E. coli* DH5α, selecting for kanamycin resistance. Plasmids were isolated and the mutation verified by sequencing. The correct plasmids were introduced into the Δ*ompA* strain E98 and the growth pattern of the bacteria was examined. The expression of OmpA in each strain was examined by Western blotting using an anti-OmpA antibody.

c) Molecular Dynamics simulations of OmpA/chitobiose complex

The structure for a single OmpA molecule was extracted from the crystal structure (1.65Å)[35] and equilibrated in a box of 3000 waters (see SM for details). Then we used the GenDock method to predict the binding site and structure for chitobiose to WT OmpA (see SM for details) for both the L and B regions. For each of the 9 mutant structures we carried out the mutation for the optimized OmpA structure, modified the side chains near the mutation using the SCREAM side chain optimization method, and equilibrated in vacuum using the Poisson-Boltzmann solvation method in the APBS software package[36]. Then we used the GenDock method to predict the binding site and structure for chitobiose to both the L and B regions for each of the nine mutants.

Then each of the two chitobiose/protein complexes (bound to the B and L regions) for OmpA and the 9 mutants were each centered in a pre-equilibrated box of solvated POPC, removing overlapping membrane/solvent molecules as described in the SM. Each system was then neutralized by adding an appropriate number of Na^+ or Cl^- replacing waters with the best electrostatic potential. In all cases the initial system size was 75Å x 75Å x 80Å (with the POPC bilayer membrane centered at z=0. The chitobiose and POPC molecules were all described using the GAFF[37] forcefield. We re-minimized the entire simulation cell (rms force tolerance of $1x10^{-4}$ kcal/mol), assigned Maxwell-Boltzmann velocities at 50K and then heated the system to 310K over 100ps of NVT dynamics, consistent with our previous studies[38-40]. Each of the WT and mutant systems was then equilibrated for 5ns MD at 300K using a Nose-Hoover thermostat[41-44] with a temperature-coupling constant of 100 fs. We then carried out NPT dynamics at a pressure of 1 bar using the Anderson barostat[45] with a 2.0 ps damping constant, requiring that all three dimensions be dilated or contracted together (isotropic pressure) . The average volume of the cell during the last 500 ps of MD was calculated and the snapshot of the system closest to this average volume was selected. This "best volume" structure was then simulated for a further 5 ns of MD at constant volume constant temperature (NVT).

IV. Acknowledgements

This project was partially supported by grants to Caltech from NSF (CMMI-072870, CTS-0608889) and by grants to USC-Children's Hospital (NIH AI 40567). The computational facilities were provided by ARO-DURIP and ONR-DURIP. TAP thanks the DOE CSGF for a graduate fellowship.

V. References

(1) Unhanand, M.; Mustafa, M. M.; McCracken, G. H.; Nelson, J. D. *Journal of Pediatrics* **1993**, *122*, 15.

(2) Gladstone, I. M.; Ehrenkranz, R. A.; Edberg, S. C.; Baltimore, R. S. *Pediatric Infectious Disease Journal* **1990**, *9*, 819.

(3) Koedel, U.; Pfister, H. W. *Infectious Disease Clinics of North America* **1999**, *13*, 549.

(4) Leib, S. L.; Tauber, M. G. *Infectious Disease Clinics of North America* **1999**, *13*, 527.

(5) Taziarova, M.; Holeckova, K.; Lesnakova, A.; Sladeckova, V.; Bartkovjak, M.; Seckova, S.; Bukovinova, P.; Hvizdak, F.; Svabova, V.; Findova, L.; Kisac, P.; Beno, P.; Bauer, F.; Bauer, M.; Karvai, M.; Rudinsky, B.; Sabo, I.; Bielova, M.; Luzica, R.; Wiczmandyova, D.; Huttova, M.; Ondrusova, A. *Neuroendocrinology Letters* **2007**, *28*, 18.

(6) Stoll, B. J.; Hansen, N.; Fanaroff, A. A.; Wright, L. L.; Carlo, W. A.; Ehrenkranz, R. A.; Lemons, J. A.; Donovan, E. F.; Stark, A. R.; Tyson, J. E.; Oh, W.; Bauer, C. R.; Korones, S. B.; Shankaran, S.; Laptook, A. R.; Stevenson, D. K.; Papile, L. A.; Poole, W. K. *New England Journal of Medicine* **2002**, *347*, 240.

(7) Prasadarao, N. V.; Wass, C. A.; Weiser, J. N.; Stins, M. F.; Huang, S. H.; Kim, K. S. *Infection and Immunity* **1996**, *64*, 146.

(8) Mittal, R.; Prasadarao, N. V. *Journal of Immunology* **2008**, *181*, 2672.

(9) Sukumaran, S. K.; Shimada, H.; Prasadarao, N. V. *Infection and Immunity* **2003**, *71*, 5951.

(10) Altmeyer, A.; Maki, R. G.; Feldweg, A. M.; Heike, M.; Protopopov, V. P.; Masur, S. K.; Srivastava, P. K. *International Journal of Cancer* **1996**, *69*, 340.

(11) Prasadarao, N. V.; Wass, C. A.; Kim, K. S. *Infection and Immunity* **1996**, *64*, 154.

(12) Prasadarao, N. V.; Srivastava, P. K.; Rudrabhatla, R. S.; Kim, K. S.; Huang, S. H.; Sukumaran, S. K. *Infection and Immunity* **2003**, *71*, 1680.

(13) Prasadarao, N. V. *Infection and Immunity* **2002**, *70*, 4556.

(14) Datta, D.; Vaidehi, N.; Floriano, W. B.; Kim, K. S.; Prasadarao, N. V.; Goddard, W. A. *Proteins* **2003**, *50*, 213.

(15) Privalov, P. L.; Makhatadze, G. I. *J. Mol. Biol.* **1993**, *232*, 660.

(16) Boczko, E. M.; Brooks, C. L. *Science* **1995**, *269*, 393.

(17) Spolar, R. S.; Record, M. T. *Science* **1994**, *263*, 777.

(18) Hummer, G.; Garde, S.; Garcia, A. E.; Paulaitis, M. E.; Pratt, L. R. *J. Phys. Chem. B* **1998**, *102*, 10469.

(19) Lin, S. T.; Blanco, M.; Goddard, W. A. *J. Chem. Phys.* **2003**, *119*, 11792.

(20) Lin, S. T.; Jang, S. S.; Cagin, T.; Goddard, W. A. *J. Phys. Chem. B* **2004**, *108*, 10041.

(21) Lin, S. T.; Maiti, P. K.; Goddard, W. A. *J. Phys. Chem. B* **2005**, *109*, 8663.

(22) Li, Y. Y.; Lin, S. T.; Goddard, W. A. *Journal of the American Chemical Society* **2004**, *126*, 1872.

(23) Jang, S. S.; Lin, S. T.; Maiti, P. K.; Blanco, M.; Goddard, W. A.; Shuler, P.; Tang, Y. C. *J. Phys. Chem. B* **2004**, *108*, 12130.

(24) Jana, B.; Pal, S.; Maiti, P. K.; Lin, S. T.; Hynes, J. T.; Bagchi, B. *J. Phys. Chem. B* **2006**, *110*, 19611.

(25) Viktor Hornak; Robert Abel; Asim Okur; Bentley Strockbine; Adrian Roitberg; Carlos Simmerling *Proteins: Structure, Function, and Bioinformatics* **2006**, *65*, 712.

(26) Cho, A. E.; Wendel, J. A.; Vaidehi, N.; Kekenes-Huskey, P. M.; Floriano, W. B.; Maiti, P. K.; Goddard, W. A. *J. Comput. Chem.* **2005**, *26*, 48.

(27) Kam, V. W. T.; Goddard, W. A. *Journal of Chemical Theory and Computation* **2008**, *4*, 2160.

(28) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. *The Journal of Physical Chemistry* **2002**, *97*, 10269.

(29) U. Chandra Singh; Peter A. Kollman *J. Comput. Chem.* **1984**, *5*, 129.

(30) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833.

(31) Bray, J. K.; Goddard, W. A. *Journal of Molecular Graphics & Modelling* **2008**, *27*, 66.

(32) Breslauer, K. J.; Remeta, D. P.; Chou, W. Y.; Ferrante, R.; Curry, J.; Zaunczkowski, D.; Snyder, J. G.; Marky, L. A. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84*, 8922.

(33) Gilli, P.; Ferretti, V.; Gilli, G.; Borea, P. A. *J. Phys. Chem.* **1994**, *98*, 1515.

(34) Williams, D. H.; O'Brien, D. P.; Bardsley, B. *Journal of the American Chemical Society* **2001**, *123*, 737.

(35) Pautsch, A.; Schulz, G. E. *J. Mol. Biol.* **2000**, *298*, 273.

(36) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 10037.

(37) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157.

(38) Maiti, P. K.; Pascal, T. A.; Vaidehi, N.; Goddard, W. A. *NUCLEIC ACIDS RES* **2004**, *32*, 6047.

(39) Maiti, P. K.; Pascal, T. A.; Vaidehi, N.; Heo, J.; Goddard, W. A. *Biophys. J.* **2006**, *90*, 1463.

(40) Maiti, P. K. P., Tod A.; Vaidehi, Nagarajan; Goddard, William A. *J. Nanosci. Nanotechnol.* **2007**, *7*, 1712.

(41) Hoover, W. G. *Physical Review A* **1986**, *34*, 2499.

(42) Hoover, W. G. *Physical Review A* **1985**, *31*, 1695.

(43) Nose, S. *J. Chem. Phys.* **1984**, *81*, 511.

(44) Nose, S. *Molecular Physics* **1984**, *52*, 255.

(45) Andersen, H. C. *J. Chem. Phys.* **1980**, *72*, 2384.

VI. Tables

Table 1: Cavity analysis of the chitobiose interaction with the WT OmpA in loop region L (1-2-3 loops) and region B (loop 4), ordered by binding energy. The residues on the OmpA were selected to be within 10Å of the chitobiose, and have a binding energy of at least 1 kcal/mol. The individual components of the energy (van der Waals, coulombic and hydrogen bonding) are shown in kcal/mol. The residues to be alanized in a specific mutant are labeled.

Region L						**Region B**					
Residue	**VdW**	**Coul**	**HB**	**Total**		**Residue**	**vdW**	**Coul**	**H-Bond**	**Total**	
Asn70	0.00	-7.37	-1.01	-8.38	**2c**	Thr145	-3.74	-13.82	-1.32	-18.88	
Tyr108	0.00	-8.03	-0.01	-8.04		Thr153	-2.34	-10.49	-2.31	-15.14	**4b**
Asn115	-0.02	-7.85	0.00	-7.87		Ile148	-1.96	-11.94	-0.02	-13.92	**4a**
Glu33	-0.01	-7.66	0.00	-7.67	**1b**	Trp103	-8.06	-4.56	-0.02	-12.64	
Gly29	-0.02	-7.30	0.00	-7.32	**1a**	Met101	-1.51	-2.85	-5.01	-9.37	
Lys114	-0.15	-6.21	-0.01	-6.37		Val78	0.00	-9.33	0.00	-9.33	
Ser67	-0.01	-5.80	0.00	-5.81	**2a**	Ser121	-2.51	-1.83	-4.82	-9.16	
Gly23	-0.07	-4.78	0.00	-4.85		Thr118	-0.08	-6.74	0.00	-6.82	
Asn26	-0.11	-1.88	-2.50	-4.49		Arg160	-0.09	-5.90	0.00	-5.99	**4c**
Arg61	-0.15	-0.67	-3.18	-4.00		Arg61	0.00	-5.81	0.00	-5.81	
Leu24	-1.63	-1.39	-0.01	-3.03		Asn115	0.00	-3.98	0.00	-3.98	
Lys74	0.00	-2.91	0.00	-2.91		Pro122	-0.36	-3.55	-0.01	-3.92	
Tyr112	-0.39	-2.33	0.00	-2.72	**3a**	Gly99	0.00	-3.64	0.00	-3.64	
Asn28	-0.13	1.88	-4.00	-2.25	**1a**	Arg104	-0.05	-3.57	0.00	-3.62	
Asp21	-0.02	-1.45	0.00	-1.47	**1c**	Val102	-0.12	-3.44	-0.01	-3.57	
Ser109	-1.07	-0.31	-0.01	-1.39		Val120	-0.57	-2.92	-0.01	-3.50	
Glu69	-0.15	-1.18	0.00	-1.33	**2c**	Trp144	-0.37	-2.61	-0.01	-2.99	
Pro63	-0.02	-1.10	0.00	-1.12	**2b**	Arg97	0.00	-2.34	0.00	-2.34	
						Gly77	-0.03	-2.16	0.00	-2.19	
						Gln79	-0.06	-1.73	0.00	-1.79	
						Leu80	0.00	-1.75	0.00	-1.75	
						Asp150	-0.01	-1.70	0.00	-1.71	**4a**
						Gln76	-0.04	-1.63	0.00	-1.67	
						Gly119	-0.25	-1.32	-0.01	-1.58	
						Ala38	0.00	-1.29	0.00	-1.29	
						Phe41	0.00	-1.18	0.00	-1.18	
Total[a]	**-4.17**	**-73.89**	**-10.85**	**-85.63**		**Total**[a]	**-22.16**	**-113.11**	**-13.56**	**-148.83**	

[a] The totals listed here includes contributions from residues not listed with binding energies < 1 kcal/mol

Table 2: Total interaction energy (kcal/mol) of critical residues in region L of the WT compared to the 9 mutants considered in this study (negative indicates binding). The residues selected to be listed here include those that form a strong HB to the structure or otherwise had a total interaction energy more favorable than 5kcal/mol, or if that residue was alanized for any of the mutants. For each mutant, residues affected by the specific mutations are bolded and italicized.

	wild	**1a**	**1b**	**1c**	**2a**	**2b**	**3a**	**4a**	**4b**	**4c**
Asp21	-1.47	0.78	0.79	***0.00***	-0.44	1.62	0.67	-0.56	-4.69	3.70
Asn26	-4.49	-0.40	-0.20	-0.17	-0.14	-0.08	-0.24	-1.29	-0.22	-0.50
Asn28	-2.25	***0.15***	0.12	-0.67	-0.50	-0.04	-0.63	-12.78	-0.51	-4.91
Glu33	-7.67	0.56	***0.00***	1.11	-1.16	0.81	0.54	-0.89	-2.57	-0.23
Arg61	-4.01	-0.94	-1.59	-1.92	0.61	-1.93	-0.68	0.37	-0.08	-9.00
Pro63	-1.12	-0.04	-0.10	-0.22	0.00	***-0.38***	-0.05	0.05	0.10	-0.80
Ser67	-5.81	-0.23	-0.64	0.03	***-0.42***	0.00	-0.02	0.36	0.17	-9.95
Glu69	-1.33	-0.44	1.77	-2.43	-21.92	-7.77	0.34	-4.40	-4.02	-0.37
Asn70	-8.37	-0.61	-0.13	-5.73	-4.08	-3.23	-0.19	-0.77	-3.42	-10.29
Tyr108	-8.04	-0.63	-1.17	-1.15	-7.24	-0.68	-16.67	-0.23	-10.42	-1.82
Tyr112	-2.72	-3.29	-0.67	-0.12	-0.13	-0.14	***0.00***	-0.07	-1.36	-0.49
Lys114	-6.37	-0.73	-1.22	0.14	-3.69	0.00	0.42	2.29	-4.16	-3.07
Asn115	-7.87	-0.24	-0.11	-0.06	-0.03	-0.01	-0.14	-0.10	-0.12	-0.21
Total	**-85.63**	**-14.43**	**-13.54**	**-32.60**	**-73.70**	**-22.99**	**-35.36**	**-76.28**	**-83.44**	**-89.45**

VII. Figures

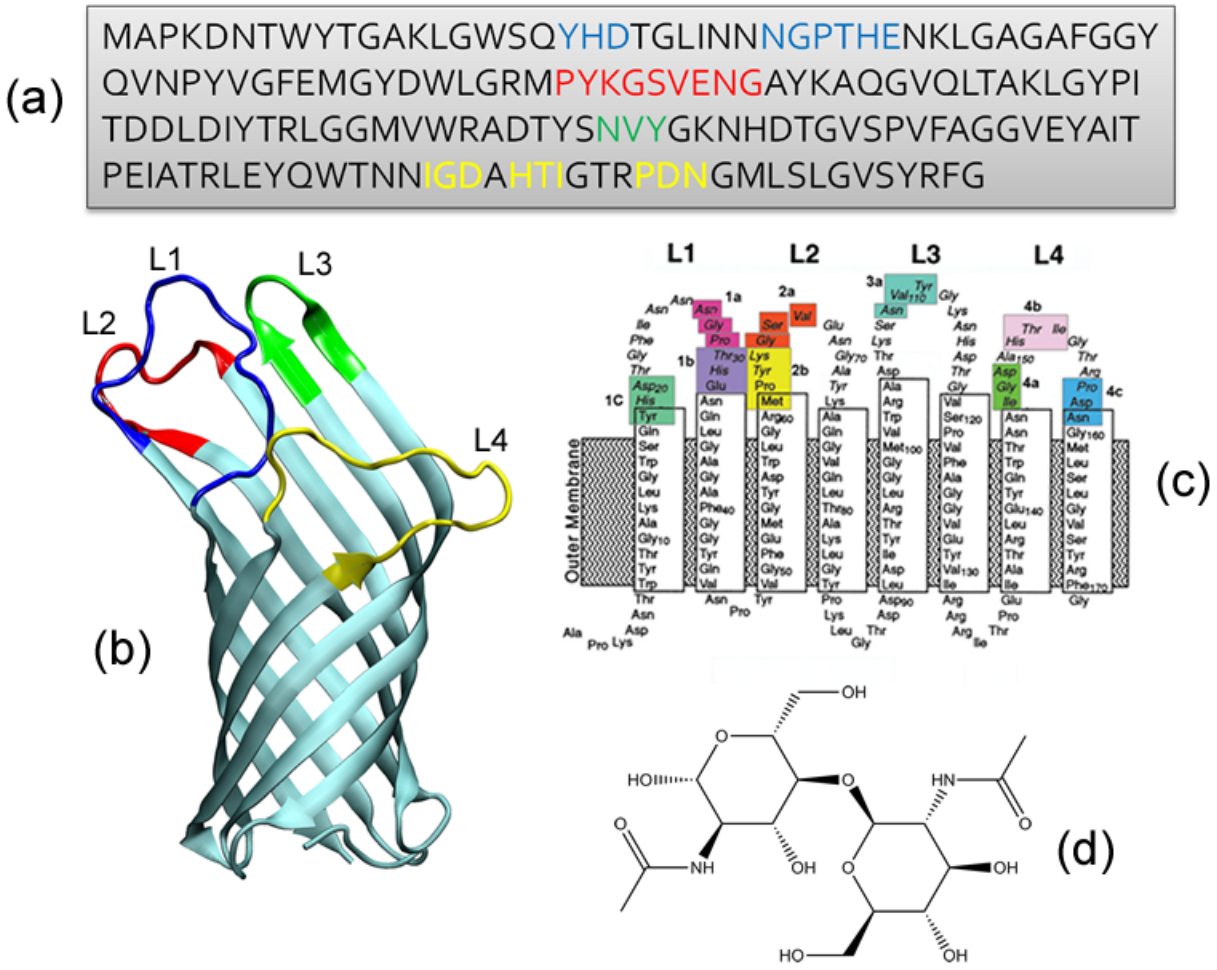


Figure 1 Overview of OmpA structure used in this study.
(a) Amino acid sequence with the three sets of amino acid residue mutations (always to alanine) color-coded according as follows:
Loop 1 (blue): 1a: N28-G29-P30, 1b: T31-H32-E33, 1c: Y19-H20-D21.
Loop 2 (red): 2a: X66-X67-X68, 2b: P63-Y64-K65, 2c: E69-N70-G71.
Loop 3 (green): 3a: N110-V111-Y112.
Loop 4 (yellow): 4a: I148-G149-D150, 4b: H152-T153-I154, 4c: P158-D159-N160.
(b) OmpA crystal structure (1QJP, 2.5Å resolution) with the individual loops numbered and color-coded as in a.
(c) Schematic of OmpA structure showing each of the 10 mutation set, plus the trans-membrane and loop regions.
(d) The chitobiose ligand.

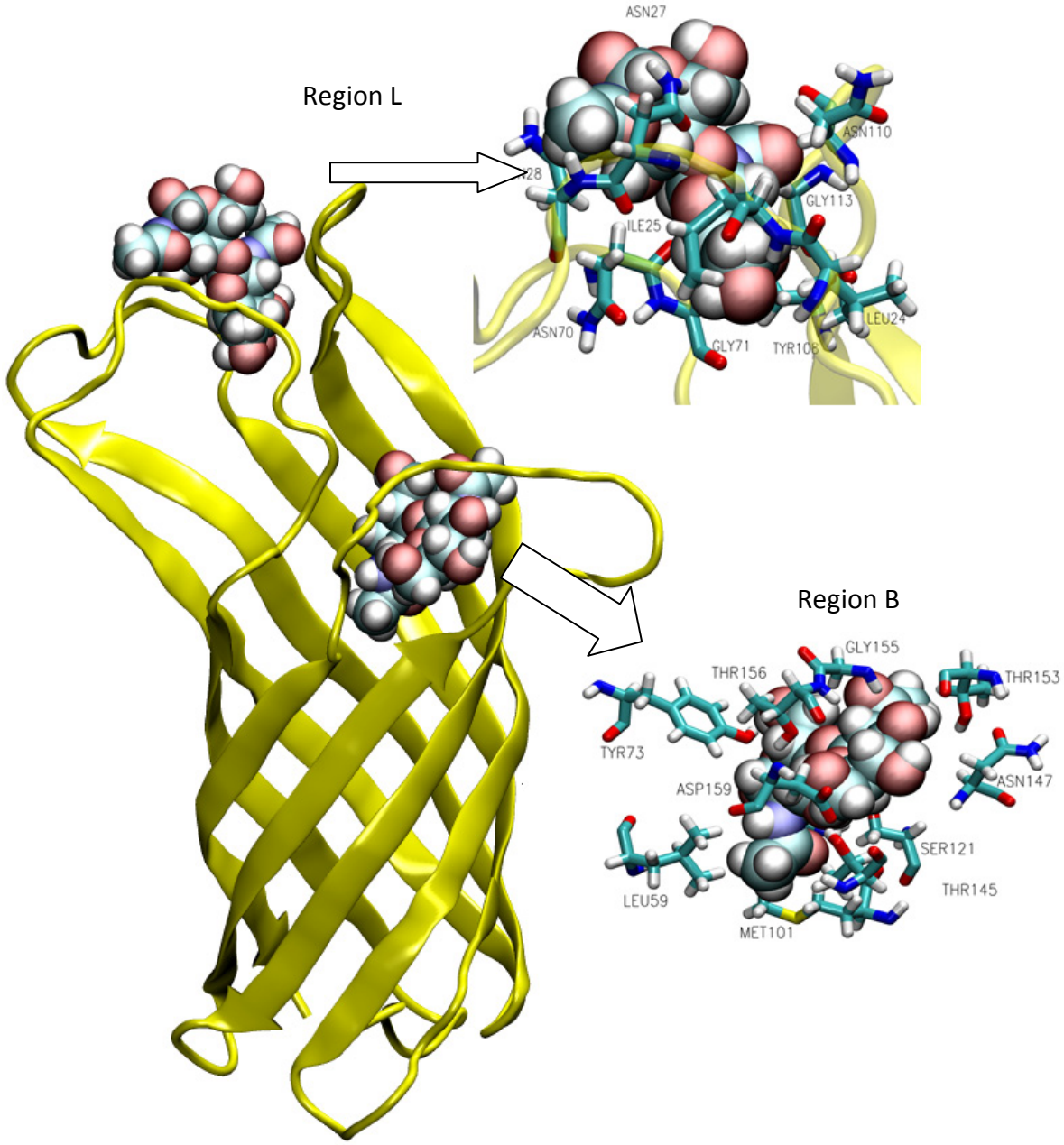
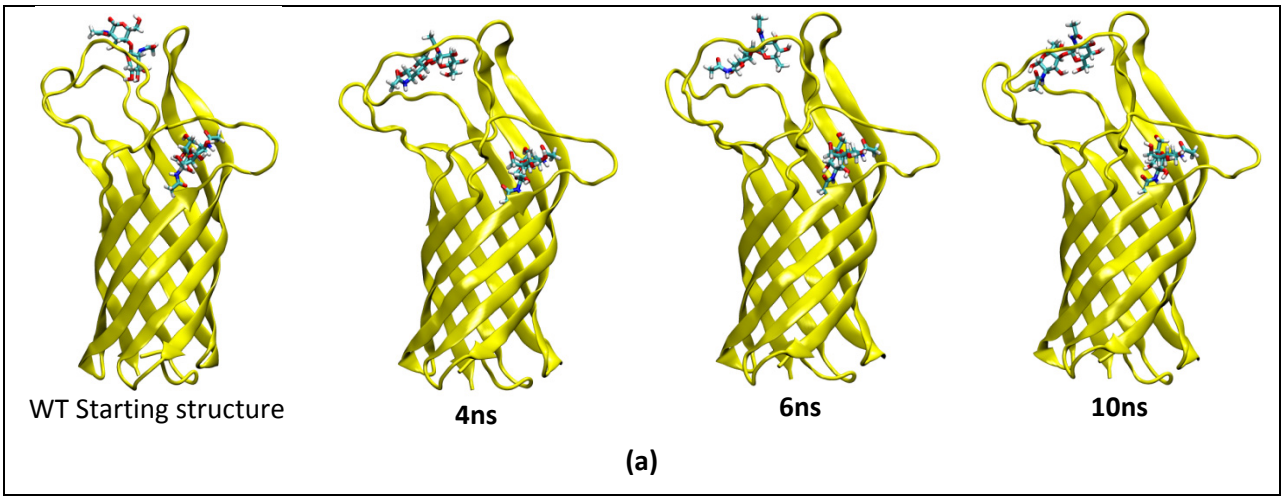
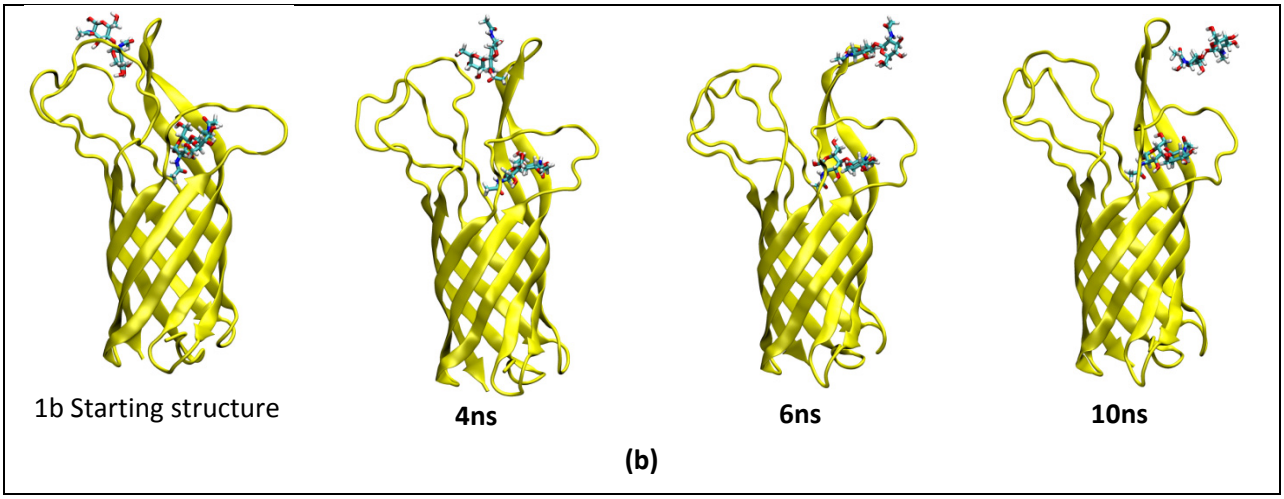


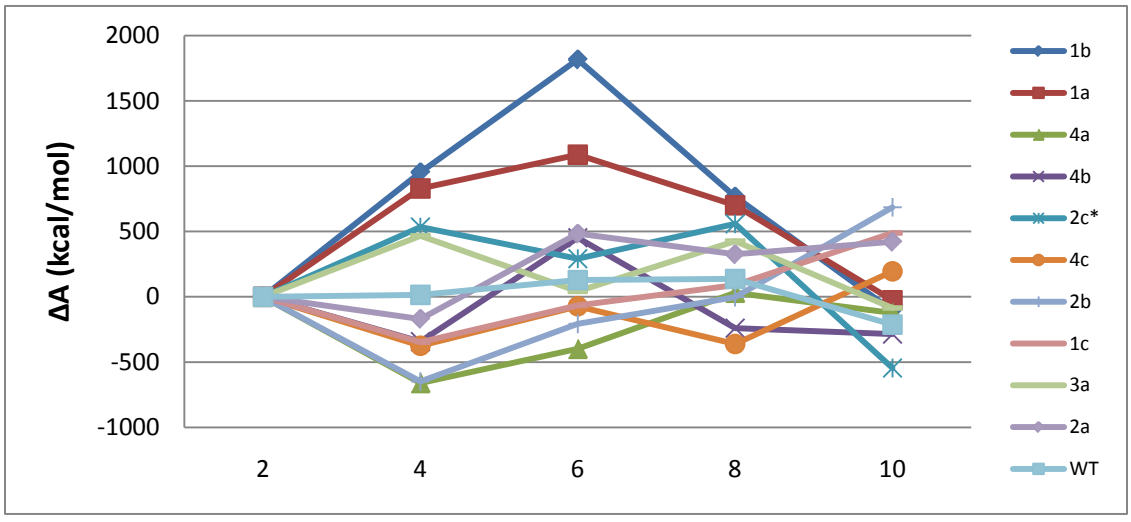
Figure 2 Structure of the chitobiose/OmpA (WT) complex after 30 ns of MD in explicit solvent, counterions, and membrane. These MD studies started with the two structures from docking:
Pose L (for Loops) with chitobiose bound to Loops 1, 2 and 3)
Pose B (for Barrel) with chitobiose bound to Loop 4.
Upper right inset: L binding site, showing all residues within 3Å. There are 4 HBs, 2 of which are removed upon mutation (1a and 2c mutants).
Lower right inset: B binding site, showing all residues within 3Å. There is a strong HB to Thr153, which is removed in the 4b mutant.


WT Starting structure
4ns
6ns
10ns

(a)

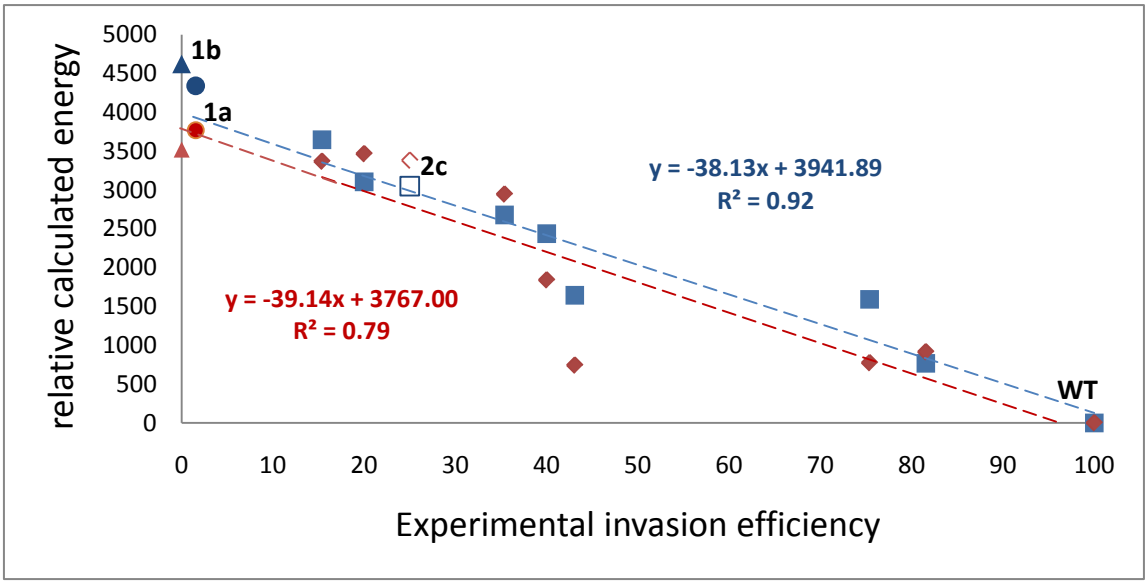

1b Starting structure
4ns
6ns
10ns

(b)

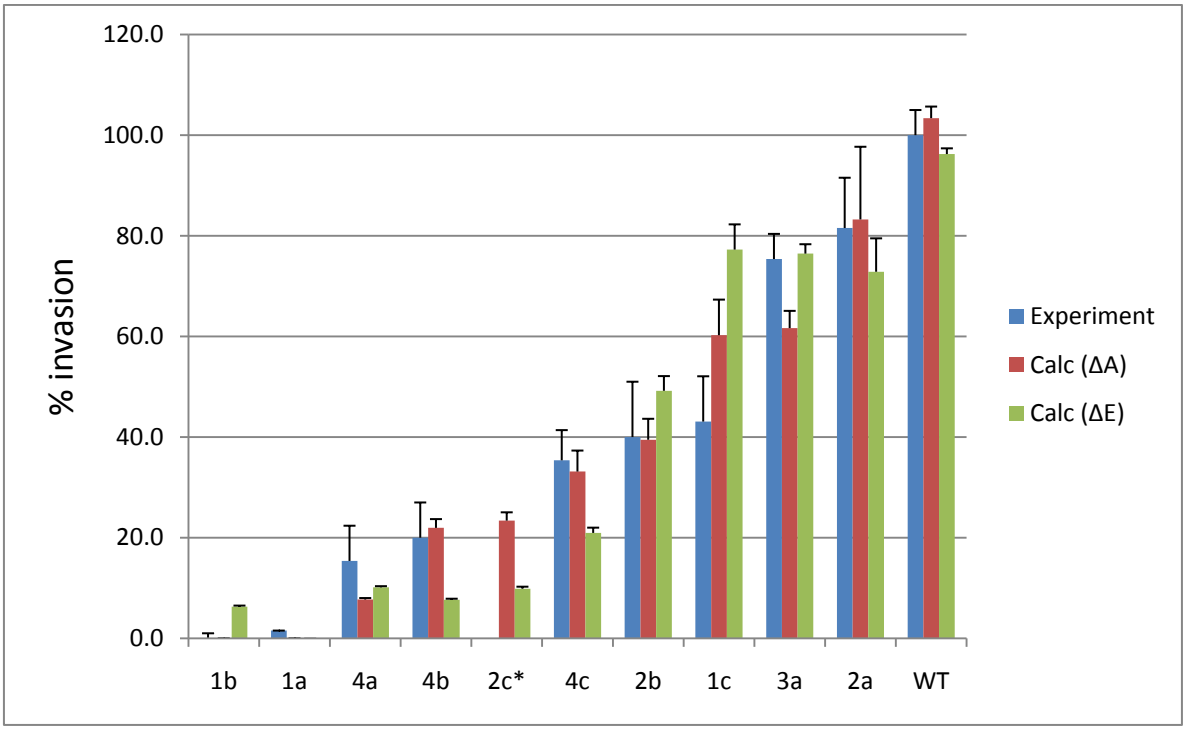

ΔA (kcal/mol)
2000
1500
1000
500
0
-500
-1000
2
4
6
8
10
1b
1a
4a
4b
2c*
4c
2b
1c
3a
2a
WT

(c)

Figure 3 Snapshots of OmpA/chitobiose complexes during the MD, with calculated thermodynamic quantities (kcal/mol) relative to the starting structure (A=Helmholtz free energy, E_0=Zero-Point corrected energy and TS = temperature times entropy). *(a)* **WT OmpA**. The chitobiose in both L and B regions are flexible but remain strongly bound. The largest change in free energy is < 1.6% of the total. *(b)* **1b mutant.** The chitobiose in region L gets ejected from the structure after 6 ns: the calculated energy (enthalpy) is 3.1% less favorable than for WT, but the total entropy (the TS term) is more favorable by 2.3%, due mainly to the increased motions of the chitobiose. The net result is that the free energy is only 0.7% less favorable than the WT. After 10ns, the system has a less favorable free energy than WT by 16% due to the less favorable energy (20%), but more favorable TS (3.7%). Similar results are found for the 1a mutant (increase in A of 26%, becoming unbound after 6.7 ns). Indeed our experiments show no invasion of the bacteria experimentally for mutants 1b and 1a. *(c)* **Helmholtz free energy profile** for WT and 10 mutants during dynamics. The energies are reported relative to the free energy of the system at 2ns. The profile of the 1a and 1b mutants is unique: increasing until 6ns then falling dramatically once the chitobiose becomes unbound.



(a)



(b)

Figure 4 *(a)* Correlation between free energies from theory and experimental invasion activity. Abscissa: experimental invasion activities (% relative to WT) of the nine OmpA mutants. Ordinate: Filled red squares – Calculated free energies (kcal/mol) relative to WT, Filled blue diamonds –

calculated energies (enthalpies). The special cases of WT and the 1a (filled red and blue circles), 1b (filled red and blue triangles) and 2c (open symbols) mutants are shown. For the 2c mutant open symbols are shown since there are no experimental data (the abscissa value chosen based on the fit, which constitutes a prediction). The lines are fitted to the 8 cases for which there are experimental data. The 1a and 1b mutants are predicted not to invade since the chitobiose in region L becomes unbound during dynamics. This is consistent with observation. We find excellent correlation of the free energy to the experimental invasion (92%). Ignoring the entropy (energy curve) leads to a poorer correlation (79%). *(b)* Comparison between experimentally observed invasions, and that predicted from the free energy best-fit line. The error bars are shown. We correctly predict no invasion for the 1a and 1b mutants since they are unbound. Overall we find excellent agreement between the experimentally observed and predicted invasion values.

SI. Supplementary Methods

a) Initial equilibration of the OmpA and mutant protein structures:

The starting structure for OmpA was obtained from the 1.65Å high resolution X-ray crystal structure[1] (PDB Code: 1QJP). This protein structure was immersed in a waterbox of 3000 TIP4P[2] rigid water molecules, removing all water molecules that overlapped the protein atoms. NaCl salt was added to a mimic the experimental concentration of 0.1M with extra sodium ions added to neutralize the system as needed. Using the AMBER all atom 1999 force field[3] with the recent SB modifications[4], structure was minimized to an RMS force of 500 kcal/mol, as in our previous studies[5-7] using the LAMMPS[8] MD simulation engine. This was the starting protein structure for simulations. Each of the ten mutants was created by mutating the specified three residue sequence to alanines and re-optimizing the neighboring side chains using the Side-Chain Rotamer Energy Analysis Method (SCREAM).

The solvated protein was simulated for 5ns of NPT dynamics at 300K, with a temperature-coupling constant of 100.0 femtosecond, and a 2.0 ps damping constant. The short-range non-bonded interactions (van der Waals and coulombic) were computed with a real space cutoff of 10Å.

This Amber force field uses the 12-6 Lennard Jones function to describe the van der Waals interaction, with any interaction not explicitly defined obtained using the geometric combination rules $r_{AB} = \sqrt{r_{AA} r_{BB}}, \sigma_{AB} = \sqrt{\sigma_{AA} \sigma_{BB}}, where\ E_{LJ} = \sum_i \sum_j \sigma_{ij} \left[\left(\frac{R_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{ij}}{r_{ij}} \right)^{6} \right]$. The long range electrostatics were evaluated with the particle-particle particle mesh ewald (PPPM) method[9] with an interpolation order of 4 and an tolerance of 10^{-4}. The MD was

performed using integration timestep of 2fs, employing SHAKE[10] constraints on all bonds involving hydrogens and on the HOH angle, with the neighbor list rebuilt every 10 steps.

b) Scanning the entire receptor for putative binding region

For a protein, there may be more than one binding region for one or more ligands. Thus our first step is to scan the entire protein for potential binding regions with no assumption on the binding site. The entire molecular surface of the predicted structure is mapped[11] and spheres representing the empty volume of the receptor are generated (currently using the Sphgen program in DOCK4.0 suite of programs[12]. The entire set of receptor spheres is partitioned into ~30 to 50 overlapping regions and the ligand being docked is used to scan for the putative binding region. The consensus of ligand structures corresponding to the most energetically favorable sites is used to determine one or more putative binding region. For some ligands this scan of the entire receptor may yield two or three putative binding regions, with similar binding energies in each region. In such cases we merge the spheres of all the regions with similar binding regions and perform docking over this larger region.

c) Docking of chitobiose units to OmpA protein structures:

The two chitobiose binding sites on OmpA (L and B) were determined above. The chitobiose unit was then docked into docked into each of these regions, with the top energy structure for each region retained.

The GenMSCDock method described below was used to obtain ligand poses bound to OmpA. The method efficiently samples ligand conformational space and takes into account protein side-chain flexibility during the ligand binding process. The method is an extension of MPsim-Dock[13], and uses a Monte Carlo method to generate various possible ligand

conformations in the field of the protein. The conformations are selected based on diversity of the conformations from each other to cover the conformational space. We call these conformations as “family heads” and they differ from each other by at least 0.6Å in CRMS (RMSD in coordinates). Next the energy of interaction of each family head with the protein is calculated, and about 10% of the good energy “family heads” are selected for further enrichment of these conformations. The enrichment is done by generating conformations using Monte Carlo procedure and selecting only those conformations that are close (within 0.6 Å CRMS) to the good energy family heads. A ligand conformation that is within 0.6 Å of the family head is known as a child of that family head. The enrichment cycle is performed until each chosen good energy family head has an average of at least 6 children. We then calculate the ligand protein interaction energy for all the children of each family head. The conformations (family heads and children) are sorted by energy and the best 50 ligand conformations are chosen, each of which is then taken through an optimum protein-side-chain placement using SCREAM side-chain optimization method[14]. These structures are then conjugate-gradient minimized in two steps: first ligand-movable/ protein-fixed and second all-movable. Then the binding energy is calculated for these conformations.

This method was further improved by using a protein with bulky residues alanized to accommodate ligands. This replaces all bulky residues (Phe, Trp, Tyr, Val, Ile, and Leu) by Ala and all the docking is done to this bulky-alanized structure (other residues can be alanized if needed). The residues are dealanized to their original type, using SCREAM for each docking ligand conformation. It also has the option to neutralize charged residues that usually results in less noise in binding energies after correcting for the pKa of exposed residues. This allows for a

better SAR (structure-activity relationship studies) as a smoother electrostatic potential in the binding region of the proteins leads to much smaller solvation energies and interaction energies both of which change much less for small changes in ligands.

SII. Supplementary tables

Table S1: Primers for site-directed mutagenesis.

Name	Sequence (5'-3')	Mutation
Omp-F1a	GGTTTCATCAACAAC ***GCTGCCGCG*** ACCCATGAAAACC	Loop 1a
Omp-R1a	GGTTTTCATGGGT ***CGCGGCAGC*** GTTGTTGATGAAACC	
Omp-F1b	ACAACAATGGCCCG ***GCCGCTGCA*** AACCAACTGGGCG	Loop 1b
Omp-R1b	CGCCCAGTTGGTT ***TGCAGCGGC*** CGGGCCATTGTTGT	
Omp-F1c	GGCTGGTCCCAG ***GCCGCTGCC*** ACTGGTTTCATCA	Loop 1c
Omp-R1c	TGATGAAACCAGT ***GGCAGCGGC*** CTGGGACCAGCC	
Omp-F2a	CGTATGCCGTACAAA ***GCCGCCGCTG****CA* AACGGTGCATAC	Loop 2a
Omp-R2a	GTATGCACCGTT ***TGCAGCGGCGGC*** TTTGTACGGCATACG	
Omp-F2b	CTGGTTAGGTCGT ***GCGGCGGCCGCA*** GGCAGCGTTGAAA	Loop 2b
Omp-R2b	TTTCAACGCTGCC ***TGCGGCCGCCGC*** ACGACCTAACCAG	
Omp-F3a	CAGACACTAAATCC ***GCCGCTGCT G***GTAAAAACCACGA	Loop 3a
Omp-R3a	CGTGGTTTTTACC *AG****CAGCGGC*** GGATTTAGTGTCTGC	
Omp-F4a	CAGTGGACCAACAAC ***GCCGCTGCC*** GCACACACCATCG	Loop 4a
Omp-R4a	CGATGGTGTGTGC ***GGCAGCGGC*** GTTGTTGGTCCACTG	
Omp-F4b	ACATCGGTGACGCA ***GCTGCCGCCGCC*** ACTCGTCCGGAC	Loop 4b
Omp-R4b	GTCCGGACGAGT ***GGCGGCGGCAGC*** TGCGTCACCGATGT	
Omp-F4c	CATCGGCACTCGT ***GCGGCTGC****C* GGCATGCTGAGC	Loop 4c
Omp-R4c	GCTCAGCATGCC ***GGCAGCCGC*** ACGAGTGCCGATG	

* Changed nucleotides are in bold and italics.

Table S2: Properties of OmpA and its 9 mutants during dynamics (without chitobiose units bounds). Statistics are obtained over the entire 10ns MD trajectory. We present two measures of the heavy atom CRMS 1) using the starting structure as the reference, to indicate how much the system relaxes from the initial minimized structure during dynamics and 2) using the equilibrated structure as a reference, to indicate convergence. Equilibration in all the structure is seen after 2ns of MD.

	SASA (Å^2)	**Rgyr (Å)**	**CRMS (Å)**	
			Starting	***Equilibrated***
WT	11196.9±103.1	18.38±0.07	3.30±0.90	1.68±0.20
1a	11531.3±220.7	18.56±0.06	3.58±0.90	1.77±0.14
1b	11510.8±303.1	18.39±0.09	3.61±1.11	2.19±0.34
1c	11688.0±156.3	18.38±0.08	3.59±0.73	1.44±0.11
2a	11545.8±101.0	18.54±0.07	3.31±0.70	1.31±0.14
2b	11880.8±64.6	18.50±0.02	3.07±0.58	1.16±0.03
3a	11676.6±134.4	18.70±0.14	3.27±0.86	1.71±0.17
4a	11674.7±92.3	18.90±0.13	4.57±1.24	2.13±0.53

4b	11302.7±169.6	18.44±0.70	3.41±0.60	1.18±0.13
4c	11782.7±224.822	18.38±0.08	3.17±0.69	1.26±0.13

Table S3: Interaction energies (kcal/mol) of chitobiose in residues in region B within 10Å. The wild type energies are given are taken from table 1 and compared to the 9 mutants in this study. The residues selected to be listed here include those that form a strong HB to the structure or otherwise had a total interaction energy more favorable than 5kcal/mol, or if that residue was alanized for any of the mutants. For each mutant, residues affected by the specific mutations are bolded and italicized.

	WT	1a	1b	1c	2a	2b	3a	4a	4b	4c
ARG 61	-5.82	-0.85	-0.60	-0.76	0.43	-0.86	2.16	-1.34	-4.51	0.13
VAL 78	-9.33	-0.52	-0.42	-0.34	0.02	-0.31	-0.16	-0.34	-0.52	-0.19
MET 101	-9.36	-5.79	-6.36	-7.96	-9.11	-7.34	-6.59	-6.42	-8.06	-2.60
TRP 103	-12.65	-13.96	-13.31	-10.15	-19.02	-22.06	-19.72	-16.61	-15.69	-13.80
THR 118	-6.83	-1.10	-1.12	-0.69	-0.68	-1.30	-0.76	-0.58	-0.06	-0.44
SER 121	-9.16	-13.96	-4.88	-9.36	-17.37	-13.27	-19.78	-9.47	-5.63	-1.81
THR 145	-1.03	-15.84	-14.38	-20.94	-5.92	-10.08	-5.29	-10.80	-13.84	-10.90
ILE 148	-13.91	-1.32	-0.99	-1.24	-2.10	-0.29	-1.44	***-0.05***	0.24	0.04
ASP 150	-1.71	-1.37	1.79	-0.05	2.19	0.89	0.88	***0.10***	-0.23	0.66
THR 153	-15.14	-0.08	-1.28	-13.39	-6.92	0.11	-5.85	-1.21	***0.02***	0.82
Total	-148.83	-133.03	-145.00	-137.93	-150.69	-155.80	-128.29	-94.83	-109.59	-91.64

Table S4: Profile of thermodynamics observables during dynamics for WT and the 10 OmpA mutants considered in this study. All values are reported relative to the value of the system evaluated at 2ns.

	ΔA (kcal/mol)				**ΔE_0 (kcal/mol)**				**TΔS (kcal/mol)**			
	Time (ns)				**Time (ns)**				**Time (ns)**			
	4	**6**	**8**	**10**	**4**	**6**	**8**	**10**	**4**	**6**	**8**	**10**
1b	953.2	1816.8	766.8	-106.2	1165.0	1318.0	1813.0	670.0	211.8	-498.8	1046.2	776.2
1a	829.7	1087.8	700.2	-26.3	921.0	981.0	1348.0	234.0	89.5	-106.9	647.8	258.9
4a	-660.6	-398.8	30.8	-123.7	-560.0	-343.0	47.0	-140.0	98.5	55.0	17.0	-16.4
4b	-342.9	452.4	-240.0	-285.4	-123.0	377.0	-19.0	-436.0	219.9	-75.4	221.0	-150.6
2c*	534.6	291.8	558.1	-545.1	588.0	385.0	486.0	-275.0	53.4	93.2	-72.1	270.1
4c	-375.1	-73.1	-361.6	194.8	-100.0	144.0	-254.0	292.0	275.9	218.8	109.5	98.0
2b	-648.7	-207.0	-2.3	685.3	-787.0	-237.0	-89.0	450.0	-138.3	-30.0	-86.7	-235.3
1c	-348.2	-66.5	89.9	488.9	-245.0	133.0	64.0	563.0	103.8	197.5	-26.6	74.1
3a	466.2	41.7	425.9	-86.1	381.0	-89.0	328.0	624.0	-130.1	-130.1	-97.6	709.9
2a	-169.8	483.4	325.3	420.9	-7.0	329.0	415.0	347.0	162.8	-154.4	89.7	-73.9
WT	14.9	128.1	136.7	-213.0	10.3	38.0	193.9	-172.9	-4.6	-90.1	57.2	40.1

Table S5: Thermodynamics of chitobiose binding in OmpA and the 9 mutants considered in this study. The computed Helmholtz free energy (ΔA), Enthalpy (ΔE_0) and Entropy (ΔS at T = 300K) for each of the mutants are given with respect to the wild type (WT). The free energy and enthalpies of all the mutants is less favorable than the wild type, while the entropies are more favorable (except in the case of the 2a, 4b and 4c mutants). For mutants 1a and 1b, the chitobiose in region L is unbound leading to an increase in the entropy of the system. The chitobiose is still bound in all other mutants, and is more flexible due to the alanization of bulky amino acids. There is a 92% correlation between the experimentally observed invasions and the free energy of binding (ignoring the unbound 1a and 1b mutants which do not invade). Considering only the energy (enthalpy) leads to a 79% correlation. The best fit line is used to predict a theoretical rate of invasion.

	% Invasion						Energy (kcal/mol)					
	Exp		Calc (ΔA)		Calc (ΔE_0)		ΔA		ΔE_0		TΔS	
	avg	±	avg	±	avg	±	avg	±	avg	±	avg	±
1b	0.0	1.0	0.0	0.0	6.3	0.2	4623.1	51.6	3521.4	42.9	-1103.1	16.4
1a	1.5	0.5	0.0	0.0	0.0	0.0	4342.2	74.0	3770.1	59.0	-570.2	15.0
4a	15.4	7.0	7.7	0.3	10.1	0.2	3647.8	25.8	3370.0	25.0	-293.9	7.5
4b	20.0	7.0	22.0	1.7	7.6	0.2	3104.9	54.4	3467.6	33.7	361.4	20.6
2c*	-	-	23.4	1.7	9.9	0.4	3050.0	48.3	3381.4	48.4	330.1	4.6
4c	35.4	6.0	33.1	4.2	20.9	1.0	2678.2	85.6	2947.0	57.6	267.5	30.1
2b	40.0	11.0	39.5	4.2	49.2	2.9	2437.1	70.8	1841.4	67.1	-595.7	7.2
1c	43.1	9.0	60.2	7.1	77.3	5.0	1644.7	78.6	743.0	73.3	-901.7	9.3
3a	75.4	5.0	61.6	3.4	76.5	1.9	1591.3	37.3	774.4	27.6	-816.9	12.6
2a	81.5	10.0	83.3	14.4	72.9	6.6	767.4	115.5	915.4	102.7	148.1	13.4
WT	100.0	5.0	103.4	2.3	96.2	1.1	0.0	14.1	0.0	13.0	0.0	5.7

*the 2c mutant was not determined experimentally, we present a prediction here that can be tested

SIII. Supplementary Figures

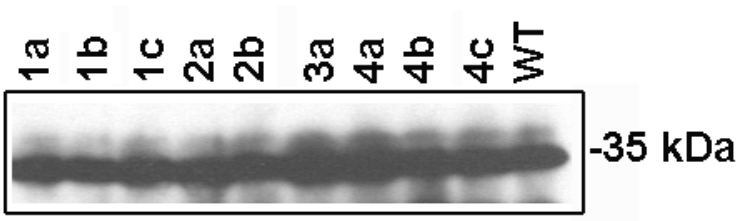


Figure S1: Western blot analysis of OmpA proteins from *E. coli* mutants. *E. coli* strains were grown overnight in Luria broth, washed and outer membrane proteins isolated. Equal amounts (10 μg per lane) was loaded and separated on SDS-polyacrylamide gel. The proteins were then transferred to a nitrocellulose membrane and immunoblotted with anti-OmpA antibodies. The bound antibody was probed with HRP-coupled secondary antibody followed by ECL reagent.

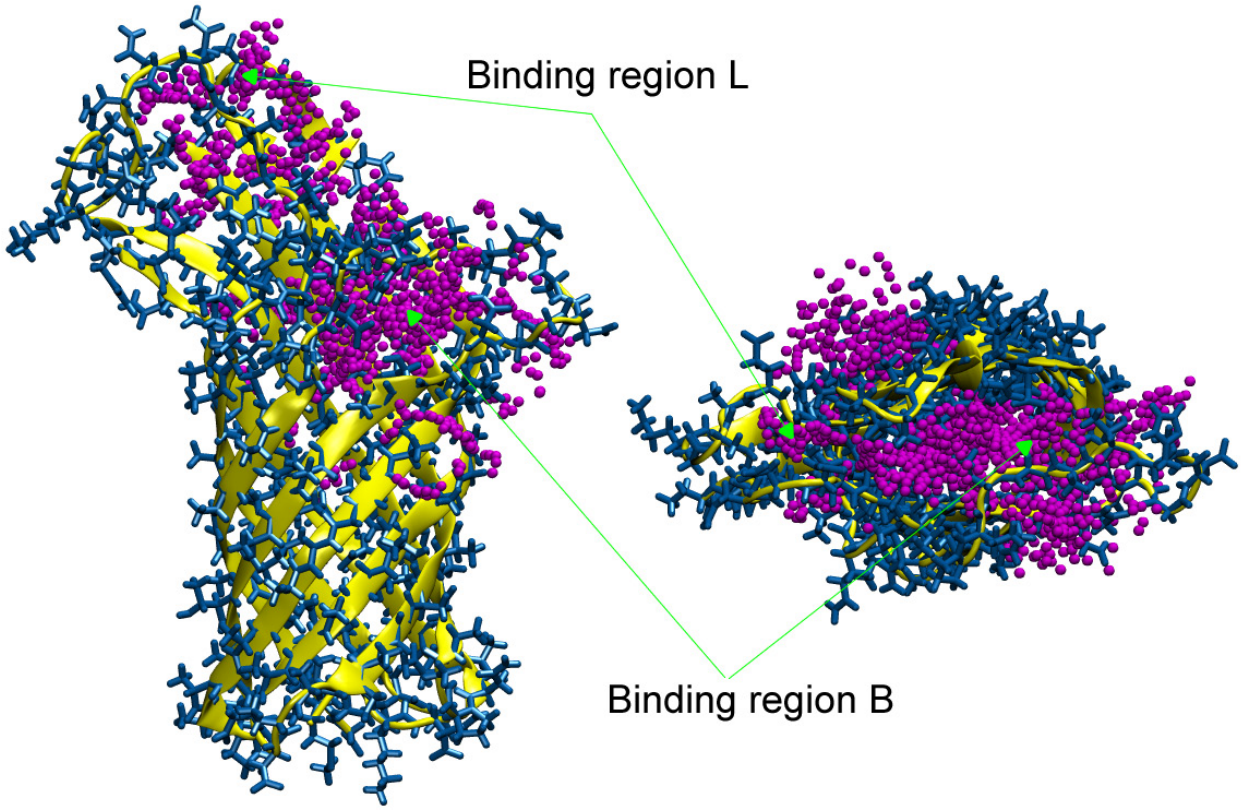


Figure S2: Volume filling spheres, with the chitobiose binding site (L and B) shown.

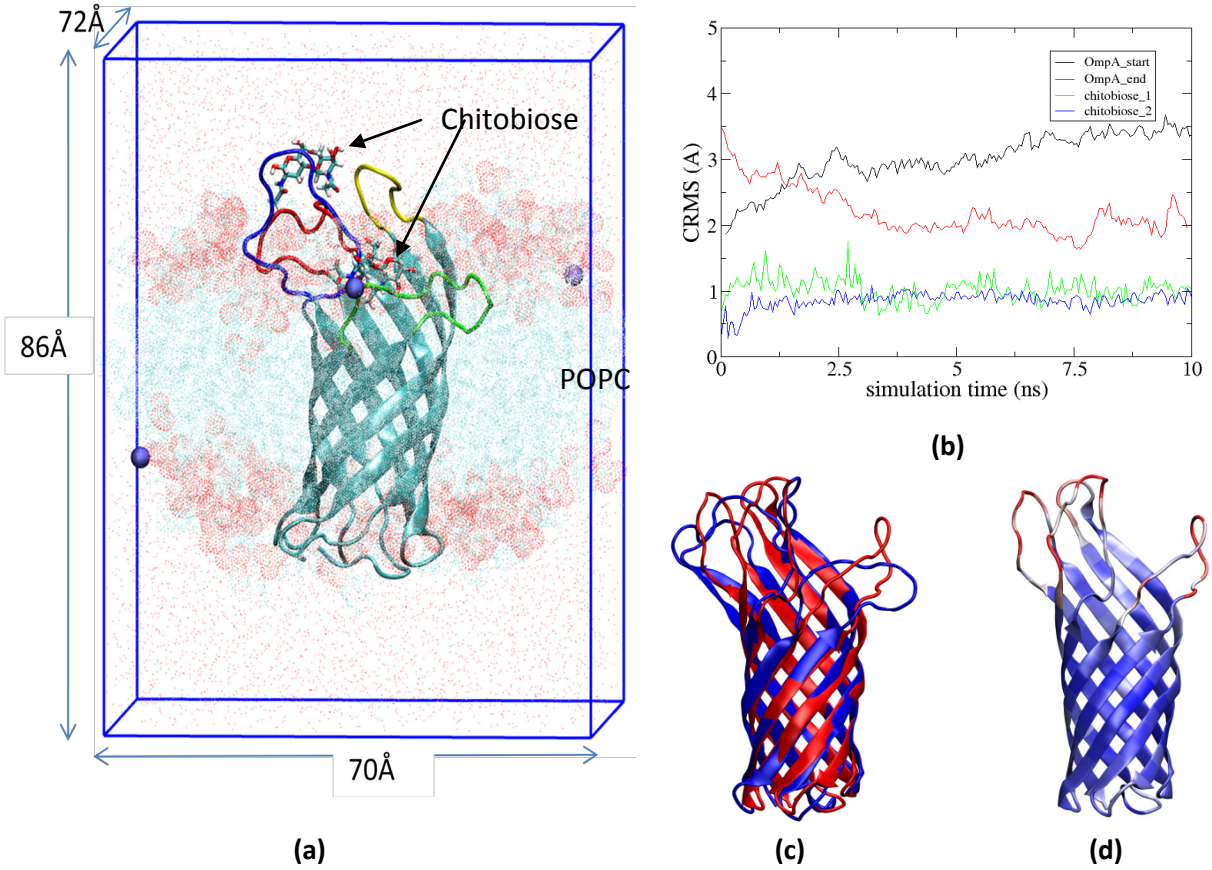


Figure S3: MD simulation of OmpA wild type with chitobiose units bound in region L and region B. the label says chitobiose 1 and 2 **(a)** Final 3D simulation cell (86Å x 70Å x 72Å) with the OmpA (the loops are color coded as Figure 1b), two bound chitobiose molecules, 144 POPC membrane molecules and three Na+ counter-ions (blue spheres) shown. The 2589 TIP4P water molecules (not shown) were added to the simulation cell to achieve an initial density of 1 g/cc. The total system has 45303 atoms. **(b)** CRMS deviation of various parts of OmpA during dynamics. Calculations are performed using the entire 10 ns trajectory (5ns NPT and 5ns NVT dynamics), after initial minimization and system heating. The fluctuations in the protein structure are accessed using two different reference points: the start crystal structure (black line) and the converged MD structure (red line). The CRMS for each chitobiose molecule (red and blue lines) is also calculated, compared to their starting structures. Using any of these measures, we see that convergence occurs after approximately 4ns of MD. **(c)** Comparison between the starting crystal structure (red) and the equilibrated MD structure (blue) of OmpA. The two structures have a CRMS of 3.1Å, which is acceptable for these sorts of simulations and close to the experimental resolution of 2.5Å. **(d)** Per residue CRMS difference of the OmpA starting structure to the MD equilibrated structure. Here the residue coloring scheme is as follows: <0.3Å blue, 0.3 – 0.6Å: white and > 0.6Å red. We see the largest deviations in the two structures occur in the loop regions, with the smallest deviations occurring in the transmembrane barrel region.

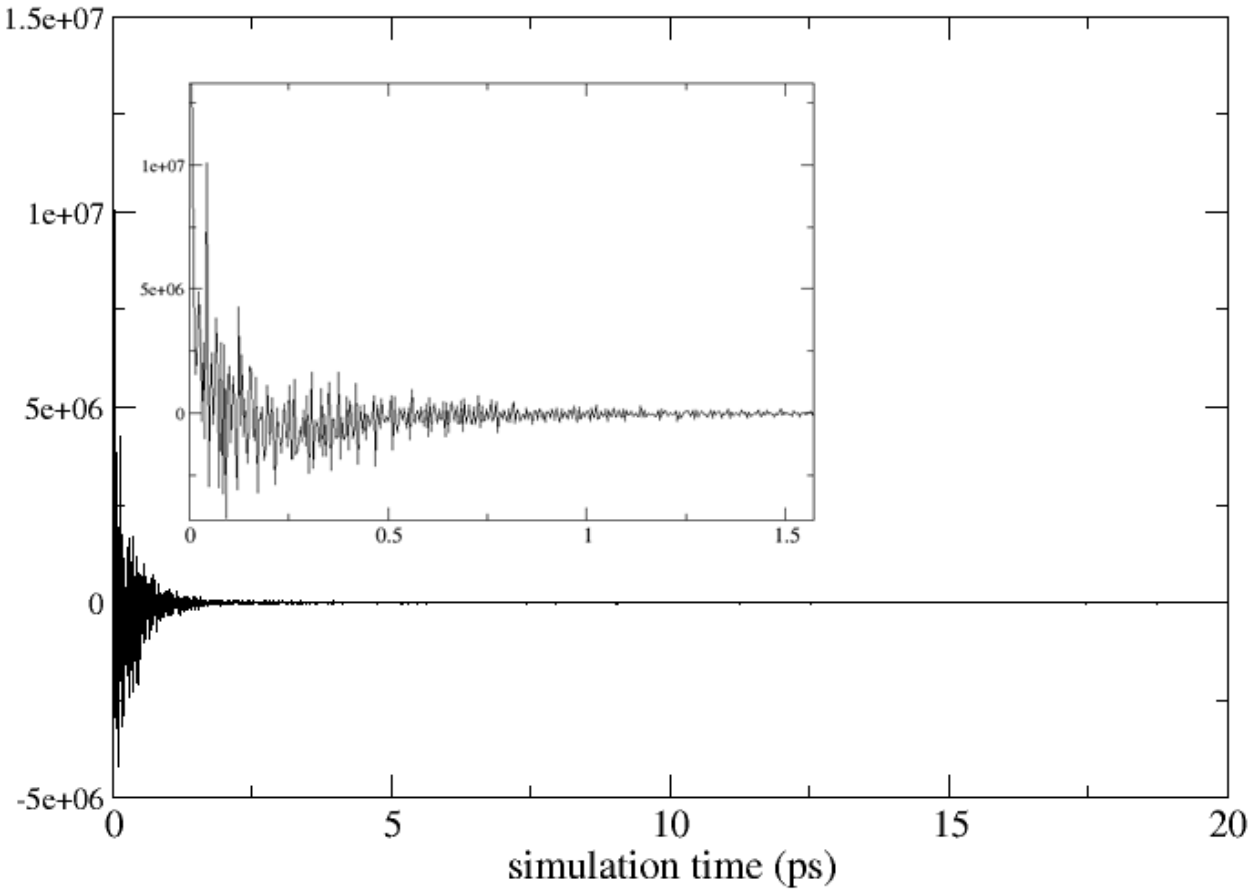


Figure S4: Velocity autocorrelation function (VAC) for WT OmpA from 2ns MD. The density of states (DoS) is obtained from the Fourier Transform of this VAC . The VAC decays to zero after 5ps. **Inset:** magnification of the first 1.5 ps.

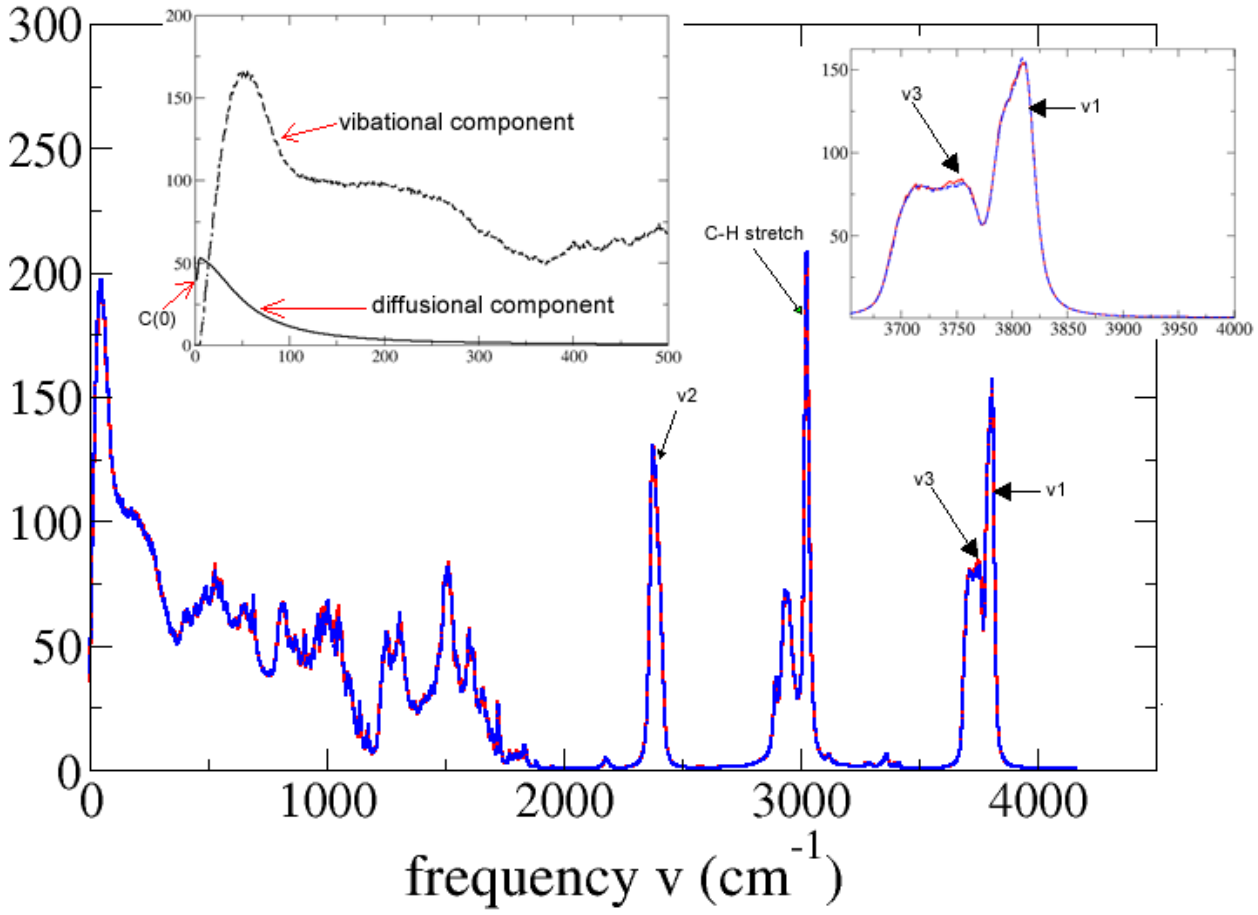
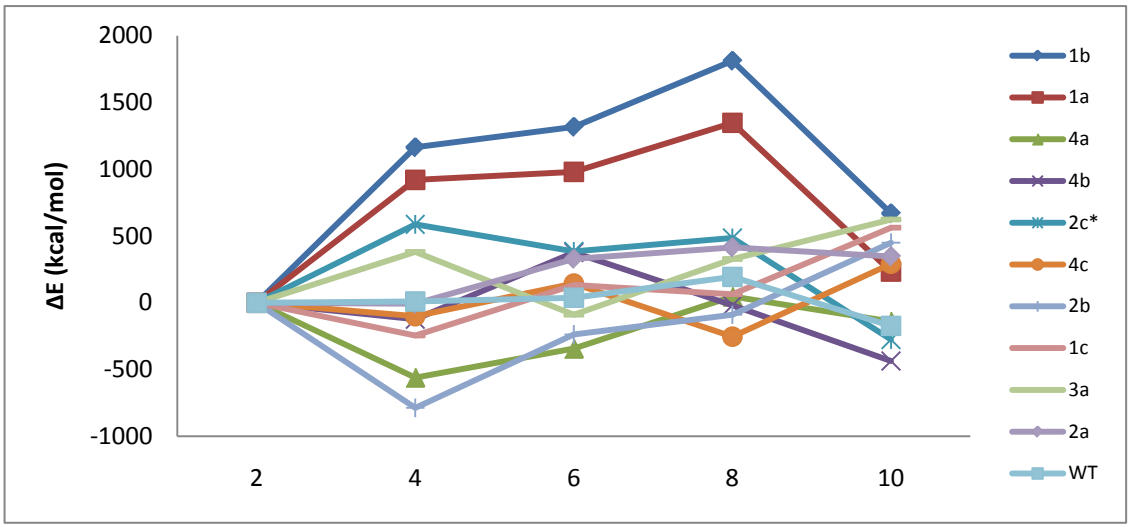
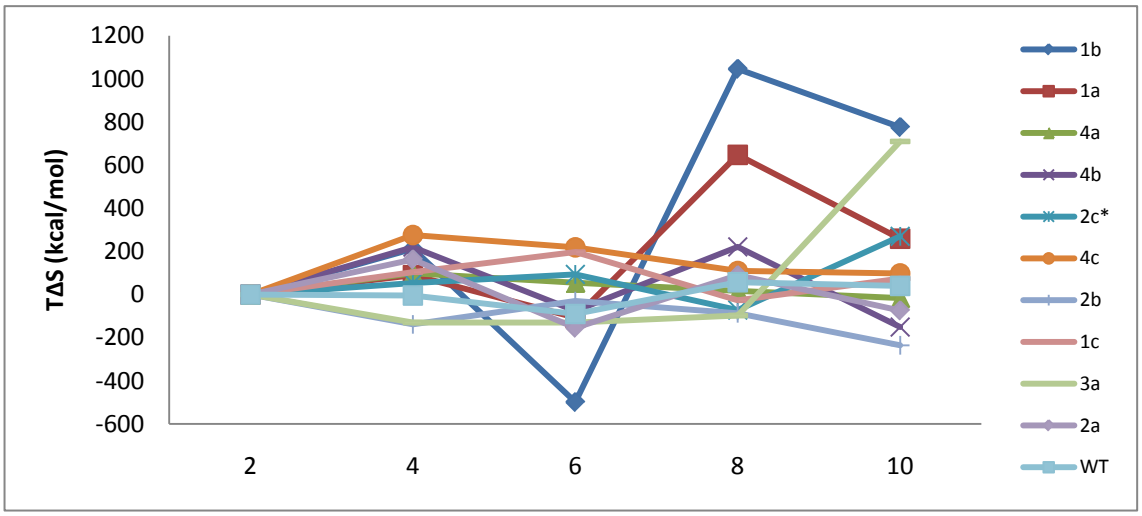


Figure S5: DoS of WT (solid red) and 1b mutant (dashed blue) calculated at 6ns during MD. This DoS is integrated with the weighting functions in equations (8) and (9) to obtain the relevant thermodynamics observables. S_0 is the value of the DoS at v = 0, and is related to the self diffusion coefficient of the system. The characteristic vibrational frequencies are labeled: O-H stretch (water symmetric v1: 3806 cm-1, water asymmetric v_3: 3750 cm^{-1}), H-O-H bend v_2 (2375 cm^{-1}) and C-H stretch (3025 cm^{-1}). The v_2 frequencies are the only ones not close to experiment (1600 cm-1) and is a deficiency of the water forcefield used (the force constant for the TIP4P H-O-H bend is 200 kcal/mol, twice the value needed to reproduce the experimental frequencies). **Insets: 1)** upper left - low frequency modes of WT, showing the decomposition of the total DoS into the diffusional and vibrational components. The $\hat{C}(0)$ parameter (the value of the DoS at 0 frequency) is 53.46 cm/mol, which is related to the self diffusion constant D=0.0973, according to equation (7). For the WT, we obtain f=0.1844 which corresponds to 25055 of the 135909 total 3N modes. The 1a mutant has more entropy than the wild type, due to the more mobile chitobiose unit (these low frequency diffusional modes contribute most to the entropy). **2)** upper right - O-H stretching modes. The broadening and splitting of the O-H band is representative is caused by the hydrogen bonding in the system.



(a)



(b)

Figure S6: Profile of OmpA thermodynamics during MD. The **(a)** enthalpy E_0 (corrected for zero-point energy) and **(b)** entropy S times temperature are shown, relative to the value at 2ns, for WT and the 10 mutants. Averages of these values are reported in table 2 of the main text (in the case of the 1a and 1b mutants, the averages were obtained from 0 – 6 ns, i.e. before the chitobiose was ejected).

SIV. Supplementary References

(1) Pautsch, A.; Schulz, G. E. *J. Mol. Biol.* **2000**, *298*, 273.

(2) Jorgensen, W. L.; Madura, J. D. *Molecular Physics* **1985**, *56*, 1381.

(3) Cheatham, T. E.; Cieplak, P.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845.

(4) Viktor Hornak; Robert Abel; Asim Okur; Bentley Strockbine; Adrian Roitberg; Carlos Simmerling *Proteins: Structure, Function, and Bioinformatics* **2006**, *65*, 712.

(5) Maiti, P. K.; Pascal, T. A.; Vaidehi, N.; Heo, J.; Goddard, W. A. *Biophys. J.* **2006**, *90*, 1463.

(6) Maiti, P. K.; Pascal, T. A.; Vaidehi, N.; Goddard, W. A. *NUCLEIC ACIDS RES* **2004**, *32*, 6047.

(7) Maiti, P. K. P., Tod A.; Vaidehi, Nagarajan; Goddard, William A. *J. Nanosci. Nanotechnol.* **2007**, *7*, 1712.

(8) Plimpton, S. *J. Comput. Phys.* **1995**, *117*, 1.

(9) Plimpton, S. J.; Pollock, R.; Stevens, M. In *Proc of the Eighth SIAM Conference on Parallel Processing for Scientific Computing* Minneapolis, MN 1997.

(10) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.

(11) Connolly, M. L. *Journal of Applied Crystallography* **1983**, *16*, 548.

(12) Ewing, T. J. A.; Kuntz, I. D. *J. Comput. Chem.* **1997**, *18*, 1175.

(13) Cho, A. E.; Wendel, J. A.; Vaidehi, N.; Kekenes-Huskey, P. M.; Floriano, W. B.; Maiti, P. K.; Goddard, W. A. *J. Comput. Chem.* **2005**, *26*, 48.

(14) Kam, V. W. T.; Goddard, W. A. *Journal of Chemical Theory and Computation* **2008**, *4*, 2160.

SECTION III. DNA NANOSTRUCTURES

Chapter 3. The PX/JX DNA Crossover Structures

[This chapter is a summary of the following paper: Pascal, T, et al. "Understanding DNA based nanostructures." Journal of Nanoscience and Nanotechnology 7.6 (2007):1712-1720. We have published two additional papers on this topic:

1. Pascal, T, et al. "Atomic-level simulations of Seeman DNA nanostructures: The paranemic crossover in salt solution." Biophysical Journal 90.5 (2006):1463-1479. Appendix II
2. Pascal, T, et al. "The stability of Seeman JX DNA topoisomers of paranemic crossover (PX) molecules as a function of crossover number." Nucleic Acids Research 32.20 (2004):6047-6056. Appendix III]

Intermolecular interactions of DNA are highly specific and readily programmable through Watson-Crick complementarity: A pairs with T, G pairs with C. This complementarity can be used to design systems in which single strands self-assemble into double strands, branched junctions, and other more complex motifs. The Seeman laboratory at New York University has exploited specificity of Watson-Crick pairing to synthesize a variety of branched DNA motifs that they have used to self-assemble both nanomechanical device and novel DNA nanostructures such as the cube and the truncated octahedron. Paranemic crossover (PX) DNA molecules and their topoisomer JX molecules, recently synthesized by Seeman and co-workers [1-2], are emerging as very important building blocks for building nanomechanical devices and for creating self-assembled DNA nanostructures [2-3]. However practical design and manufacture of nanoscale machines and devices requires overcoming numerous formidable hurdles in: synthesis, processing, characterization, design, optimization, and fabrication of the nanocomponents. Each area presents the significant challenges to the experimentalist with because the properties of nanoscale systems may differ significantly from macroscopic and molecular systems and because the manipulation and characterization of structures at the nanoscale is difficult. Even so, branched motifs of DNA that provide components for the self-assembly of 2D and 3D arrays at the nanoscale have been synthesized. The concepts of

crossover points that connect one double helical strand to a strand in a second double helix have been demonstrated to yield rigid motifs of DNA. Thus these crossover points connect two very flexible double helical structures to form a single rigid structure. Such rigid units as the DAO and DAE motif double crossover (DX) molecules [4] are critical in making a nanomechanical device.

Recently, the Seeman group discovered a new DNA motif, paranemic crossover (PX) DNA and one of its variants, JX_2 DNA, that provides the basis for a robust sequence-dependent nanomechanical device [1]. Using the sequence dependence of this device an array of such devices could be organized so that each device would respond individually to a unique set of signals. The crossover points in PX structures occur at every point at which the two strands from each double helix come together, as shown in **Figure 1**. As illustrated here the PX motifs can be built with varying number of nucleotides in the major and minor grooves. For example PX55 is the PX structure with five nucleotides in the major groove and five in the minor groove. Thus there are 10 nucleotides in one turn of the double helix. Various PX structures such as PX55, PX65, PX75, PX85 and PX95 have been synthesized in solution. These PX structures consist of four individual strands that are designed to complement in only one way. The W and N notations in the center of the molecule indicate the wide and narrow groove juxtapositions of the two helices.

Sometimes such experiments fail to make the desired structures while often they work as intended, but there is no clear cut understanding of why certain sequences and structures form stable conformations while others do not. Because of this difficulty in experimentally controlling nanostructures and in measuring the properties of nanosystems, it is possible for

modeling and simulation to play an essential role in nanoscale synthesis and design. Earlier [5-6] we have reported the atomistic level structural properties of the PX and JX structure in the presence water and monovalent Na+ ions and demonstrated how simulation can elucidate the structure-property relationships in such DNA nanostructure. In particular we have developed a strain energy analysis method based on the nearest-neighbor interaction and computed the strain energy for the PX molecules compared to the B-DNA molecules of the same length and sequence. We found that PX65 has the lowest calculated strain energy (~ –0.77 kcal/mol/bp), and the strain increases dramatically for PX75, PX85 and PX95. PX55 has the highest strain energy (~1.85 kcal/mol/bp) making it unstable, which is in accordance with the experimental results. We also found that PX65 has helical twist and other helical structural parameters close to the values for normal B-DNA of similar length and sequence. Vibrational mode analysis shows that compared to other PX motifs, PX65 has the smallest population of the low frequency modes which are dominant contributor for the conformational entropy of the PX DNA structures. All these results indicate that PX65 is structurally more stable compared to other PX motifs in agreement with experiments. These results should aid in designing optimized DNA structures for use in nanoscale components and devices.

Thus these previous simulations also helped explain some of the experimentally observed results in such systems. However, there was some uncertainty in our conclusions because the simulations used Na+ based salts for the counter ions, whereas the experiments were carried out using Mg+2 based salts. Indeed the experimental studies indicate that the PX structures are more stable and more reproducible when carried out in Mg++ based solutions rather than Na+

In this paper we extend these computational studies of the structural properties of the PX structures in the presence of divalent Mg++ ions rather than Na+. . Here we carried out molecular dynamics simulations in explicit water with Mg++ salt to predict the structural properties of the PX structures. This is the first time that such large DNA based nanostructures have been simulated in explicit water for such long time scales. The paper is organized as follows:

The structure building and the simulation methods are presented in section 2. The results from the molecular dynamics simulation on various PX structures are presented in section 3. Finally, a summary of the main results and the conclusions drawn from these are given in section 4.

I. Methods

a) Building and Simulation Details for the PX structures

The details of the construction procedure and simulation for the PX and JX structures has been published elsewhere [5-6]. Here we outline the basic steps involved. We first constructed two regular B-DNA structures with the base pair sequence shown in **Figure 1** and accommodated different number of base pairs per helical turn by adjusting the twist angle of a selected number of base pairs. The individual double helices were built using Namot2 (version 2.2.) [7]. The two double helices thus built in Namot2, were oriented so that the 5’ and 3’ ends of the double helices are parallel to the y-axis and the individual helices rotated so that the desired crossover points are at the closest distance to each other. Once we had identified the suitable crossover points, we created the crossovers using the “nick” and “link” commands in

Namot2. These structures were saved in the Protein Database (PDB) file format. **Figure 2** shows the computer-generated snapshot of the built PX structures.

b) Simulation Details for the PX – JX structures

All MD simulations reported in this paper used the AMBER7 software package [8] with the all-atom AMBER95 force field (FF)[9]. For Mg++ we used the Aqvist [10] interaction parameters. The electrostatics interactions were calculated with the Particle Mesh Ewald (PME) method [11-12] using a cubic B-spline interpolation of order 4 and a 10^{-4} tolerance set for the direct space sum cutoff. A real space cut off of 9Å was used both for the electrostatics and van-der Waals interactions with a non-bond list update frequency of 10.

Using the LEAP module in AMBER, the PX/JX nanostructures were immersed in a water box using the TIP3P model for water. The box dimensions were chosen in order to ensure a 10Å solvation shell around the DNA structure. In addition, some waters were replaced by Mg^{+2} counter ions to neutralize the negative charge on the phosphate groups of the backbone of the PX/JX structures. This procedure resulted in solvated structures, containing approximately 42,000 atoms in a box of dimensions 45 Å x 65 Å x 196 Å. These PX and JX structures were subjected to the equilibration protocol outline in our previous work [5-6].

II. Results and Discussion

a) Differences in Flexibility of the PX structures

MD simulations in the presence of monovalent Na+ ions and divalent Mg(+2) ions have been reported previously for the crystal structure of B-DNA to validate the AMBER force field (FF) [13], also using explicit salt and water and the particle mesh Ewald method for calculating the non-bond electrostatic interactions [14-17]. The simulations on crystalline B-DNA led to an overall

calculated CRMSD for all atoms of 1.0-1.5 Å [14-17]. For the solution phase, there are no reliable experimental structures with which to compare the simulations, which generally lead to RMSD differences of 3.6-4.2 Å from the crystal [16-17]. However, the effect of Mg(+2) ions on the structure of DNA in water has not been reported. A few simulations have been reported on the influence of Mg(+2) ions on the DNA structure [17-18]. Our present simulation study also helps us to understand the binding of Mg(+2) ions to DNA.

We carried out MD simulations for 2.5 to 3 ns in explicit salt and water for each of the five PX nanostructures (PX55, PX65, PX75, PX85, and PX95) at 300K. In each case we define an average MD structure by averaging the coordinates for various snapshots for the last 1 ns at an interval of 1 ps. This structure represents the time averaged solution structure of the PX nanostructures (that one would compare to an NMR structure). These averaged structures for various PX structures are shown in **Figure 3(a)** (in presence of Na+) and **3(b)** (in presence of Mg(+2)).

In the presence of monovalent Na+ ions the base stacking and Watson-Crick hydrogen bonding are well maintained in the solution structure for all the PX molecules. However the PX55, PX85, and PX95 structures undergo substantial writhing (as seen from the side view of the PX structures shown in **Figure 4(b)),** whereas PX65 and PX75 do not. Based on this feature PX65 and PX75 should be most suitable for creating 2-D arrays of nanostructures.

We also find that the presence of divalent Mg(+2) ions leads to significant changes in the solution properties of PX molecules compared to Na+ containing solutions. Thus Mg(+2) ions lead to a significantly distortion in the Watson-Crick hydrogen bonding among the bases and also to large bending of the double helix for PX95. Further quantitative insight into the base

pairing and other structural features is obtained from the calculated helical parameters shown in **section 3.2.**

To obtain some measure of the flexibility of these structures, **Figure 4(a) and (b)** show the CRMSD of the instantaneous PX snapshots from the time averaged solution structure as a function of time. This CRMSD was also calculated for the entire 3 ns MD runs, representing the fluctuations of a PX structure in solution. We see that PX55 and PX65 have the least fluctuations in solution with an average CRMSD of 2.0-2.5 Å over the final 1 ns both in the presence of Na+ as well as Mg(+2) ions. The other three cases, PX75, PX85 and PX95, have larger fluctuations of 2.5-3.0 Å, revealing a lower structural ordering in solution.

The CRMSD with respect to the initial minimized canonical structure is shown in **Figure 5(a)** for all the PX structures. The root-mean-square deviation in coordinates (CRMSD) from the initial canonical structure remains within 3-4 Å for PX65 over the dynamics, but it goes up to 8-12 Å for PX55, PX75, PX85 and PX95. For comparison we show the CRMSD of the PX structure in presence of Na+ ions in **figure 5(b)**. Here again CRMSD from the initial canonical structure remains within 3-4 Å for PX65 over the dynamics, but goes up to 7-8 Å the other PX structures. This indicates the intrinsic stability of PX65 structure compared to other PX structures, which agrees with the experimental results.

b) Comparison of the helicoidal parameters for the PX nanostructures

More critical structural characterization of the PX/JX DNA nanostructure can be made by calculating various helicoidal structural parameters such as roll, rise, tilt, and twist. These parameters were calculated for each base pair, averaged over last 400 ps of the 3 ns dynamics run. The Curve algorithm [19-20] was used to calculate the various helicoidal parameters.

Divalent cations like Mg(+2) are known to have sequence specific binding to the DNA and to affect the major/minor groove binding properties [21]. They are also known to produce significant bending of the double helical geometry. **Figures 6 shows** the rise, tilt, roll, and twist calculated for every base pair for the PX65 structure in the presence of Mg(+2) ions. The helical twist angle for the two double helices of the PX65 fluctuates around 31° and the base tilt angle fluctuates around 0.31° (essentially zero). These values are close to the values (30° for twist and 0.22° for tilt) obtained from simulation of the two separated double helices of PX65 [5]. Thus the helical parameters for PX65 are close to that of B-DNA double helix and hence PX65 should be a very stable structure like a B-DNA. On the other hand, for the PX55, PX75, PX85 and PX95 structures the helical twist and tilt angles show large fluctuations about the corresponding values in B-DNA. These fluctuations are especially large at the crossover points. The presence of Mg(+2) has resulted in significant distortion of the PX95 structure as is evident from various helicoidal parameters.

Figures 7(a) and 7(b) show the variation of the width of the major and minor grooves in each of the two double helices for PX65 in presence of Na+ and Mg(+2) respectively.. The major and minor grooves show a steady increase in the width in going from the PX55 to the PX95 structure. The average major groove width for a B-DNA is around 11.7 Å, which can widen to 15 Å on binding a protein or drug [22]. The minor groove width in B-DNA is 5.7 Å. The PX55 structure shows average width of 12.9 Å for major groove and 5.5 Å for minor groove, which is close to the values for B-DNA. These values increase for PX65 [13.9 Å for major groove and 6.4 Å for minor groove] eventually going up to 15-17 Å for the major groove of PX95. The instantaneous

major groove width deviates significantly from the average value for each nucleotide, especially at the crossover points. We see the decrease in the major groove width in presence of Mg(+2) ions.

c) Macroscopic structural properties of PX nanostructures

We calculated the macroscopic structural features such as writhing, overall bending, and the solvent accessible surface area of the PX structures. We also analyzed the vibrational modes of the PX structures to understand the relation between the low frequency modes and structural stability. These properties throw light on the nature of the PX nanostructures.

i. Writhing in longer PX DNA nanostructures

Figure 8 shows the variation of "strand shortening" for various PX structures averaged over the last 200 ps of the 3 ns MD simulation runs. Strand shortening is calculated as follows: The Curve algorithm outputs the vectorial direction of each local helical axis segment *U* and its reference point *P*. The path length between successive helical axis reference points can be calculated as

$$path = \sum_{i} \left| \vec{P}_{i} - \vec{P}_{i-1} \right|$$

and the end-to-end distance of the DNA fragment can be calculated as

$$R_e = \left| \vec{P}_1 - \vec{P}_N \right|$$

where $\vec{P}_1$ and $\vec{P}_N$ are the reference points for the two end helical axis corresponding to two terminal nucleotides. The difference between sum of all the path lengths and the end-to-end distance is a measure of the strand shortening. The strand shortening also indicates the overall flexibility of the DNA.

Figure 8 shows that the end-to-end distance decreases or the strand shortens more as the number of base pairs increase in the PX structures, indicating that strand shortening is highest for PX95. In the presence of Mg(+2) strands shorten much stronger than in the presence of Na+ ions. This could be rationalized by the fact that presence of Mg(+2) contributes to the overall charge neutralization of the phosphate backbone thereby making the helix more flexible. Also there is substantial distortion (e.g., writhing and/or bending) in the overall PX95 structure. This effect could be due to the total length or to the 95-crossover motif that places 14 base pairs in one helical turn or due to base sequence which is intrinsically unstable.

We have also calculated global helical bending for each of the two helices using the algorithm developed by Strahs and Schlick [23]. This method computes the DNA curvature by summing the projected components of local base pair step tilt and roll angles after adjusting the helical twist. Our analysis for the global angles is based on the values of local tilt and roll angles for each base pair step computed by the Curves program [24]. **Figure 9** shows the global bending for each PX molecule. The curvature of the double helix axis is similar for both helices in the PX55 molecule. On the other hand the curvature of the two helical axes show a bending angle differing by 10-40° for the PX65, PX75, PX85, and PX95 structures, indicating the effect of writhing of the helical axis for these structures. Combining the effect of strand shortening with the bending, we infer that PX95 shows a larger writhing in its solution structure compared to the other PX structures. The effect of writhing is likely to be an important structural feature in designing nanostructures. For example, because of the writhing PX95 may not be a good choice for constructing a 2-D array using PX nanostructures.

d) Relative Stability of the PX/JX nanostructure

i. Relative stability of the JX structures:

The PX DNA is a four-stranded molecule in which two parallel double helices are joined by reciprocal exchange of strands at every point at which the strands come together [1-2]} The JXM structure is related to PX by containing M adjacent sites where backbones of the two parallel double helices *juxtapose without crossing over*. Earlier we used Molecular dynamics simulation to demonstrate [5] that in the presence of the Na+ ions, JX motifs are not able to maintain the parallel double helix crossover structures. Thus the two helical domains of the crossover structure move increasingly far apart with decreasing number of crossover points. However, experimentally JX molecules are found to form in the presence of Mg(+2) ions. To test the stability of the JX stricture we have calculated [25] the relative stability of the JX structure as a function of crossover points in the presence of Mg(+2) ions and find that Mg(+2) ion induces an effective attraction between the two helical domains to maintain their crossover structures.

ii. Solvent accessibility surface of PX structures:

The thermodynamic stability of the PX/JX nanostructure is also greatly influenced by its interaction with the surrounding solvent medium as well as its interaction with the counterions. The solvent accessible surface area (SASA) gives useful insight into the nature of interaction of the PX motifs with the water. The SASA of the PX structures was calculated using the Analytical volume generalized Born (AVGB) method developed in Goddard group [26]. AVGB is very fast and accurate and has been applied successfully to study solvation effects in biological systems [27]. **Figures 10(a) and (b)** shows the SASA of the PX molecules in the presence of Na(+) and Mg(+2)

respectively. In the presence of Na(+) SASA increases from PX55 to PX75 but decreases for PX85 and then increases further for PX95. This observation is consistent with the number of water molecules at the surface of each PX structure shown in **Figures 11(a)**. **Figure 11(a)** shows that the number of water molecules in the first solvation shell of the PX structures increases steadily from PX55 to PX95 with a break at PX85. This could be caused by local kinks present in the PX85 structure although the overall bending of PX85 is not as high as PX95. The presence of Mg(+2) apparently reduces the number of water in the vicinity of the DNA as the degree of bending of the helical axis increases compared to case when Na(+) ions are present.

III. Summary and Conclusions

We have demonstrated that MD simulations over several ns provide critical information on the structural features and relative stability of the various DNA motifs that provide the building blocks for DNA based nanostructures. Our present study fills the gap in the existing literature to provide a comprehensive understanding of the effect of binding of the Mg(+2) ions on the DNA structural properties. Our study clearly demonstrates how the binding of the Mg(+2) ions affect the structural properties of DNA nanostructures. Calculated CRMSD values along with various helicoidal parameters suggest that PX65 molecule is the most stable molecule in accordance with the experimental findings. Our strain energy analysis based on the nearest neighbor interactions model shows that the molecule with a higher number of crossovers has a higher thermodynamic stability, but that Mg(+2) ions help stabilize those JX structures with fewer crossover points. .

IV. References

(1) Yan, H.; Zhang, X. P.; Shen, Z. Y.; Seeman, N. C. *Nature* **2002**, *415*, 62.

(2) Shen, Z. Y.; Yan, H.; Wang, T.; Seeman, N. C. *Journal of the American Chemical Society* **2004**, *126*, 1666.

(3) Seeman, N. C. *Biochemistry* **2003**, *42*, 7259.

(4) Fu, T. J.; Seeman, N. C. *Biochemistry* **1993**, *32*, 3211.

(5) Maiti, P. K.; Pascal, T. A.; Vaidehi, N.; Goddard, W. A. *NUCLEIC ACIDS RES* **2004**, *32*, 6047.

(6) Maiti, P. K.; Pascal, T. A.; Vaidehi, N.; Heo, J.; and Goddard, W. A. *Biophys. J.* **2006**, *90*, 1463.

(7) Tung, C. S.; Carter, E. S. *CABIOS* **1994**, *10:*, 427.

(8) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E.; Wang, J.; Ross, W. S.; Simmerling, C.; Darden, T.; Merz, K. M.; Stanton, R. V.; al., e.; 7 ed.; University of California, San Francisco.: 1999.

(9) Cornell, W. D., et al. *J. Am. Chem. Soc.* **1995**, *117*, 5179.

(10) Aqvist, J. *J. Phys. Chem. B* **1990**, *94*, 8021.

(11) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.

(12) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577.

(13) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *Journal of the American Chemical Society* **1995**, *117*, 5179.

(14) Lee, H.; Darden, T.; Pedersen, L. *Chem. Phys. Lett.* **1995**, *243*, 229.

(15) York, D. M.; Yang, W. T.; Lee, H.; Darden, T.; Pedersen, L. G. *Journal of the American Chemical Society* **1995**, *117*, 5001.

(16) Miaskiewicz, K.; Miller, J.; Cooney, M.; Osman, R. *Journal of the American Chemical Society* **1996**, *118*, 9156.

(17) Bevan, D. R.; Li, L. P.; Pedersen, L. G.; Darden, T. A. *Biophys. J.* **2000**, *78*, 668.

(18) MacKerell, A. D. *J. Phys. Chem. B* **1997**, *101*, 646.

(19) Lavery, R.; Sklenar, H. *J. Biomol. Struct. Dyn.* **1989**, *6*, 655.

(20) Lavery, R.; Sklenar, H. *J. Biomol. Struct. Dyn.* **1989**, *6*, 63.

(21) Chiu, T. K.; Dickerson, R. E. *J. Mol. Biol.* **2000**, *301*, 915.

(22) Schumacher, M. A.; Miller, M. C.; Grkovic, S.; Brown, M. H.; Skurray, R. A.; Brennan, R. G. *Embo Journal* **2002**, *21*, 1210.

(23) Strahs, D.; Schlick, T. *J. Mol. Biol.* **2000**, *301*, 643.

(24) Lavery, R.; Sklenar, H. *J. Biomol. Struct. Dyn.* **1988**, *6*, 63.

(25) Maiti, P. K.; Pascal, T. A.; Vaidehi, N.; and Goddard, W. A. I. *Nanotechnology* **2006**.

(26) Zamanakos, G., Caltech, 2002.

(27) Vaidehi, N.; Floriano, W. B.; Trabanino, R.; Hall, S. E.; Freddolino, P.; Choi, E. J.; Zamanakos, G.; Goddard, W. A. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12622.

V. Figures

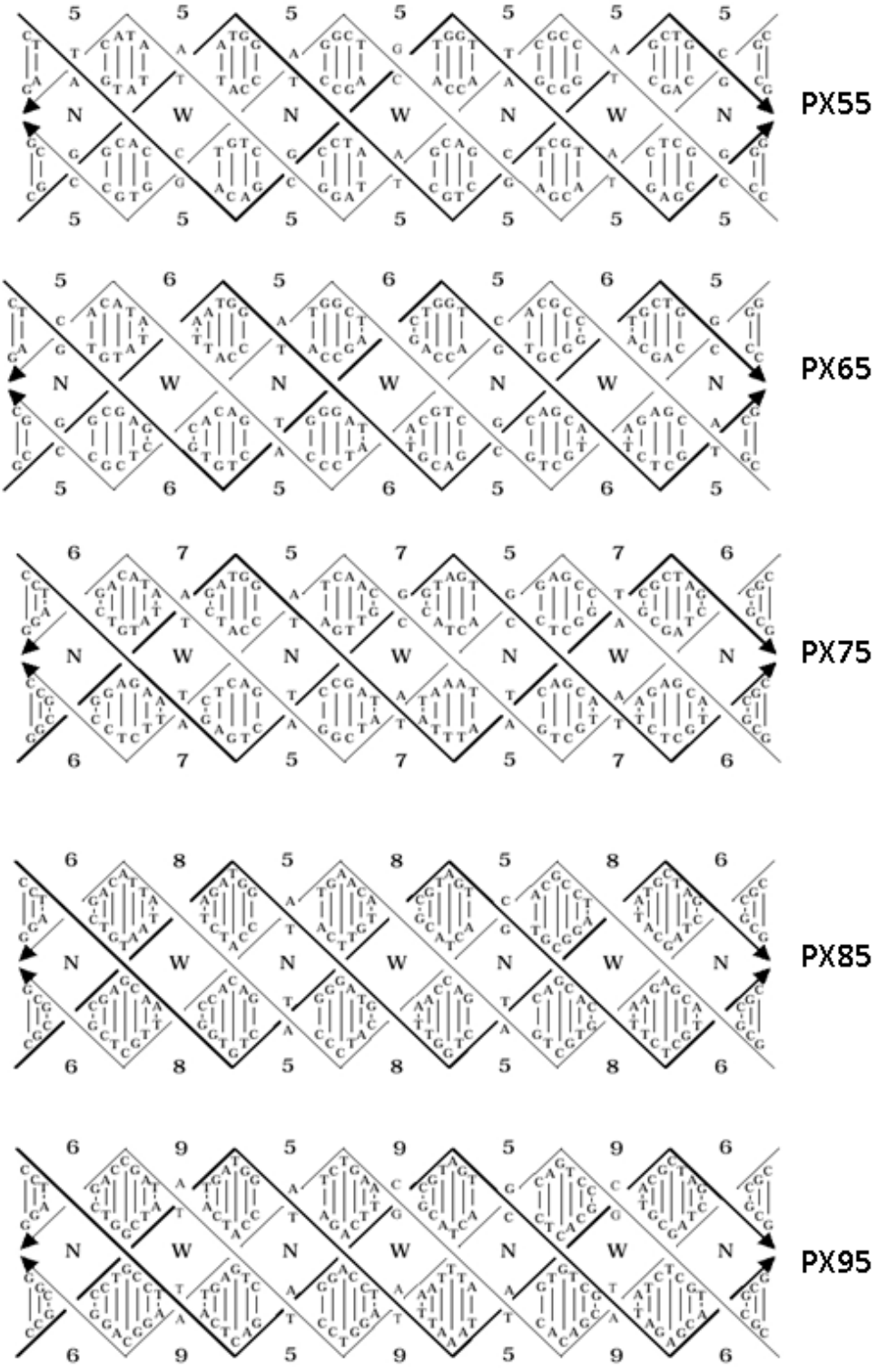


Figure 1 The base pair sequences used in the generations of PX55, PX65, PX75 PX85, and PX95.

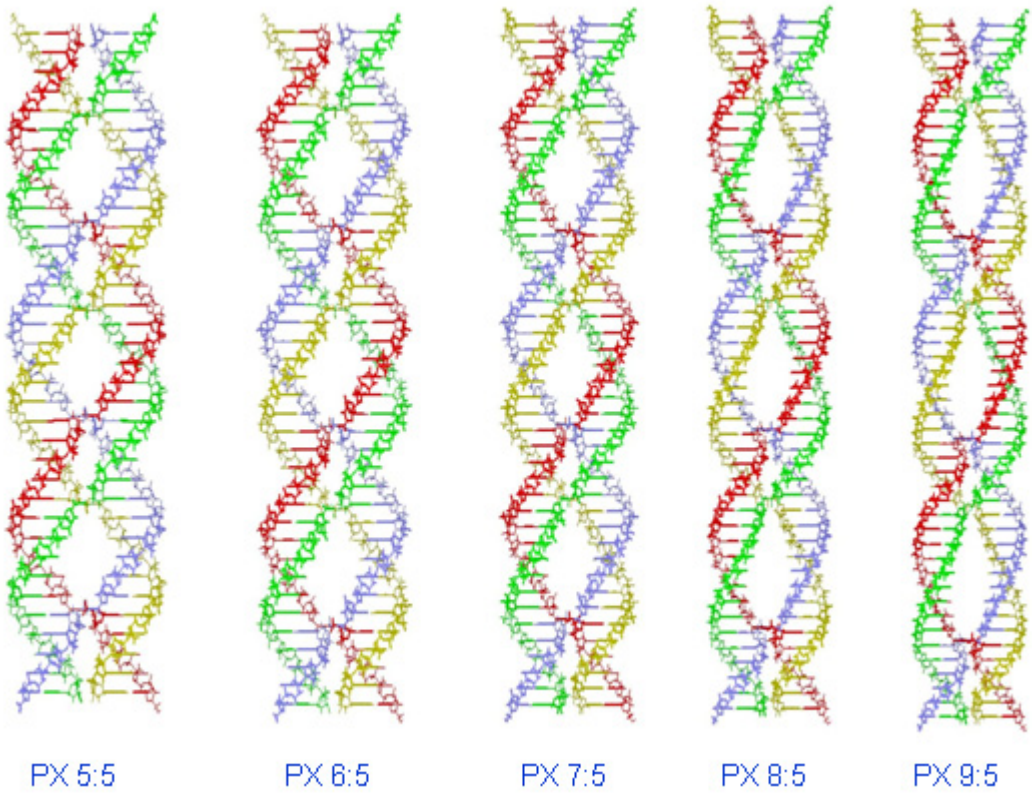


Figure 2 Computer generated built initial structure for PX55, PX65, PX75 PX85, and PX95.

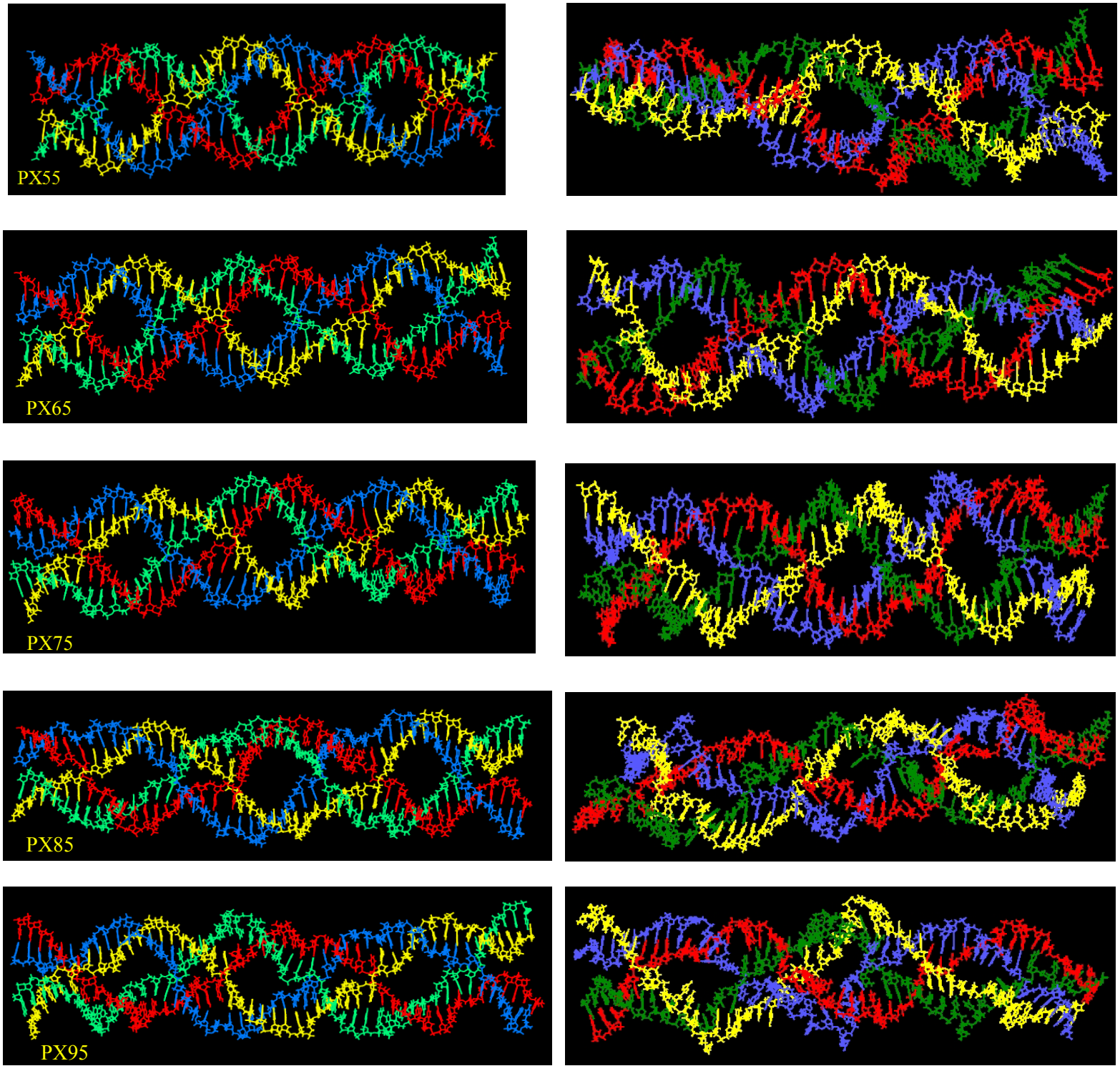


Figure 3 structures for various PX molecules averaged over the last 10 ns of molecular dynamics **(a)** in presence of Na(+) and **(b)** in presence of Mg(+2). For clarity water molecules and counter ions are not shown.

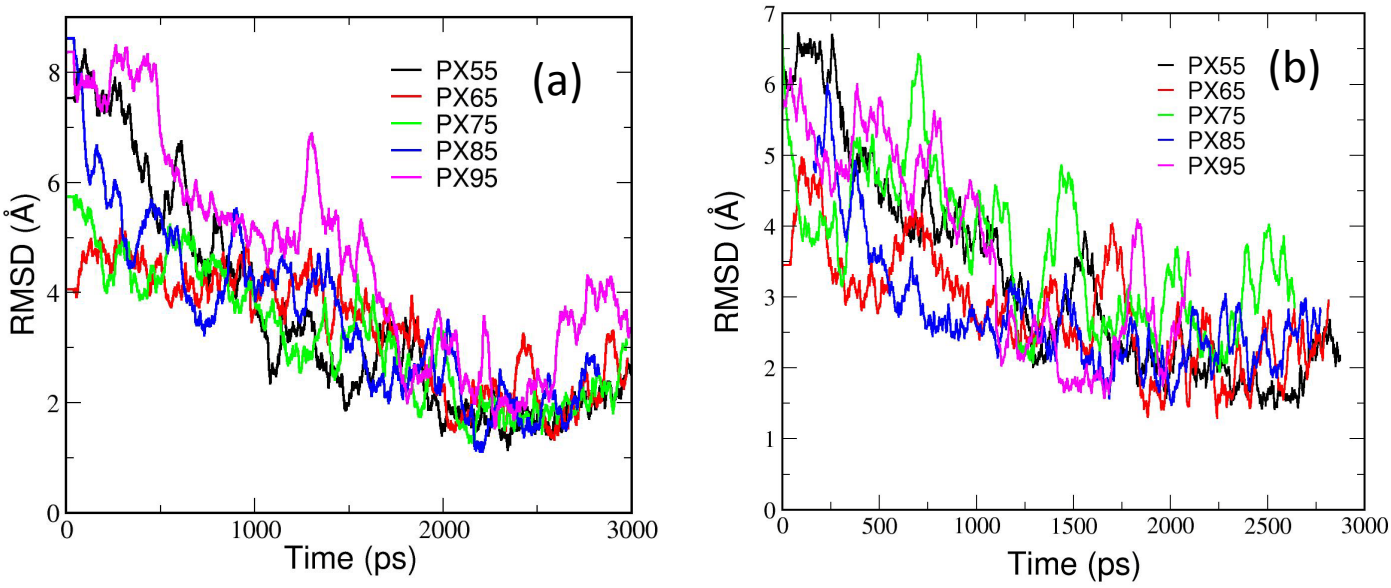


Figure 4 Variation of the coordinate rms displacement (CRMSD) of all atoms of various snapshots from the MD simulation run with respect to the average dynamics structure **(a)** in presence of Na+ and **(b)** in presence of Mg(+2)

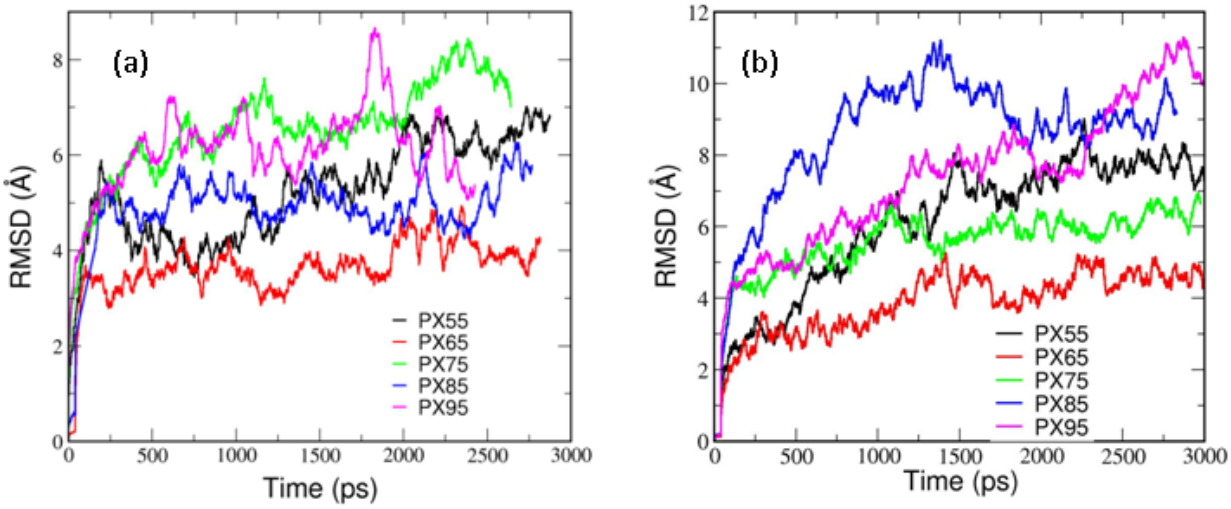


Figure 5 CRMSD for individual bases for PX55, PX65 PX75, PX85 and PX95 from the starting dynamics structure, in **(a)** Na+ and **(b)** Mg(+2). The data has been averaged for last 200 ps of the dynamics run.

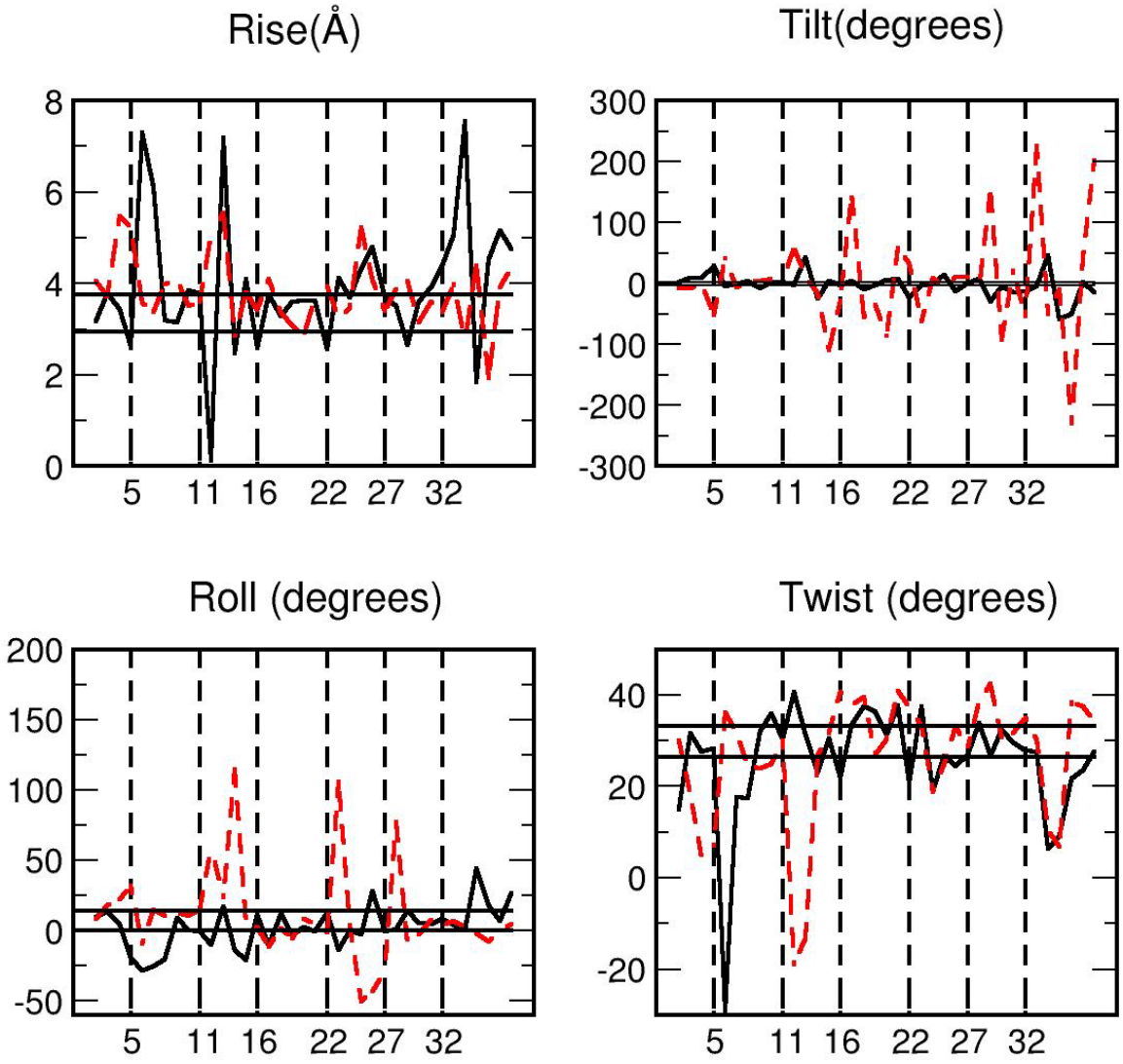
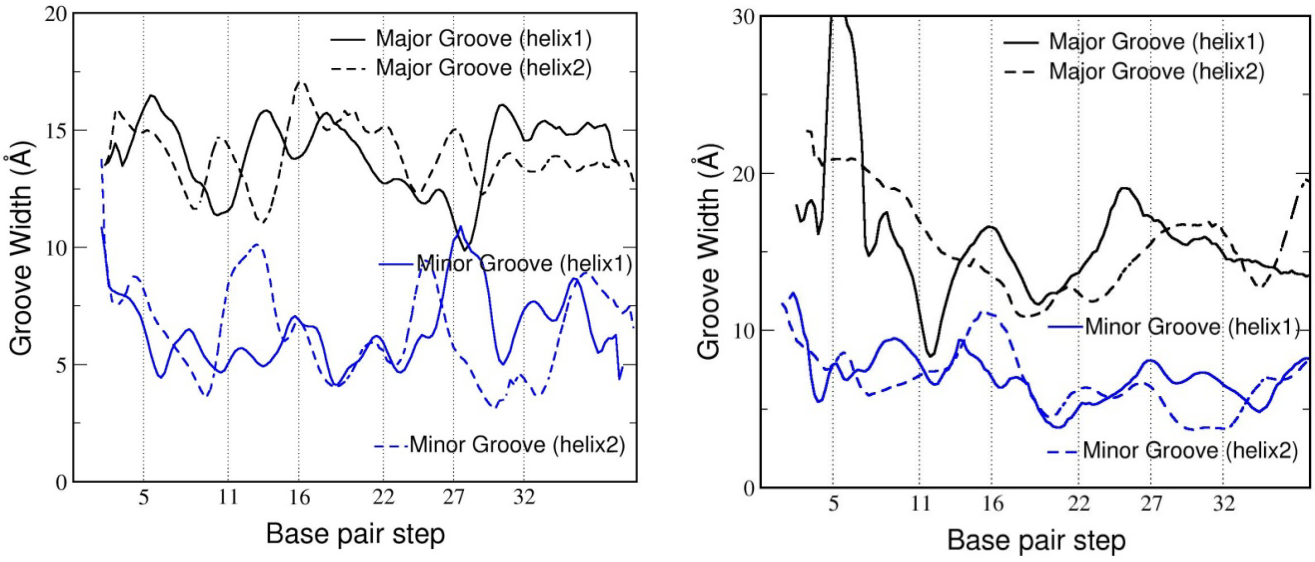


Figure 6 Average Rise, Tilt, Roll, and twist for for PX55, PX65 PX75, PX85 and PX95 from the starting dynamics structure, in **(a)** Na+ and **(b)** Mg(+2). Solid line is for helix1 and broken line is for helix2. The vertical lines corresponds to the crossover points. The data has been averagred over last 200 ps of the 3ns long dynamics. The horizontal solid lines gives the upper bound and lower bound for the corresponding quantities expexted for the helices in their B-DNA form (non-crossover form) during the dynamics. This shows that PX65 retains the B DNA much better than PX95..



a **b**

Figure 7 Average major groove width and minor groove width for PX65 in presence of Mg(+2) Solid black and blue lines represent major groove and minor groove width respectively for helix1. Broken black and blue lines is for helix2. The data have been averagred over last 200 ps of the 3ns long dynamics

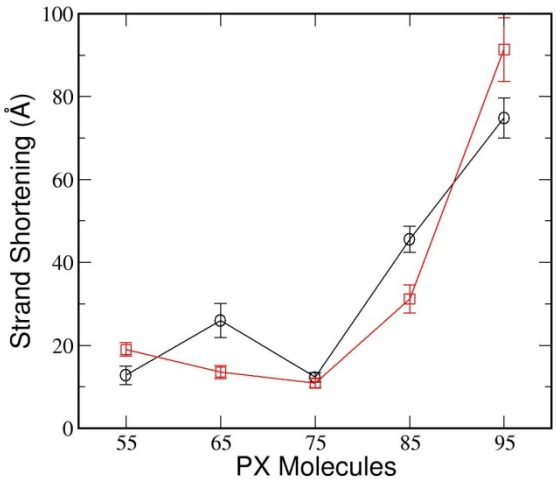


Figure 8 Strand shortening as defined in text for various PX molecules in **(a)** Na+ and **(b)** Mg(+2). The data have been averaged over last 200 ps of the dynamics run.

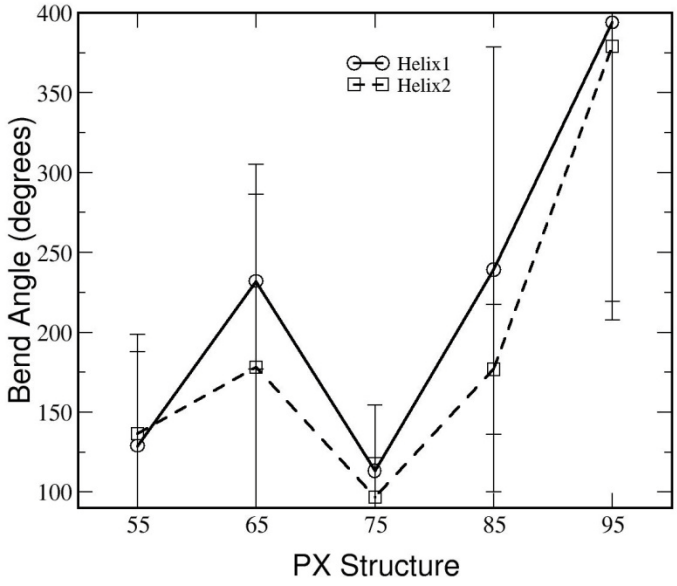


Figure 9 Global bend angle calculated for each double helix of each PX structure. The error bars indicate the fluctuations that occur in the molecular dynamics simulation.

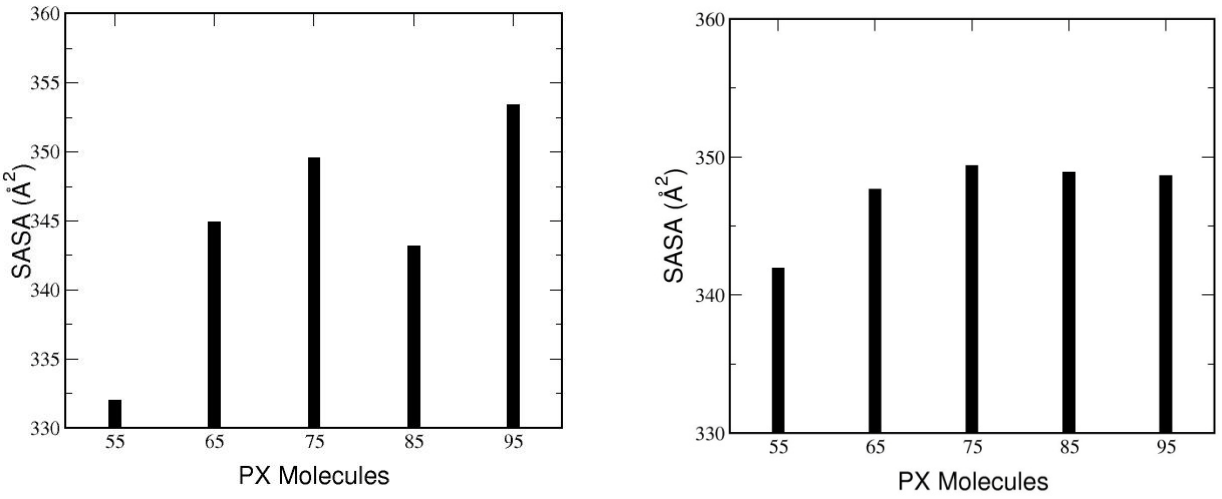


Figure 10 Solvent accessible surface area (SASA) per base pair for various PX molecules. The area has been avergaed over last 200 snapshots of the dynamics. The area has been calculated by a very fast and accurate AVGB alogorithm [1]

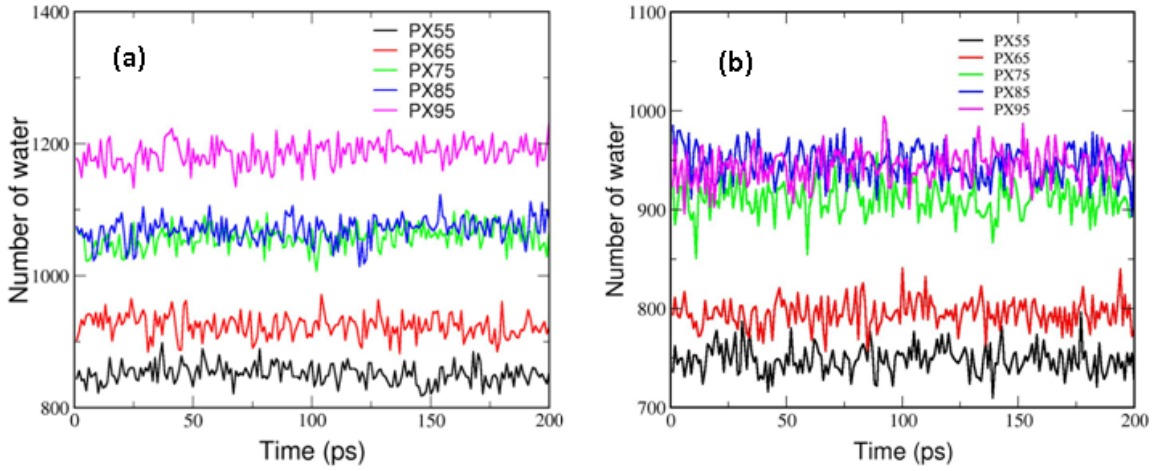


Figure 11 Number of water in the first solvation shell (within a shell of 2 Å).

(1) Zamanakos, G., Caltech, 2002.

SECTION IV. EXPLORING THE HYDROPHOBIC EFFECT

Chapter 4. The QM-FF force field for Carbon

Recently there has been substantial progress in developing DFT methods (M06, XYG3) sufficiently accurate to provide reliable results for weakly bond systems, dominated by London dispersion forces. This has motivates us to initiate a program to develop a new generation of forcefields based soley on accurate QM data. In this paper, we develop this QM-FF forcefield for carbon and validate it by comparison to the mechanical and thermodynamics properties of graphite, which provides a strong test of the quality of the forcefield because of the dominance of London dispersions for the inner-plane interactions.

There has been considerable computation effort expended towards obtaining analytic potentials to graphitic systems[1-18]. The methods employed are as varied as the potentials themselves, ranging from so called *ab-initio* methods[2-5], semi-continuum[6-8] methods, semi-empirical methods[19] to empirical parametric bond order methods[9-18]. Tersoff-Brenner (TB) potentials[14-15] have recently been shown to produce reproduce the equilibrium lattice constants and cohesive energy of graphene, but have been limited in their inability to reproduce the experimentally observed elastic constants. Recently, Tewary and Yang[11] extended the TB model by adding in a radial term in order to reproduce the correct elastic constants. Finally, Perebeinos and Tersoff[18] improved on the TB potential to calculate the phonon spectrum of graphene and the nanotubes breathing-mode energy, with good comparison to experiment. Their model however still relies on some information related to the under lattice structure of the system being considered. We are interested in predicting the bulk properties of graphitic systems[20],

and accurately modeling the interface with solvents[21], biological molecules[22-23] and polymers[24], using smooth two-body potentials. Thus, approaches that rely on non differentiable radial terms or knowledge of the underlying structure of the system being considered are not suitable for our purposes.

Many common two body forcefield, such as AMBER[25-28], CHARMM[29-31], OPLS[32], MM2/MM3[33-34], and DREIDING[35], use generalized sp^2 carbon atom types for describing graphitic systems, in particular benzene. While these forcefields' graphite parameters can be shown to be applicable for discerning useful trends in these systems, they cannot provide the necessary resolution to calculate more sensitive properties (such as elastic constants and phonon modes) adequately[36]. The recent interest in graphene[37-39] and carbon nano-ribbons[40-41] as the foundation for future electronics underscore the need to obtain accurate forcefields that fit into established Molecular Mechanics/Dynamics and Monte-Carlo codes, facilitating the investigation of these systems on time and length scales not available to *ab-initio* methods. In these systems, theoretical investigations may be the only tool to uncover the underlying physics.

To this end, there have been several attempts to develop more accurate two body potentials for graphitic systems based on QM and/or available experimental data. Our previous effort[20] treated the nonbonded interactions using a Lennard-Jones (LJ) 12-6 potential, and correctly predicted the correct heat of formation of C_{60} and C_{70} crystals. Girifalco et al. produced a universal LJ potential for describing graphitic systems[42] and Ulbricht et al.[43] used thermal desorption spectroscopy and C60 interacting with graphite to obtain a LJ potential. We will show that the last two potentials do not reproduce the

elastic constants of graphite, and therefore cannot be expected to accurately determine the thermodynamic properties of the bulk system.

We have developed a new forcefield (QM-FF) with analytic two body potentials describing the nonbonds from accurate electronic structure calculations on the coronene ($C_{24}H_{12}$) dimer. The spherically symmetric coronene ($C_{24}H_{12}$) is a polycyclic aromatic hydrocarbons[44] (PAH) made of seven benzene rings and frequently used as a computational model for graphite[45-51]. PAH molecules are planar molecules, comprised of fused aromatic rings in various orientations: from elongated naphthalene ($C_{10}H_8$) and anthracene ($C_{14}H_{10}$), to symmetric pyrene ($C_{16}H_{10}$). Graphene can therefore be thought of as the ultimate PAH, existing as an infinite 2-D sheet of benzene units. PAHs are known carcinogens[52] and are ubiquitous; occurring in most organic substances[53-54], in significant quantities in fossil fuels[55], and has even been implicated in the earliest interstellar and planetary processes[56-58]. Recently, Zarcharia et al.[59] experimentally determined the cohesive energy (per carbon atom) of graphite to be 1.27±0.1 kcal/mol by studying the interaction of various PAHs with the basal plane of graphite.

Lee and Kim[60] developed a potential for PAH interactions based on Møller-Plesset second-order perturbation theory (MP2)[61] theory using PAHs up to pyrene, finding that their C-C interactions were twice as large as other reported values. This is not surprising, as the propensity of MP2 calculations to systematically overbind is well documented[62-64]. Density-functional theory[65-66] (DFT) based methods are more efficient, but traditionally suffer from an inability to accurately describe nonbonded van der Waals interactions[67-69]. Means of correcting this deficiency range from adding empirical

VDW terms[64], performing calculations with excitation to virtual orbitals[70] and mixing in Hartree-Fock (HF) exchange terms (M0X[71-74] series of functionals).

The M06-2X[74] functional has been shown to be accurate at describing weakly interacting, dispersion dominated systems. Further, Zhao and Truhlar[75] recently mapped the potential energy surface of the coronene dimer for the six most energetically favorable configurations, using the M06-2X DFT functional. We extend their calculations to find the distance dependence of the two most relevant configurations to graphite: parallel displaced along the *x* axis (PD-X) and parallel displaced along the *y* axis (PD-Y). We use the coulomb corrected dispersion curves, combined with the two experimental lattice vibrations, to develop a classical force field (QM-FF) that reproduces the elastic constants and phonon modes of graphite. Using QM-FF various thermodynamic quantities of the graphite crystal are calculated from 1 – 2500K.

The remainder of the paper is organized as follows: in Section II we outline the selection of experimental values and corrections to 0 K. Section III describes the process used in obtaining the force field for graphite. In Section IV we discuss the predictions of the less well characterized properties (e.g., C_{44}, C_{13}). In Section V we report a more detailed analysis of properties for graphite based on the predicted force field. This includes phonon dispersion curves and prediction of thermodynamic properties. In Section VI we consider the discrepancy between experimental values of C_{44} (which differ by a factor of 15) and the comparison with available theory.

II. Experimental Data for Graphite

a) Crystal Structure

The space group of graphite was taken as $P6_3/mmc$ (D_{6h}^{14}), which assumes flat layers. The lattice parameters at 25°C are[76] a = 2.4612 Å, c= 6.7090 Å. To convert the lattice parameters at low temperature, we used the thermal expansion data from Bailey and Yates[77] (20-270 K), extrapolated and interpolated to obtain

$$\delta a = -0.0072 \text{ Å}$$
$$\delta c = +0.0369 \text{ Å}$$

for 0 K to 298 K. This leads to

$$a(O\ K) = 2.4684 \text{Å}$$
$$c(O\ K) = 6.6721 \text{ Å}$$

b) Elastic Constants

A complete study of the elastic constants of compression annealed pyrolytic graphite (at room temperature) was carried out by Blakslee et al.[78] who find elastic constants (stiffness), Young's moduli, and Poisson ratios as listed in **Table S1** of supplementary materials. The elastic constants C_{11} and C_{12} relate to in-plane deformations, C_{33} is a direct measure of the force constant for the C – C van der Waals interactions, and C_{44} relates to shear of one plane with respect to the next. The quantities C_{11}, C_{12}, C_{33}, and C_{44} are derived from direct experiments; however, C_{13} (which involves coupling of stress in the plane to the spacing between the planes) is indirect and rather uncertain (33% quoted uncertainty).

The biggest variations in the literature on graphite occur for C_{44} where values from 0.18 to 0.35 GPa are obtained[78] from mechanical studies on compression annealed samples (the values were independent of external compression), while neutron

scattering studies[79-80] lead to values of[79] C_{44} = 4.2 ± 0.2 and 4.6 ± 0.2, or 15 times larger. The neutron-irradiated samples are referred to as "dislocation-free graphite", since it is believed that neutron damage (or impregnation with boron) impedes dislocation motion[79,81-82]. Further, by considering the specific heat at low temperatures, Komatsu[83-84] determined the C_{44} of natural graphite to be 4.05 and 4.52, while Bowman and Krumhansl[85] calculated the value as being 2.3. The measured value of the C_{44} in graphite therefore has not been unambiguously resolved. Our calculations agree with the low values from the mechanical studies (see discussion in Sections IV.B and VI).

The temperature dependence of the elastic constants was measured by Gauster and Fritz[86] on compression-annealed graphite from 4 to 300 K, yielding changes as listed in **Table S1** of supplementary materials. The values for δC_{11}, δC_{12}, δC_{33}, and δC_{44} are obtained from independent experiments; however, δC_{13} is obtained from a complex mode after subtracting the contributions from the other quantities (see Section IV.C). As a result, δC_{13} has an uncertainty comparable to its magnitude. Combining the above results leads to the total values in **Table S1**.

c) Lattice Vibrations

With four atoms per unit cell, graphite has 12 vibrational branches. For the Γ point of the Brillouin zone (k = 0), this leads to the following 12 modes[87]:

- E_{1u}, A_{2u}: transverse (TA) and longitudinal (LA) acoustic (at 0 cm^{-1})
- E_{2g} , B_{2g} : transverse (TO) and longitudinal (LO) modes for sheet-sheet interactions (at ~10 cm^{-1} and ~140 cm^{-1})
- A_{2u}, B_{2g} : out-of-plane (ZO) crinkling of graphite (at ~868 cm^{-1})

- E_{1u},E_{2g}: in-plane ring modes (at ~1588 cm^{-1})

The 868 cm^{-1} mode (A_{2U}) was used to determine the torsional parameters for our force field. The 1588 cm^{-1} mode (E_{1u}) was used to help determine the in-plane force constants. The sheet-sheet modes (E_{2g}, B_{2g} or TO, LO) were also obtained from neutron scattering studies on "dislocation-free" graphite at room temperature, yielding[79] 45 cm^{-1} and 126 cm^{-1} for the TO (E_{2g}) and LO (B_{2g}) transition at the Γ point. These neutron scattering studies lead to elastic constants of C_{33} = 37.1 ± 0.5 GPa, C_{44} = 4.6 ± 0.2 GPa, and C_{11} = 1440 ± 200 GPa. The value for C_{33} is consistent with the mechanical studies (36.5), but the other values are significantly higher than mechanical results (C_{44} = 0.27 and C_{11} = 1060). We believe that this may be due to the neutron damage and have not used the neutron scattering values for the TO and LO modes in our fits.

d) Thermal Properties

The specific heat of graphite has been extensively studied, with values obtained for both pure and impure graphite crystals (see the work of Nihira and Iwata[88] and the references therein). The specific heat is reported as 1.92 – 2.06[89] cal/mol/K at 298K. There is an almost linear increase with temperature up to 1500K, and then a plateau to approximately 0.53 kcal/mol/K. Below room temperature, the specific heat is sensitive to the purity of the graphite, the size of the crystal and the differences in stacking faults[88]. Above room temperature, it is believed to be relatively insensitive to any of these differences, with reported attributed to experimental variations.

In testing the accuracy off the QM-FF X6 potential, we use a variety of experimental data. For low temperatures (< 300K) we use the data of Desorbo and

Tyler[90] on pile graphite, Canadian natural graphite and graphitized lampblack. From 300 – 500K, we used the JANAF[91] and data from Hultgren[92] from 750K – 2500K. A comparison of the experimental values and those calculated by the X6 QM-FF is shown **Figure 7** and **Table S5** of supplementary materials. To convert the calculated values of C_v to C_p, we used the thermal expansion data of Bailey and Yates[77].

There is a non-linear, temperature dependent increase in the carbon atom interatomic spacing in graphite. This expansion is highly anisotropic: low in-the-plane (the *ab* direction) and an order of magnitude larger out-of-the-plane (the *c*) direction[80]. In fact, there is an observed thermal *contraction* in the *ab* direction from 0 – 400K, presumably due to the internal stress resulting from rapid expansion in the *c* direction (the Poisson effect)[93]. This leads to a coefficient of thermal expansion of 0.09 and 0.38 deg^{-1} K at 30K for the ab and c directions respectively and 1.27 and 2.72 at room temperature. The calculated in-plane lattice constants and the corresponding coefficient of thermal expansion is compared to available experimental data in **Figure 8**. The out of plane lattice constants is recorded is compared to experiment in **Figure S3** of supplementary materials.

III. Calculations

a) Electronic structure calculations on the Coronene dimer

The crystal structure for the three coronene structures were obtained from the supplementary material of the recent paper of Zhao and Truhlar[75]. Starting with these structures, we carried out M06-2X[74] calculations of the coronene dimer using the Jaguar7.0[94] ab-initio package, with the functional. M06-2X is a hybrid meta-GGA

exchange-correlation functional with 54% HF exchange that has be shown to describe weakly bonded dispersion effect rather impressively[74] around the van der Waals minima.

Three sets of coronene structures can be considered relevant for describing the electronic structure of graphite: eclipsed, parallel displaced along the x axis (PD-X) and parallel displaced along the y axis (PD-Y) (**Figure 1**). Of these, the PD-X and PD-Y cases were used to obtain the van der Waals parameters for carbon.

We now describe our procedure for obtaining the dispersion curve of coronene. The z displacement of the dimer is discretely ($\Delta z = 0.01$Å) varied and the geometry optimized at that distance spring (energy tolerance $5x10^{-5}$ hartrees, rms density matrix change $5x10^{-6}$). We constrain the center of mass of each monomer with a 500 kcal/mol harmonic spring, with the augmented polarized triple-ζ 6-311+G(2df,2p)[95] in all DFT calculations. We removed any bias due to basis set superposition error (BSSE) by performing counterpoise corrections[96-97] to all our stated energies, and take this value to be the correct interaction energy (**Table S3** of supplementary materials).

We consider the binding energies obtained above to be a combination of electrostatic (coulombic) and dispersion (van der Waals) terms. The purely electrostatic components were obtained from a single-point, molecular mechanics calculation, using the atom-centered point charges of a single coronene unit. We derived these charges from the electrostatic potential charge model, using the slightly smaller 6-311G(2df,2p) basis set. The resulting binding curve (without the electrostatic interactions) was taken to be the true dispersion curve of the coronene dimer. The charges on the monomer

unit used in this study, as well as the charges on the dimer are shown in **Table S2** of supplementary materials. Also recorded is a comparison with two other charge models: Mulliken population analysis[98] and Charge Model 4 (CM4)[99].

b) The Force Field

i. ***van der Waals Terms.***

We considered the Exponential-6 potential (X6) for describing the van der Waals interactions:

$$E = Ae^{-CR} - {}^{b}/_{R^6} = \frac{D_v}{(\zeta - 6)}\{6e^{\zeta(1-\rho)} - \zeta\rho^{-6}\} \quad (1)$$

where D_v is the well depth,

$$\rho = \frac{R}{R_v} \quad (2)$$

is the scaled distance, R_v is the well radius (distance between the carbons at the well minimum), and ζ is a dimensionless parameter. Equation (1) leads to a force constant at the minimum of the form

$$\mathrm{k}_v = \left(\frac{\delta^2 E}{\delta R^2}\right)_{R_v} = \frac{6\zeta(\zeta - 7)D_v}{(\zeta - 6)R_v^2} \quad (3)$$

so that ζ is directly related to the dimensionless force constant,

$$\bar{\kappa} = \frac{\mathrm{k}_v R_v^2}{D_v} = \frac{6\zeta(\zeta - 7)}{(\zeta - 6)} \quad (4)$$

We also considered two alternate potentials for describing the van der Waals interactions: the Lennard-Jones potential

$$E = D_v[\rho^{-12} - 2\rho^{-6}] \quad (5)$$

where ρ is given by (2), leading to $\bar{\kappa} = 72$

and the Morse potential

$$E = D_v\left[e^{-\alpha_v(R-R_v)} - 1\right]^2 \quad (6)$$

where α_v is the dimensionless curvature parameters.

ii. Valence Terms.

The valence interactions were described using:

A Morse potential ($E_b = D_b\left[e^{-\alpha_b(R-R_b)} - 1\right]^2$) for the C-C bond stretch,

where $k_v = 2D_v\alpha_v^2$ is the force constant.

A cosine angle bend with angle-stretch *($k_{r\theta}$)* and stretch-stretch (k_{rr}) coupling

$$E_a = \frac{1}{2}[cos\theta - cos\ \theta_a]^2 + k_{rr}(R_1 - R_b)(R_2 - R_b) + D(cos\theta - cos\ \theta_a)[(R_1 - R_b) + (R_2 - R_b)] \quad (7)$$

for the C-C-C interaction, where ϑ_a is the equilibrium angle, $k_{\theta\theta} = Csin^2\theta_a$ is the diagonal angle force constant, and $k_{r\theta} = -Dsin\theta_a$ is the angle-stretch force constant.

- **A two-fold torsion**

$$E_t = \frac{1}{2}V_t(1 - cos2\theta) \quad (8)$$

where Φ is the torsional angle

V_t is the barrier and the minimum is for Φ = 0 (planar).

c) Fitting Procedure

The M06 class of DFT functional have been shown to provide accurate representations around the van der Waals minima[74]. We use the Powell conjugate gradient method[100] and Boltzmann weighting to minimize the energy difference between the three points (3.13 – 3.53Å *z* displacement) defining the minimum of the QM and FF van der Waals curves. We constrained the H-H interaction to be consistent with our previous work on polyethylene[24].

In graphite, we assume that the in-place valence interactions are balanced by the out of plane van der Waals interactions. Using this hypothesis, we employed the Biased-Hessian Method[101] to obtain the R_b, k_b, $k_{\theta\theta}$, $k_{r\theta}$ and k_{rr} valence parameters, requiring zero stress in the plane for the experimental a lattice parameter and matching the experimental 868 cm^{-1} 1588 cm^{-1} lattice modes. We did not optimize θ_a, D_b, or k_{rr}. All classical calculations were done using Ewald summation for the R^{-6} dispersion term[102] with a convergence tolerance of 0.001 and a real space cutoff of 4Å.

The final parameters are listed in **Table 2**, and the calculated properties are in **Table 3**. **Figure S1** of the supplementary materials shows a comparison between the QM and QM-FF van der Waals curves.

d) Thermodynamic properties

Our standard approach for calculating the thermodynamic properties is to use a uniform grid (N_g = 20 and 100) along all three directions in reciprocal space (a total of N_g^3 points). The vibrational states for each such point were considered as independent harmonic oscillators in calculating the quantum partition function and thermodynamic properties. In these calculations we used 1/24 of the Brillouin zone with proper weights to obtain the sum over the full Brillouin zone. The three zero frequency modes for the central cell are ignored cell are ignored in all calculations.

At T = 10K we tested these N_g = 100 results by using N_g = 120 (1728000 total points). Here we find that C_v increases by 3×10^{-8} cal/mol K or 0.08%. Based on comparisons to the calculations described below it appears that the C_p from N_g = 100 is accurate down to about 2K (**Table S5** of supplementary materials).

To obtain more accurate thermodynamic properties at lower temperature (T < 10K), we used the frequency distribution functions derived by Komatsu[103]. These functions were derived from the equations for vibrations of thin plates with coupling terms between the plates. This approximation describes only the acoustic modes and is accurate for frequencies below about 130 cm^{-1}. Thus for low temperatures (where contributions from high frequencies are small) it is adequate to use these functions for the thermodynamic properties. The frequency distribution functions are[103]

$$f(v) = \frac{4V}{c'}\left(\frac{1}{v_1^2} + \frac{1}{v_t^2}\right) v sin^{-1}\left(\frac{v}{v_g}\right), for\ v \leq\ v_g$$

and

$$f(v) = \frac{4V}{c'}\left(\frac{1}{v_1^2} + \frac{1}{v_t^2}\right) v \frac{\pi}{2}, for\ v \geq\ v_g \quad (9a)$$

for waves with polarization vectors in the plate and

$$f(v) = \frac{V}{\pi c' \kappa} x \int_0^{\sin^{-1}\left(1/\sqrt{1+\left(\frac{x_0}{x}\right)^2}\right)} \frac{d\varphi}{\sqrt{1 - x^2(1 + \left(\frac{x_0}{x}\right)^2 \sin^2\varphi}},$$

$$for\ v \leq\ v_z' = \frac{\mu}{\pi} \quad (9b)$$

$$= \frac{V}{\pi c' \kappa}\left[1 + \left(\frac{x_0}{x}\right)^2\right]^{-0.5} \int_0^{\pi/2} \frac{d\varphi}{\sqrt{1 - x^2(1 + \left(\frac{x_0}{x}\right)^2 \sin^2\varphi}}, \quad for\ v \geq v_z'$$

for waves with polarization vector out of the plate, where $x = v/v'_z$ and $x_0 = \zeta/4\pi\kappa v'_z$. Our force field leads to the following parameters in these calculations:

c' = c/2 = 3.33605 x 10^{-8} cm
V = volume per mole = 5.3004 cm^3/mol
v_1 = longitudinal wave velocity in the plate
$= \sqrt{C_{11}/\rho}$ = 2.229 x 10^6 cm/s
V_t = transverse wave velocity in the plate
$= \sqrt{C_{66}/\rho}$ = 1.433 x 10^6 cm/s
κ = bending modulus of the plate = 5.616 x 10^{-3} cm^2/s
$\mu = (1/c')\sqrt{C_{33}/\rho}$ = 1.2704 x 10^{13} s^{-1}
$\zeta = C_{44}/\rho$ = 1.1562 x 10^9 cm^2/s^2.

Here, the bending modulus κ, was obtained by fitting the equation $\nu = \kappa k_x^2/2\pi$ to the frequencies calculated for the out-of-the-plane modes at $k_x = 0.1\ k_a$. In this calculation, we used a single isolated infinite layer of graphite.

These thin plate results for T ≤ 10K are shown in the inset of **Figure 7**. For T ≤ 2K we see that the thin plate approximation is superior to using N_g = 100 uniform grid.

e) Thermal Expansion

We investigate the changes in the graphite lattice parameters as a function of temperature(**Figure S3** of supplementary materials) using the LAMMPS[104] MD engine and the QM-FF X6 potential. Here we construct a unit cell of 6 sheets in graphite (800 atom each) in the hexagonal (ABABAB…) orientation, giving an initial cell dimension of 42.8 x 49.4 x 20.0 Å^3. The van der Waals were evaluated with Ewald summation for the R^{-6} dispersion term[102] with a convergence tolerance of 0.0001 and a real space cutoff of 10Å. For the X6 potential, this leads to an Ewald parameter of 0.273155. The MD was performed using integration time step of 1fs and the neighbor list rebuilt every 10 steps.

The cell was first minimized at 0K, followed by 100ps of NPT dynamics at constant pressure (1 atmosphere) and temperature (NPT)[105-108] using an Nose-Hoover thermostat with a temperature-coupling constant of 100.0 fs, and an Berendsen barostat with a 2.0 ps damping constant. The cell coordinates were saved every 1 ps at each discrete temperature, and averaged (table S4 of supplementary materials). We then obtain the in-plane (α_{ab}) and out-of-plane (α_c) linear coefficients of thermal expansion by taking numerical derivatives of the lattice parameters:

$$\alpha_a(T) = \frac{1}{\mathrm{a}_0}\left(\frac{\partial \mathrm{a}}{\partial T}\right)_{\mathrm{P}} \qquad (10a)$$

$$\alpha_c(T) = \frac{1}{c_0}\left(\frac{\partial c}{\partial T}\right)_P \qquad (10b)$$

where a_0 and c_0 are the lattice parameters at 0K.

IV. Results

a) Energetic of two of the coronene stationary points

Six stationary points were obtained by Zhao and Truhlar from M06-2X DFT studies on the coronene dimer[75]. Of these six, we considered two as representative of the graphite structure: the top coronene unit displaced in the *x* (PD-X) and *y* (PD-Y) directions (**Figures 1a/b**). PD-X is topologically similar to the graphite structure, while PD-Y should represent a low energy excited state. Indeed we find that PD-X has the lowest QM energy, 0.51 kcal/mol lower than PD-Y (**Table 1**).

We find that the counterpoise corrected energy of the PD-X structure is 0.52 kcal/mol lower than the PD-Y structure, which compares well with the 0.63 kcal/mol difference obtained at a similar level of theory by Zhao and Truhlar[75]. It has been shown[75] that the M06-2X results on the coronene dimer are similar to those of DFT-D methods (BLYP-D[64] and B97-D[109]), which employs empirical corrections for dispersion interactions.

The magnitude of the counterpoise corrections is larger in our calculations (5.63 and 5.64 kcal/mol for PD-X and PD-Y respectively) that those of Zhao and Truhlar (3.18 and 3.10 kcal/mol respectively), although smaller than 12.64 kcal/mol for the PD-X structure from MP2[50] (using a smaller 6-31G* basis set). Further comparison reveals that the counterpoise corrected energies were similar in both cases; however our non-

counterpoise energies were larger (by 2.48 and 2.92 kcal/mol for PD-X and PD-Y respectively).

There are subtle yet important differences in the two sets of calculations. In our calculations, we performed full geometry optimization at each discrete plane-plane distance, allowing the structures to relax to their energy minima, thereby increasing the binding energy (Zhao and Truhlar performed single point energies at each distance). Additionally, Zhao and Truhlar reported energies obtained using a slightly smaller basis set [6-31G+(d,p)] after the plane-plane distance was obtained using the 6-311G+(2df,2p) basis set. In our study, both the geometry optimization and the final reported energies were done using the 6-311G+(2df,2p) basis set.

b) QM-FF parameters

The atom-centered charges on the coronene monomer and equilibrium PD-X structure were evaluated using the ESP and Mulliken charge methods, and compared to the CM4 results reported by Zhao and Truhlar (**Table S2** of supplementary materials). Here we see little differences in the monomer and dimer charges, across all three methods, indicating that inter-plane charge polarization is small and can be neglected. We selected the ESP charge model for our subsequent calculations, which gave monomer charges of -0.012, 0.172, -0.24 and 0.16 for the inner 6 carbons, the 6 sp2 carbons, the 12 carbons attached to hydrogens and the hydrogens respectively.

The lack of charge transfer in the PD-X dimer is an important, allowing for the use of the monomer charges in evaluating the classical electrostatic component of the binding energy. Fixed charges are one of the central tenets of standard molecular

dynamics calculations. We evaluated the coulombic interactions at each plane-plane distance using the ESP charges of the monomer, removed these interactions from the counterpoise corrected binding curve and report the true van der Waals dispersion curve for the PD-X and PD-Y cases (**Figure S1** of supplementary materials).

We find a van der Waals binding energy of -20.91 kcal/mol for the PD-X structure (**Table S3** of supplementary materials). Interestingly, we obtain the same binding energy for the PD-Y structure, implying that any differences observed in the total binding energy of these structures are due to Pauli repulsion. Additionally, we find that the eclipsed (sandwiched) structure is higher in energy (-15.33 and -10.29 kcal/mol for QM binding energy and counterpoise corrected binding energy respectively) than either the PD-X or PD-Y structure, consistent with the results of Zhao and Truhlar (-12.64 and -9.77 kcal/mol respectively).

The optimized force field parameters obtained from fitting the vdW curves and the two (868 cm^{-1} 1588 cm^{-1}) experimental lattice modes are reported in **Table 2**. As seen in **Table 1**, we are able to describe the PD-X and PD-Y curves accurately with the X6 (-18.56 and -18.08 kcal/mol) and Morse potentials (-18.51 and -18.05 kcal/mol), but not as well with the LJ potential (-18.55 and -18.49 respectively). None of the potentials are able to reproduce the energy of the high energy eclipsed case however (-14.17, -15.90, -16.03 and -16.57 kcal/mol for the QM vdW, X6, Morse and LJ potentials respectively).

Using our optimization procedure, we match the 3 points defining the minima for the PD-X and PD-Y curves to the QM data to an absolute error of < 1%. Further from the minima, significant deviation from the QM vdw curve is observed for each of the

QM-FF potential (**Figure 2**). In particular, differences are observed in the curvature away from equilibrium: the M06-2X curve does not display the $1/R^6$ behavior for the dispersion seen in the QM-FF potentials. We believe this to be indicative of the fact that curvature information was not used when determining the parameters of the M06 class of functionals, and that the energies of these functionals are most valid in the vicinity of the van der Waals minima.

i. Comparison to experiment

The parameters of all three of the QM-FF potentials are similar, with the exception of the curvature parameter, where the X6 and Morse potentials (ζ = 17.64 and 17.29 respectively) are significantly larger than the LJ potential (ζ = 12 by definition). This allows for a better fit to the QM energies (**Figure 2**) and experimental properties (**Table 3**). The *c* lattice constant is directly related to the van der Waals radius of the potential use and we obtain errors of < 0.01% for all three cases, compared to experiment. The C_{33} parameter is related to the curvature of the potential, and here we find errors of 0.12%, 0.4% and 1.9% for the X6, Morse and LJ potential respectively. More relevant to test the parameters would be the cohesive energy, which was determined to be 1.27± 0.12 kcal/mol by a recent experiment[59], where we get errors of -7.1%, +30.7% and -21.3% respectively. We conclude that the X6 potential most closely reproduces the experimental results.

ii. Comparison with other 2-body van der Waals Parameters

In **Table 4** we include the predicted properties for graphite from other sets of van der Waals parameters[28,31-33,35,42-43]. For each case we used the valence parameters

for the van der Waals parameters and readjusted the bond radius R_b to obtain zero stress for the experimental *a* lattice parameter. Then we optimized the unit cell (and internal atoms) and calculated the properties. Girfalco[42] and Ulbricht[43] (the only potentials derived for graphitic systems) did not report any valence interactions, so we assumed harmonic potentials with standard values initially for the valence terms.

The errors in the c lattice constant are: MM2/MM3 (-0.08 Å), Dreiding (+0.11 Å), CHARMM/OPLS (-0.34 Å), AMBER (+0.03 Å), Girifalco (-0.001 Å), Ulbricht (+0.03 Å) and GraFF (+0.004). All parameters except for CHARMM/OPLS and Dreiding could be considered adequate.

Next, the errors in the C_{33} elastic constant (out of 40.7): MM2/MM3 (-41%), Dreiding (+48%), CHARMM/OPLS (+4%), AMBER (+23%), Girifalco (+21%), Ulbricht(-14%) and GraFF (-0.5%). Here, the GraFF forcefield[20] performs most admirably. The next best is the CHARMM/OPLS parameters. The Dreiding and MM2/MM3 parameters show large errors, and the next three are marginally acceptable.

Using the experimental cohesive energy, the errors are: MM2/MM3 (-15%), Dreiding (+62%), CHARMM/OPLS (+51%), AMBER (+71%), Girifalco (+70%), Ulbricht (+17%), and GraFF (+37%). The MM2/MM3 and Ulbricht parameters are in good agreement with the experimental value, the GraFF parameters are acceptable. Large errors are seen in the Dreiding, CHARMM/OPLS, AMBER and Girifalco parameters.

Combining all three criteria, we summarize that the most accurate van der Waals parameters for carbon is the GraFF force fields, with the Ulbricht parameters showing acceptable results and the CHARMM/OPLS and MM2/MM3 parameters barely

acceptable. The GraFF forcefield is our purely empirical effort, which has been shown to give the correct cohesive energy of stacked C_{60} molecules, so the acceptable errors in predicting the properties of graphite are encouraging. We consider none of the other parameters to be adequate for accurate calculations on graphite and find that the new QM-FF potentials (**Table 2**) improves on the GraFF forcefield and therefore should be useful for describing graphene, nanotubes, fullerenes and other graphitic materials.

iii. Graphite PES

We map the entire potential energy surface (PES) of graphite by sliding a flat sheet of graphene over another in 0.1Å increments in the x and y directions (starting from the eclipsed structure). At each displacement, we obtained the singlepoint vdW energy with the LAMMPS[104] MD engine and the X6 potential. The unit cell comprised of 1600 atoms (800 per sheet) infinitely parallel in the xy plane, with a cell dimension of 42.8 x 49.4 x 10.0 $Å^3$. The vertical *z* displacement was optimized for each displacement and resulting surface shown in **Figure 3a**. The high energy eclipsed structure (+5.6 kcal/mol) is evident at position (0,0) and $\Delta x = \pm 2.125Å$, $\Delta y = \pm 1.25Å$. An example of the minimum energy pathway for sliding in the X and Y directions is a tortured path along the edges of the hexagon as shown by the yellow dashed lines in **Figure 3b**. The barrier for sliding is 0.5 and 1.2 kcal/mol for the X and Y directions respectively (along the edges of the hexagon as shown by the minimum energy path), compared to 5.1 kcal/mol for the (1,0,0) and (0,1,0) directions over the eclipsed structure.

c) Mechanical properties of graphite

The full set of elastic constants (stiffness C_{ij} and compliance S_{ij}) are given in **Table S4** of supplementary materials (note that only the upper half of the matrices are given).

i. C_{44} Shear Stiffness

In mechanical studies of graphite[78] the observed shear elastic constant C_{44} ranged from 0.18 to 0.33 GPa, whereas the neutron-irradiated samples[79] led to 4.2 to 4.6 GPa. It is widely believed that neutron irradiation pins dislocations, leading to parameters appropriate for the intrinsic system; hence the name "dislocation-free graphite" for these materials[81-82]. Boron implantation also leads to the larger values of C_{44}[81-82].

There is evidence against this belief. Thus, in studying C_{44} down to 2 K, Gauster and Fritz[86] observed a nearly constant low value, whereas if the low value were due to dislocations, one would have expected an increase for temperatures too low to activate the dislocations. Similarly, Ayasse et al.[82] measured C_{44} down to 0.1 K and found no evidence of the hardening expected if the low C_{44} were due to dislocations. Our calculations at 0K lead to value of C_{44} (0.45 to 0.74), providing evidence for the smaller value of C_{44} from neutron-free samples is intrinsic, suggesting that the neutron irradiation (and boron implantation) causes a dramatic increase in shear resistance.

ii. C_{13} Stiffness

The calculated values for C_{13} are small and track the values for C_{44} (in all cases C_{13} is within 2% of C_{44}). Experimental values are much larger but the measurements are indirect. Thus in the mechanical studies of Blakslee et al[78], it was not possible to

measure C_{13} directly nor was it possible to measure the related compliance S_{13} in ultrasonic experiments. Instead, S_{13} was obtained by measuring the Poisson ratio in static experiments and C_{13} was calculated using

$$C_{13} = -S_{13}/[S_{33}(S_{11} + S_{12}) - 2S_{13}^2]$$

leading to C_{13} = 15 ± 5 GPa at room temperature. In the temperature-dependence studies of Gauster and Fritz[86], C_{13} could be obtained only from crosscut samples (45° from the c axis), leading to a velocity of sound v given by

$$pv^2 = \frac{1}{4}\left\{C_{11} + C_{33} + 2C_{44} - \sqrt{(C_{11} - C_{33})^2 + 4(C_{13} + C_{44})^2}\right\}$$

(where *p* is the density). Since C_{11} is much larger than C_{13}, this leads to considerable uncertainty in C_{13} (the uncertainty is comparable to the magnitude). Interestingly, the values for C_{13} deduced from these experiments are independent of temperature from 0 to 250 K and then change rapidly for higher temperature.

Our calculated (small) values of C_{13} are consistent with flat graphite planes. For higher temperature the mean fluctuations in crinkling of the sheets should increase, leading to an increased c lattice constant (as observed). Applying tension parallel to the sheets might decrease these fluctuations, thereby decreasing the c axis. Thus it is plausible that C_{13} might be small and constant for low temperature and large for high temperature (as observed[86]).

Consistent with the small calculated value of C_{13}, we obtain small values for the Poisson ratios coupling strain in the *a* and *c* directions. Thus, for strain in the *x* direction we find μ_{21} = 0.171, 0.165, 0.174 for the X6, LJ and Morse potentials respectively (observed values at room temperature = 0.16 ± 0.06). For the *y* direction μ_{31}, we obtain the same values (observed values at room temperature = 0.34 ± 0.08). For stress in the

z direction we find almost no effect in the x and y directions, μ_{13} = 0.0004, 0.0004, 0.0006 respectively (the estimate from observations at room temperature is = 0.012 ± 0.003). Experiments at low temperature on these Poisson ratios would be most valuable in resolving these issues.

iii. Lattice Modes

The mean value of the out-of-plane crinkling modes (868 cm^{-1}) was adjusted to fit experiment. This leads to a rotational barrier of 20.00, 20.50 and 18.74 kcal/mol for the X6, LJ and Morse potentials respectively, about 1/3 the value for ethylene (65 kcal), as might be expected (since the resonance structures have one π bond for each three C-C bonds). There are two modes at 868 cm^{-1} (one infrared-allowed), and we calculate the (Davydov) splitting to be 2.2, 1.7 and 2.9 cm^{-1} respectively. Similarly, the splitting of the in-plane modes (1588 cm^{-1}) is predicted to be 0.2 cm^{-1}.

The 10 mode (B_{2g}) is predicted at 137.5, 139.4 and 139.5 cm^{-1} respectively and is observed (room temperature) by neutron scattering at 126 cm^{-1}. Expansion of the *c* lattice parameter from the value at 0 K to the value at 300 K (see Section V.A) leads to a prediction value of v = 132 cm^{-1} at 300K, explaining most of the discrepancy. The frequency of this mode is directly related to C_{33} which we calculate to 0.2%.

The TO mode (E_{2g}) is directly related to the C_{44} elastic constant. Thus neutron-irradiated samples that yield C_{44} = 4.4 lead to a much higher frequency (45 cm^{-1}) than the calculations (which yield 14.2 to 18.2 cm^{-1} and C_{44} of 0.45 to 0.74 GPa). As discussed above, we believe that the discrepancy is due to neutron damage. Experimental studies

of the TO mode for low temperature and low irradiation would be most valuable to help resolve these uncertainties.

Based on the above comparisons, we conclude that the best description of the van der Waals interactions is the X6 potential with ζ = 17.6, and we have used this to calculate several properties for graphite.

d) Phonon Dispersion Curve

In **Figure S2** we show the calculated phonon dispersion curves for the lattice modes of graphite at 0 K. The vibrational frequencies are plotted in THz and cm^{-1} (to convert THz to cm^{-1}, multiply by 33.35641). The left half is for waves along the *c*-axis, while the right half is for waves along the *a*-axis. The acoustical modes are denoted TA, ZA and LA (transverse in-plane, transverse out-of-plane and longitudinal), while the sheet-sheet optical modes are denoted TO, ZO and LO.

The velocity of sound from these calculations is

$$v_T = \sqrt{C_{44}/\rho} = 0.44\ km/sec$$
$$v_L = \sqrt{C_{33}/\rho} = 4.24\ km/sec$$

for modes along the c axis and

$$v_{T_1} = \sqrt{C_{44}/\rho} = 0.44\ km/sec$$
$$v_{T_2} = \sqrt{C_{66}/\rho} = \sqrt{(C_{11} - C_{12})/2\rho} = 13.90\ km/sec$$
$$v_L = \sqrt{C_{11}/\rho} = 21.59\ km/sec$$

for modes along the *a* axis. In **Figure 4** we show all phonon modes. For vibrations within the sheets, there is little dispersion for waves in the *c* direction.

There are direct experimental data on the lattice modes from neutron scattering[79-80]; however, these data are for room temperature. To better compare our results with these data, we recalculated the lattice modes using the lattice parameters

for room temperature. To do this we modified the valence and van der Waals parameters to fit the room temperature lattice constants (*a* and *c*), elastic constants (C_{11}, C_{12}, C_{33}), and sheet vibrational modes (868 and 1588 cm^{-1}) as in Section III.C for 0 K. We note that technically this is comparison is still not entirely consistent, since the calculated phonon frequencies do not take into account anharmonicity due to zero-point motions. The temperature effect on the phonon have been known to soften the frequencies by approximately 1% for the hexagonal MgB_2 crystal[110].

The modified force field is in the last column of **Table 2** and the various properties are summarized in **Table 3**. The room temperature lattice modes are plotted in **Figure 5** along with the experimental data. For waves in the *a* direction, there is good agreement with the ZO and ZA modes (traverse modes with amplitudes perpendicular to the sheets). The data for longitudinal modes lie slightly higher than the calculated LO and LA modes. This is expected since the experimental dispersion curve leads to an elastic constant of C_{11} = 1440 GPa, whereas the calculated dispersion leads to C_{11} = 1061 GPa, in agreement with the experimental value from mechanical studies (C_{11} = 1060 ± 20 GPa).

For the TA mode at the *M* k-point, we find a convergence towards 800 cm^{-1}, in good agreement with High-resolution electron energy loss spectroscopy (HREELS) on graphite[111], but larger than that 626 cm^{-1} and 634 cm^{-1} calculated from LDA and GGA DFT calculations (**Table 5**). The TA and TO modes are apparently not observed in the neutron scattering experiments. Additionally, consistent with a recent inelastic x-ray

scattering study by Maultzsch and coworkers[112], we find crossing of the LO and TO branches between the Γ-M and Γ-K directions.

For waves along the *c* direction, the predicted LA and LO modes are in good agreement with experiment. As discussed earlier, the TA modes for neutron-irradiated samples lead to a C_{44} that is 15 times larger than the C_{44} from mechanical experiments on irradiated samples. Our calculations support the low values for C_{44} and lead to TA and TO modes much lower in energy than for neutron experiments.

e) Thermodynamic Properties of graphite

The predicted values of various properties using each of these three forcefields are listed in **Table 3**. for T > 20K.

i. Cohesive and Surface Energy

The experimentally determined cohesive energy graphite of 1.27±0.1 kcal/mol[59] is lower than that of similar hydrocarbons: the experimental cohesive energy (at 0 K) of polyethylene[113] is 1.838 ± 0.032 kcal/mol, the experimental heat of sublimation (per carbon) is 2.1 kcal/mol[114] for benzene and 2.1 kcal/mol[114] for n-hexane. The calculated cohesive energy is 1.18, 1.66 and 1.00 kcal/mol for X6, LJ and Morse potentials respectively, representing errors of -7%, +31% and -22%.

Given the cohesive energy of 1.27 kcal/mol/C, the smallest energy to create a new surface would be $0.625/(\sqrt{3}a^2/4)$ = 0.24 (kcal/mol)/Å^2 = 1.00 (kJ/mol)/Å^2 = 0.172 J /m^2= 172 erg/cm^2. This should be a lower bound on the surface energy. Using the parameters for room temperature leads to an estimated surface energy of 200 erg/cm^2. An upper bound on the energy for creating a surface perpendicular to the sheets is

$\left(\frac{1}{3}113\right)/\left(\frac{1}{2}ac\right)$ = 4.5 kcal/mol/Å^2 = 19 (kJ/mol)/ Å^2 = 3.2 J/ m^2 = 3200 erg/cm^2 [assuming each broken bond at the surface costs the average bond energy of $\frac{2}{3}$ of the total cohesive energy].

ii. Free energy and Specific Heat Capacity

The entropy (S), energy (V), and free energy (F) are plotted in **Figure 6** and tabulated in **Table S5** of supplementary materials. The most relevant quantity for comparison to experiment is the specific heat capacity, where we find that the results from the 20x20x20 grid (dotted lines) differs from the 100x100x100 grid (solid lines above T = 10K, dashed lines below) for T < 40K (see the inset of **Figure 7**). This sensitivity to grid size is primarily due to the low frequency modes in the z direction (see **Figure 4** and **Figure S2** of supplementary materials). In particular, neglecting the contribution of the three acoustical modes for the central point leads to a low C_p. However the calculated C_p for N_g = 20 is too *high* for T < 10K. This occurs because only modes with kx = ky = 0 contribute to C_p and the weights of these modes are too large due to the small number of points in kx-ky plane.

The thin plate results for T ≤ 10K are shown in the inset of **Figure 7** by the solid lines. We see that all three experiments and both theoretical cases agree above 20K. However below 10K the experimental results for pile graphite and canadian natural graphite lead to specific heats significantly below that of graphitized lampblack. Also it is clear that the calculated optimum parameters for graphite C_{44} = 0.25 GPa) lead to a C_p in excellent agreement with experiment for graphitized lampblack down to 1K. On the

other hand, the calculated C_p values are not in agreement with experiment for pile graphite and Canadian natural graphite.

iii. Lattice parameters and Thermal Expansion

As seen in **Figure 8a**, there is a decrease in the lattice constant in the ab direction from 0 – 350K, to a minimum value of 2.4680Å. This is consistent to the experimental results of Kellett and Richards[115] who reported a minimum of 2.4607 Å between 350 and 400K, and the experimental results of Nelson and Riley[93] and the theoretical results of Riley[116]. From 400 – 2500K, we observe an linear increase in the lattice parameter.

Below room temperature, we find good agreement with the overall shape of experimental coefficients of thermal expansion[77,117] (**Figure 8b**), although the X6 potential predicts a minima at 150K, whereas the experimental results predict a minima between 200 and 250 K. At the minima, we predict a value of -1.239×10^{-6} deg^{-1} K, in excellent agreement with the experimental value of 1.236×10^{-6} deg^{-1} K. This result is not surprising considering the agreement in the specific heat at this temperature observed previously.

From 300K – 100K, the X6 potential overestimates the thermal expansion coefficient by 40%, although convergence is observed at higher temperatures. The out of plane lattice parameters increases almost linearly from 0 – 2500K (**Figure S3** of supplementary materials), in excellent agreement with the experimental results[115].

f) Rhombohedral Graphite

The planes in graphite stack in the sequence ABABAB ... , leading to hexagonal symmetry. A second structure has been observed with the stacking ABCABC... , leading

to rhombohedral symmetry[76]. From experiment, the hexagonal form is clearly the more stable, but an estimate of the difference in energy is not available. Using the force fields, we calculated these forms to have energies within 0.01 kcal/mol. [These calculations use an accuracy parameter of 0.0001 kcal/mol in the convergence acceleration[102], leading to η = 1.55Å, R_{cut} = 5.55Å, and H_{cut} = 5.72Å^{-1} for the hexagonal form with ζ = 17.6]. The phonon modes for rhombohedral graphite are shown in **Figure 9**. We find that the dispersion curve is similar to that of hexagonal graphite for frequencies over 400cm^{-1}. The major discrepancies occur in the *A* – *Γ* region, with the rhombohedral graphite structure showing and additional LO band in this region. The thermodynamic properties of rhombohedral graphite are calculated and can be compared with hexagonal graphite in **Figure 7**; the differences are too small to see however. Since high quality graphite is hexagonal with but few stacking faults, we believe that the calculated energy different may be reasonable. (Bill: I'm not sure about this last sentence, as I'm not sure if the magnitude of the differences is significant enough).

V. Discussion

The experiments and calculations on graphite leave some important discrepancies unresolved. The major issue is the value of C_{44}. Ultrasonic measurements on compression annealed graphite lead to C_{44} = 0.28 ± 0.08. Our theoretical force field yields C_{44} 0.45 to 0.74, in good agreement with the above results. This force field predicts a specific heat, C_p, in agreement with experiments on graphitized lampblack. On the other hand, neutron scattering leads to C_{44} ~ 4.0 GPa.

Our speculation is that neutron-induced defects might disrupt the resonance system in graphite, leading to lower symmetry regions that might cause the planes to be "stickier". The result is a "pinned" resonance structure where bond charges in adjacent planes might couple the sheets, leading to larger C_{44} (and C_{13}). In addition, such structures could have single bond torsions smaller than in normal resonant graphite, allowing crinkling that would couple the layers. Alternatively, as the temperature is increased, the instantaneous structure should have increasing crinkling of the planes. This could increase C_{44} markedly while affecting C_{33} only a small amount. However, the observed small change in C_{44} with temperature[86] indicates that this is not an important effect.

We should emphasize here that the assumption (1) that the graphite sheets are flat and (2) that the van der Waals interactions are described in terms of X6 or LJ type two body functions combined with (3) the requirement of fitting the experimental values of c = 6.6721Å and C_{33} = 40.7 GPa leads directly to values of C_{44} = 0.2 to 0.3 and necessarily leads to (a) $C_{13} \sim C_{44}$, (b) the high values of Cp near 1K and (c) a small energy separation between hexagonal and rhombohedral graphite.

[Al-Jishi and Dresselhaus[118] have reported a set of parameters that lead to C_{44} = 4.2 and C_{33} = 36.9; however the model used considers only interactions of atoms on the same or adjacent layers (and only up to 4th neighbors) and involves individually adjusted radial and tangential pair-wise force constants. These pair-wise forces constants are not derivable from a smooth two-body potential.]

If the high value of C_{44} from neutron experiments is the correct value for perfect graphite at 0K, then we believe that either (or both) of the above assumptions (flat sheets or two body long range attractions) must be false. That is, either the graphite planes are ruffled (as suggested by Pauling, and confirmed for recently for graphene[119]) or the van der Waals interactions cannot described in terms of smooth pair-wise additive terms.

The latter hypothesis is possible. Graphite is a semimetal, i.e. the valence band maximum (highest occupied molecular orbital) is degenerate with the conduction band minimum (lowest unoccupied molecular orbital). In this circumstance there should be extra large contributions to the dispersion interactions due to the states near the band edges (small energy denominators). This might well have a major effect on C_{44} (and $C_{13)}$.These effects would not be captured by the M06-2X functional used in this study, and would lead to larger vdW energies than reported here and would necessarily not be accurately described with smooth two-body terms. Such investigations could be carried out with the new generation of doubly hybrid meta DFT functionals (such as XYG3[70]).

Another possible limitation of the two-body potentials presented here is the inability to reproduce the energy of eclipsed structure (ΔE = 10.2 and 5.5 kcal/mol for the QM and QM-FF X6 respectively). This indicates that the inner walls of these potentials are too soft (not repulsive enough). This may be remedied by a potential that describes the curvature of the inner and outer walls with two separate terms, such as a stretched exponential.

VI. Concluding remarks

We find that nonbonded parameters based upon high-quality (M06-2X) DFT, combined with two in-plane lattice modes, leads to forcefields that describe well the properties of graphite, which are sensitive to London dispersion interactions. The specific heat, thermal expansion and lattice modes in good agreement with experiment, however. This, we believe that these carbon potentials will be useful for examining surface properties and adsorbates on graphite. In addition, they should be useful in considering the intercalation of various molecules in graphite and in examining adhesion to polymers and other molecules.

VII. Acknowledgements.

This research was partially supported by MARCO-FENA. TAP thanks the National Science Foundation and US Department of Energy (CSGF) for graduate fellowships.

VIII. References and Notes

(1) Qian, D.; Wagner, G. J.; Liu, W. K.; Yu, M.-F.; Ruoff, R. S. *Applied Mechanics Reviews* **2002**, *55*, 495.

(2) Dunlap, B. I.; Boettger, J. C. *Journal of Physics B-Atomic Molecular and Optical Physics* **1996**, *29*, 4907.

(3) Kahaly, M. U.; Waghmare, U. V. *J. Nanosci. Nanotechnol.* **2007**, *7*, 1787.

(4) Mounet, N.; Marzari, N. *Physical Review B* **2005**, *71*.

(5) Wirtz, L.; Rubio, A. *Solid State Communications* **2004**, *131*, 141.

(6) Nagamiya, T.; Komatsu, K. *J. Chem. Phys.* **1954**, *22*, 1457.

(7) Komatsu, K. *Journal of the Physical Society of Japan* **1955**, *10*, 346.

(8) Komatsu, K.; Nagamiya, T. *Journal of the Physical Society of Japan* **1951**, *6*, 438.

(9) Liu, A.; Stuart, S. J. *J. Comput. Chem.* **2008**, *29*, 601.

(10) Los, J. H.; Ghiringhelli, L. M.; Meijer, E. J.; Fasolino, A. *Physical Review B* **2005**, *72*.

(11) Tewary, V. K.; Yang, B. *Physical Review B* **2009**, *79*.

(12) Yanagisawa, H.; Tanaka, T.; Ishida, Y.; Matsue, M.; Rokuta, E.; Otani, S.; Oshima, C. *Surface and Interface Analysis* **2005**, *37*, 133.

(13) Zimmermann, J.; Pavone, P.; Cuniberti, G. *Physical Review B* **2008**, *78*.

(14) Brenner, D. W. *Physical Review B* **1990**, *42*, 9458.

(15) Tersoff, J. *Phys. Rev. Lett.* **1988**, *61*, 2879.

(16) Stillinger, F. H.; Weber, T. A. *Physical Review B* **1985**, *31*, 5262.

(17) Aizawa, T.; Souda, R.; Otani, S.; Ishizawa, Y.; Oshima, C. *Physical Review B* **1990**, *42*, 11469.

(18) Perebeinos, V.; Tersoff, J. *Physical Review B (Condensed Matter and Materials Physics)* **2009**, *79*, 241409.

(19) Collignon, B.; Hoang, P. N. M.; Picaud, S.; Liotard, D.; Rayez, M. T.; Rayez, J. C. *Journal of Molecular Structure: THEOCHEM* **2006**, *772*, 1.

(20) Guo, Y. J.; Karasawa, N.; Goddard, W. A. *Nature* **1991**, *351*, 464.

(21) Hummer, G.; Rasaiah, J. C.; Noworyta, J. P. *Nature* **2001**, *414*, 188.

(22) Chen, R. J.; Bangsaruntip, S.; Drouvalakis, K. A.; Kam, N. W. S.; Shim, M.; Li, Y. M.; Kim, W.; Utz, P. J.; Dai, H. J. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 4984.

(23) Meyer, E. A.; Castellano, R. K.; Diederich, F. *Angew. Chem.-Int. Edit.* **2003**, *42*, 1210.

(24) Karasawa, N.; Dasgupta, S.; Goddard, W. A. *J. Phys. Chem.* **1991**, *95*, 2260.

(25) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E.; Wang, J.; Ross, W. S.; Simmerling, C.; Darden, T.; Merz, K. M.; Stanton, R. V.; al., e.; 7 ed.; University of California, San Francisco.: 1999.

(26) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817.

(27) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157.

(28) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *Journal of the American Chemical Society* **1995**, *117*, 5179.

(29) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.

(30) Mackerell, A. D.; Wiorkiewiczkuczera, J.; Karplus, M. *Journal of the American Chemical Society* **1995**, *117*, 11946.

(31) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.

(32) Jorgensen, W. L.; Severance, D. L. *Journal of the American Chemical Society* **1990**, *112*, 4768.

(33) Allinger, N. L. *Journal of the American Chemical Society* **1977**, *99*, 8127.

(34) Allinger, N. L.; Lii, J. H. *J. Comput. Chem.* **1987**, *8*, 1146.

(35) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. *J. Phys. Chem.* **1990**, *94*, 8897.

(36) Goddard, W. A. *Handbook of nanoscience, engineering, and technology*; CRC Press: Boca Raton, 2003.

(37) Wu, J.; Pisula, W.; Mullen, K. *Chemical Reviews* **2007**, *107*, 718.

(38) Stankovich, S.; Dikin, D. A.; Dommett, G. H. B.; Kohlhaas, K. M.; Zimney, E. J.; Stach, E. A.; Piner, R. D.; Nguyen, S. T.; Ruoff, R. S. *Nature* **2006**, *442*, 282.

(39) Venema, L. *Nature* **2007**, *446*, 36.

(40) Son, Y.-W.; Cohen, M. L.; Louie, S. G. *Nature* **2006**, *444*, 347.

(41) Barone, V.; Hod, O.; Scuseria, G. E. *Nano Lett.* **2006**, *6*, 2748.

(42) Girifalco, L. A.; Hodak, M.; Lee, R. S. *Physical Review B* **2000**, *62*, 13104.

(43) Ulbricht, H.; Moos, G.; Hertel, T. *Phys. Rev. Lett.* **2003**, *90*, 095501.

(44) Fetzer, J. C. *Polycyclic Aromatic Compounds* **2007**, *27*, 143.

(45) Marzec, A. *Carbon* **2000**, *38*, 1863.

(46) Falconer, J. L.; Madix, R. J. *Journal of Catalysis* **1977**, *48*, 262.

(47) Fichthorn, K. A.; Miron, R. A. *Phys. Rev. Lett.* **2002**, *89*.

(48) Gameson, I.; Rayment, T. *Chem. Phys. Lett.* **1986**, *123*, 150.

(49) Mercedes Rubio; Enrique Ortí; José Sánchez-Marín *Int. J. Quantum Chem.* **1996**, *57*, 567.

(50) Ruuska, H.; Pakkanen, T. A. *The Journal of Physical Chemistry B* **2001**, *105*, 9541.

(51) Rapacioli, M.; Calvo, F.; Spiegelman, F.; Joblin, C.; Wales, D. J. *The Journal of Physical Chemistry A* **2005**, *109*, 2487.

(52) Conney, A. H. *Cancer Research* **1982**, *42*, 4875.

(53) Schmidt, M. W. I.; Noack, A. G. *Global Biogeochemical Cycles* **2000**, *14*, 777.

(54) Pignatello, J. J.; Xing, B. S. *Environmental Science & Technology* **1996**, *30*, 1.

(55) Mastral, A. M.; Callen, M. S. *Environmental Science & Technology* **2000**, *34*, 3051.

(56) Langhoff, S. R. *J. Phys. Chem.* **1996**, *100*, 2819.

(57) Allamandola, L. J.; Tielens, A.; Barker, J. R. *Astrophysical Journal Supplement Series* **1989**, *71*, 733.

(58) Sagan, C. E.; Lippinco.Er; Dayhoff, M. O.; Eck, R. V. *Nature* **1967**, *213*, 373.

(59) Zacharia, R.; Ulbricht, H.; Hertel, T. *Physical Review B* **2004**, *69*, 7.

(60) Lee, N. K.; Kim, S. K. *The Journal of Chemical Physics* **2005**, *122*, 031102.

(61) Møller, C.; Plesset, M. S. *Physical Review* **1934**, *46*, 618.

(62) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M. *The Journal of Chemical Physics* **2004**, *120*, 647.

(63) Zhao, Y.; Truhlar, D. G. *Journal of Chemical Theory and Computation* **2006**, *3*, 289.

(64) Stefan Grimme *J. Comput. Chem.* **2004**, *25*, 1463.

(65) Hohenberg, P.; Kohn, W. *Physical Review B* **1964**, *136*, B864.

(66) Kohn, W.; Sham, L. J. *Physical Review* **1965**, *140*, 1133.

(67) Rydberg, H.; Dion, M.; Jacobson, N.; Schroder, E.; Hyldgaard, P.; Simak, S. I.; Langreth, D. C.; Lundqvist, B. I. *Phys. Rev. Lett.* **2003**, *91*.

(68) Sato, T.; Tsuneda, T.; Hirao, K. *The Journal of Chemical Physics* **2007**, *126*, 234114.

(69) Kohn, W.; Meir, Y.; Makarov, D. E. *Phys. Rev. Lett.* **1998**, *80*, 4153.

(70) Zhang, Y.; Xu, X.; Goddard, W. A. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 4963.

(71) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *Journal of Chemical Theory and Computation* **2006**, *2*, 364.

(72) Zhao, Y.; Truhlar, D. G. *Journal of Chemical Theory and Computation* **2005**, *1*, 415.

(73) Zhao, Y.; Truhlar, D. G. *J. Chem. Phys.* **2006**, *125*.

(74) Zhao, Y.; Truhlar, D. *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)* **2008**, *120*, 215.

(75) Zhao, Y.; Truhlar, D. G. *Journal of Physical Chemistry C* **2008**, *112*, 4061.

(76) Donohue, J. *The Structures of the Elements*; Robert E. Krieger Publishing Co.: Malabar, Florida, 1974.

(77) Bailey, A. C.; Yates, B. *Journal of Applied Physics* **1970**, *41*, 5088.

(78) Blakslee, O. L. *Journal of Applied Physics* **1970**, *41*, 3373.

(79) Dolling, G.; Brockhouse, B. N. *Physical Review* **1962**, *128*, 1120.

(80) Nicklow, R.; Smith, H. G.; Wakabaya.N *Physical Review B* **1972**, *5*, 4951.

(81) Baker, C.; Kelly, A. *Philosophical Magazine* **1964**, *9*, 927.

(82) Ayasse, J. B.; Ayache, C.; Jager, B.; Bonjour, E.; Spain, I. L. *Solid State Communications* **1979**, *29*, 659.

(83) Komatsu, K. *Journal of Physics and Chemistry of Solids* **1958**, *6*, 380.

(84) Komatsu, K. *Journal of Physics and Chemistry of Solids* **1964**, *25*, 707.

(85) Bowman, J. C.; Krumhansl, J. A. *Journal of Physics and Chemistry of Solids* **1958**, *6*, 367.

(86) Gauster, W. B.; Fritz, I. J. *Journal of Applied Physics* **1974**, *45*, 3309.

(87) Nemanich, R. J.; Lucovsky, G.; Solin, S. A. *Solid State Communications* **1977**, *23*, 117.

(88) Nihira, T.; Iwata, T. *Physical Review B* **2003**, *68*.

(89) Pierson, H. O.; Knovel; Noyes Publications: Park Ridge, N.J., 1993.

(90) Desorbo, W.; Tyler, W. W. *J. Chem. Phys.* **1953**, *21*, 1660.

(91) Prophet, H.; Stull, D. R.; United States. National Bureau of, S. *JANAF thermochemical tables*; USGPO: Washington,D.C., 1971.

(92) Hultgren, R. R.; American Society for, M. *Selected values of the thermodynamic properties of the elements*; American Society for Metals: Metals Park, Ohio, 1973.

(93) Nelson, J. B.; Riley, D. P. *Proceedings of the Physical Society* **1945**, *57*, 477.

(94) Schrödinger: New York, NY, 2007.

(95) Hehre, W. J. R., L.; Schleyer, P. v. R.; Pople, J. A.; *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986.

(96) Boys, S. F.; Bernardi, F. *Molecular Physics: An International Journal at the Interface Between Chemistry and Physics* **1970**, *19*, 553

(97) Schwenke, D. W.; Truhlar, D. G. *The Journal of Chemical Physics* **1985**, *82*, 2418.

(98) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833.

(99) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **2005**, *113*, 133.

(100) Powell, M. J. D. *Mathematical Programming* **1977**, *12*, 241.

(101) Dasgupta, S.; Goddard, W. A. *J. Chem. Phys.* **1989**, *90*, 7207.

(102) Karasawa, N.; Goddard, W. A. *J. Phys. Chem.* **1989**, *93*, 7320.

(103) Komatsu, K. *Journal of the Physical Society of Japan* **1955**, *10*, 346.

(104) Plimpton, S. *J. Comput. Phys.* **1995**, *117*, 1.

(105) Hoover, W. G. *Physical Review A* **1986**, *34*, 2499.

(106) Hoover, W. G. *Physical Review A* **1985**, *31*, 1695.

(107) Nose, S. *J. Chem. Phys.* **1984**, *81*, 511.

(108) Nose, S. *Molecular Physics* **1984**, *52*, 255.

(109) Grimme, S.; Muck-Lichtenfeld, C.; Antony, J. *The Journal of Physical Chemistry C* **2007**, *111*, 11199.

(110) Lazzeri, M.; Calandra, M.; Mauri, F. *Physical Review B* **2003**, *68*.

(111) Oshima, C.; Aizawa, T.; Souda, R.; Ishizawa, Y.; Sumiyoshi, Y. *Solid State Communications* **1988**, *65*, 1601.

(112) Maultzsch, J.; Reich, S.; Thomsen, C.; Requardt, H.; Ordejon, P. *Phys. Rev. Lett.* **2004**, *92*.

(113) Billmeyer, F. W. *Journal of Applied Physics* **1957**, *28*, 1114.

(114) Williams, D. E.; Cox, S. R. *Acta Crystallographica Section B-Structural Science* **1984**, *40*, 404.

(115) Kellett, E. A.; Richards, B. P. *Journal of Nuclear Materials* **1964**, *12*, 184.

(116) Riley, D. P. *Proceedings of the Physical Society of London* **1945**, *57*, 486.

(117) Morgan, W. C. *Carbon* **1972**, *10*, 73.

(118) Aljishi, R.; Dresselhaus, G. *Physical Review B* **1982**, *26*, 4514.

(119) Fasolino, A.; Los, J. H.; Katsnelson, M. I. *Nat Mater* **2007**, *6*, 858.

(120) Tuinstra, F.; Koenig, J. L. *J. Chem. Phys.* **1970**, *53*, 1126.

(121) Siebentritt, S.; Pues, R.; Rieder, K. H.; Shikin, A. M. *Physical Review B* **1997**, *55*, 7927.

(122) Mohr, M.; Maultzsch, J.; Dobardzic, E.; Reich, S.; Milosevic, I.; Damnjanovic, M.; Bosak, A.; Krisch, M.; Thomsen, C. *Physical Review B (Condensed Matter and Materials Physics)* **2007**, *76*, 035439.

IX. Tables

Table 1. Comparison of interaction energies (kcal/mol) for PD-X, PD-Y and Eclipsed Coronene dimer configurations. The z displacement for the PD-X and PD-Y cases is 3.33Å, and 3.56Å for the eclipsed structure.

	QM-FF[a]			M06-2X[b]		M06-2X[c]		MP2[d]	
	X6	**Morse**	**LJ**	**Cp**	**no Cp**	**Cp**	**no Cp**	**Cp**	**no Cp**
PD-X	18.56	18.51	18.55	18.58	24.21	18.55	21.73	18.68	31.32
PD-Y	18.08	18.05	18.49	18.06	23.70	17.88	20.98		
Eclipsed	15.90	16.03	16.57	10.29	15.33	9.77	12.64		

[a] fitting was done to the counterpoise corrected, vdW energies of the 5 points defining the energy minimum of the PD-X and PD-Y curves. The coulombic interactions are included for comparison sake

[b] current work. 6-311G+(2df,2p) basis set. Geometries optimized at each plane-plane distance

[c] Reference[75]. 6-311G+(2df,2p) for geometry optimization/ 6-31G+(d,p) for singlepoint calculations at each plane-plane distance

[d] Reference[50]. 6-31G* basis set

Table 2. Optimized Force Field Parameters for Graphite at 0K

Parameters	**X6 [eq (1)]**	**LJ [eq (5)]**	**Morse [eq (6)][a]**	**X6 (300K)**
van der Waals				
R_v (Å)	3.7248	3.7997	3.6530	3.7566
D_v (kcal/mol)	0.0635	0.0704	0.0609	0.0566
ζ	17.64		17.29	16.88
Bond Stretch [eq (6)]				
R_b (Å)	1.4046	1.4089	1.4174	1.4077
K_b [(kcal/mol)/Å^2]	700.36	799.946	677.33	720.48
D_b (kcal/mol)	98.38	150.617	98.94	116.40
Angle Bend [eq (7)]				
Θ_a (deg)	120.69	120.582	121.74	120.21
$k_{\Theta\Theta}$ [(kcal/mol)/rad^2]	182.74	160.300	175.09	167.06
D [(kcal/mol)/ Å]	-59.47	-72.528	-66.13	-61.51
k_{rr} [(kcal/mol)/ Å^2]	57.69	50.698	58.81	51.31
Torsion [eq (8)]				
V_t [kcal/mol]	20.00	20.50	18.74	20.00

[a] α_v was used instead of ζ in eq 6

Table 3. Calculated Properties for Graphite

	Experiment (0K)[a]	Fitted?	X6 [eq (1)]	LJ [eq (5)]	Morse [eq (6)]	X6 (300K)
Lattice Parameters (Å)						
a	2.4684	yes	2.4683	2.4685	2.4682	2.4612
c	6.6721	no	6.6726	6.6720	6.6730	6.7090
Elastic Constants (GPa)						
C_{11}	1126±22	no	1056.2	1130.6	1136.2	1024.7
C_{12}	200±20	no	180.87	186.8	197.9	164.10
C_{33}	40.7±1.1	no	40.75	41.46	40.89	36.01
C_{44}	4.51±0.5	no	0.452	0.468	0.738	0.325
C_{13}	39.5±40	no	0.452	0.468	0.738	0.324
C_{66}	(440)	no	437.6	471.9	469.2	432.4
Young's Moduli (GPa)						
E_1	1020±30	no	1025.2	1099.7	1101.8	998.5
E_3	36.5±1.0	no	40.75	40.9	43.5	29.01
Poisson Ratios						
μ_{21}	(0.16±0.06)	no	0.171	0.165	0.174	0.160
μ_{31}	(0.34±0.08)	no	0.171	0.165	0.174	0.160
μ_{13}	(0.012±0.003)	no	0.0004	0.0004	0.0006	0.0003
Compressibility (GPa^{-1})						
β	0.026[b]	no	0.0261	0.0259	0.0244	0.036
Lattice Modes (cm^{-1})[c]						
E_{2g}	(45)	no	14.2	14.5	18.2	12.2
B_{2g}	(134)	no	137.5	139.4	139.5	116.0
A_{2u}	868	yes	867.7	867.6	867.3	863.6
B_{2g}	868	no	869.9	869.3	870.2	863.9
E_{2g}	(1575[d])	yes	1588.6	1587.6	1585.9	1567.9
E_{1u}	1590[e]	no	1588.8	1587.8	1586.1	1581.7
Cohesive energy (kcal/mol/C)						
E_{coh}	1.27±0.12	no	1.18	1.66	1.00	1.20
$E_{rhom-Ehex}$			0.0104	0.0117	0.008	0.0184

[a] Parenthesis indicate values at room temperature
[b] Reference[78]
[c] See discussion in Section II.C. 8 cm^{-1} was added to correct from 300K to 0K
[d] Reference[120]
[e] Reference[111]
[f] Reference[59]

Table 4. Comparison of QM-FF with Published van der Waals Parameters for Carbon

	QM-FF (X6)	Dreiding[a]	MM2/ MM3[b]	CHARMM[c]	AMBER[d]	Girifalco[e]	Ulbricht[f]	GraFF[g]
Force field Parameters								
van der Waals								
R_v (Å)	3.7248	3.88	3.8005	3.9848	3.8160	3.8321	3.7810	3.8050
D_v (kcal/mol)	0.0635	0.095	0.0515	0.0700	0.0860	0.0847	0.0608	0.0692
ζ	17.64	-	12.5	-	-	-	-	-
Valence								
R_b (Å)	1.4046	1.39925	1.4219	1.38113	1.40668	1.40577	1.40769	1.4114
Properties								
Lattice Parameters (Å)								
c	6.6726 +0.01%	6.7773 +1.51%	6.5923 -1.20%	7.009 +5.05%	6.7010 +0.43%	6.6732 +0.02%	6.640 -0.48%	6.676 +0.06%
Elastic Constants (GPa)								
C_{33}	40.75 +0.12%	60.13 +48%	24.12 -41%	42.55 +4%	50.24 +23%	49.56 +21%	35.16 -14%	40.51 -0.5%
C_{44}	0.452	0.422	0.168	0.499	0.582	0.598	0.425	0.401
C_{13}	0.452	0.423	0.167	0.500	0.583	0.598	0.426	0.401
Compressibility (GPa^{-1})								
β	0.0261	0.0181	0.0431	0.0254	0.0213	0.0216	0.0306	0.0262
Lattice Modes (cm^{-1})								
E_{2g}	14.2	13.7	13.1	14.6	16.1	16.3	13.8	13.4
B_{2g}	137.5	167.2	161.0	138.7	154.2	152.8	129.6	138.7
A_{2u}	867.7	1623.3		976.0	1034.6	1035.8	949.4	867.8
Cohesive energy (kcal/mol/C)								
E_{coh}	1.18	1.94	1.021	1.81	2.05	2.04	1.41	1.64

[a] Reference[35]
[b] Reference[33-34]
[c] Reference[29-32]
[d] Reference [25-28]
[e] Reference[42]
[f] Reference[43]
[g] Reference[20]

Table 5. Phonon frequencies of graphite and derivatives at the high-symmetry points A, Γ, M, and K, in cm^{-1}. The X6 QM-FF (0K) is compared to results from ab-initio DFT studies and experiment.

	Hexagonal			Rhombohedral		Experiment
Functional	QM-FF	LDA†	GGA†	QM-FF	GGA†	
Lattice constants						
***a*(Å)**	2.4612	2.4395	2.4607	2.4612	2.4607	
***c*(Å)**	6.7090	6.6274	6.6274	6.7090	6.6274	
$A_{TA/TO'}$	10	31	29	12		35[a]
$A_{LA/LO'}$	96	80	96	68		89[a]
A_{LO}	825	897	878	864		
A_{TO}	1598	1598	1564	1588		
$\Gamma_{LO'}$	14	44	41	11	35	49[a]
$\Gamma_{TO'}$	138	113	135	118	117	95[b], 126[a]
Γ_{ZO}	868	899	879	864	879	861[b]
$\Gamma_{LO/TO}$	1588	1593	1559	1589	1559	1590[b], 1575[f]
	1589	1604	1567	1604		
M_{ZA}	479	478	477	440	479	471[a], 465[b], 451[d]
M_{TA}	869	630	626	865	626	630[d], 800[f]
M_{ZO}	629	637	634	640	635	670[b]
M_{LA}	1341	1349	1330	1388	1330	1290[c]
M_{LO}	1350	1368	1342	1404	1344	1321[c]
M_{TO}	1468	1430	1394	1432	1394	1388[c], 1389[b]
K_{ZA}	562	540	540	494	535	482[d], 517[d], 530[e]
K_{ZO}	562	544	542	494	539	588[d], 627[e]
K_{TA}	1164	1009	998	1139	998	
$K_{LA/LO}$	1164/1280	1239	1216	1147/1157	1216	1184[c], 1202[c]
K_{TO}	1424	1359	1319	1292	1319	1313[d], 1291[e]

† LDA and GGA DFT results from Reference[4]

[a]Reference[80], [b]Reference[111], [c]Reference[112], [d]Reference[12], [e]Reference[121], [f]Reference[120] as reported in Reference[4]

[g] Reference[111]

X. Figures

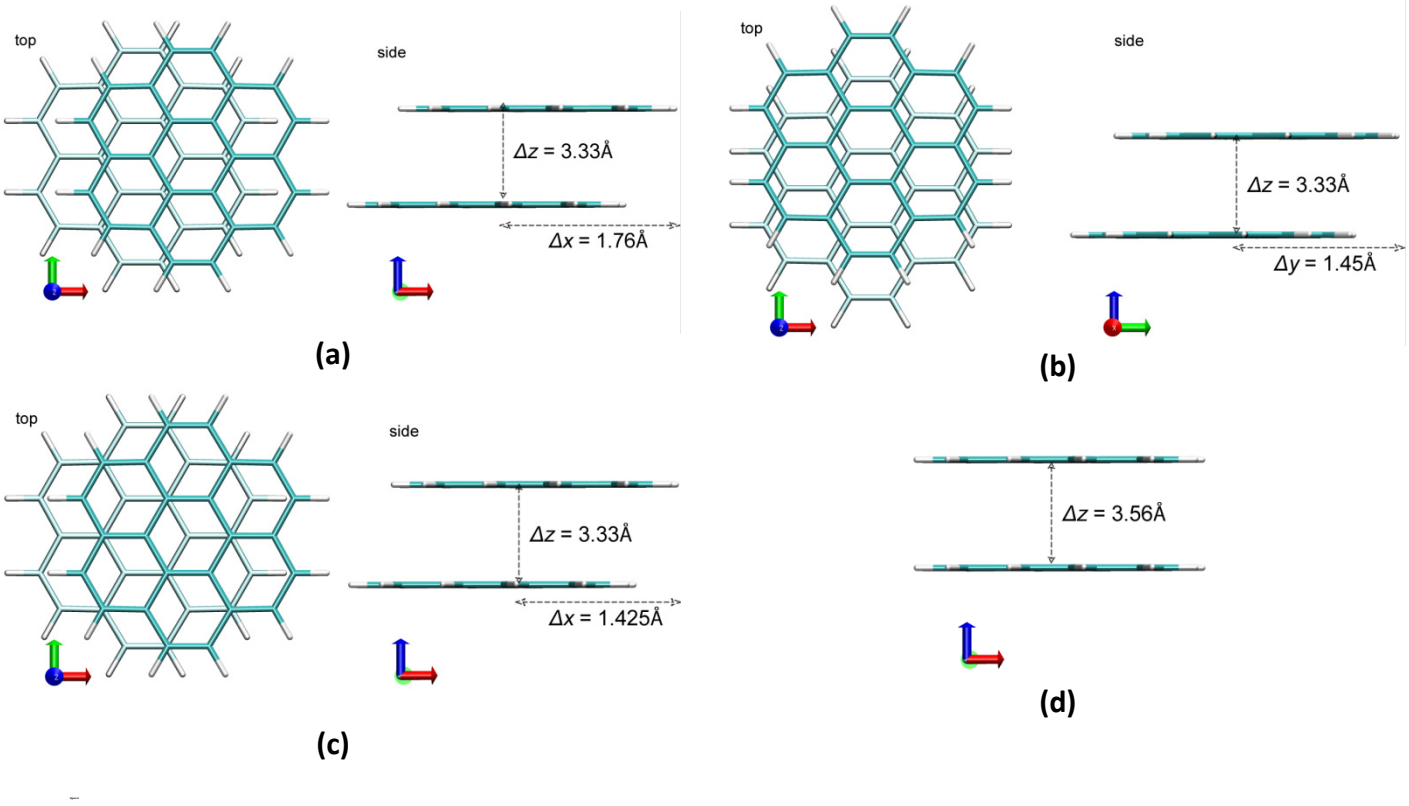


Figure 1. Side projection of the lowest energy structure for each set of coronene structure (PD-X – (a) and PD-Y – (c)) used to obtain the van der Waals parameters for QM-FF. The top view of each structure is shown to the right. Also shown are corresponding grapite structure (c) and the high energy eclipsed structure (d). For the PD-X and PD-Y structures, the displacement vectors values are given. The dispersion curves are obtained by varying the z coordinate of the top coronene structure while holding the bottom fixed.

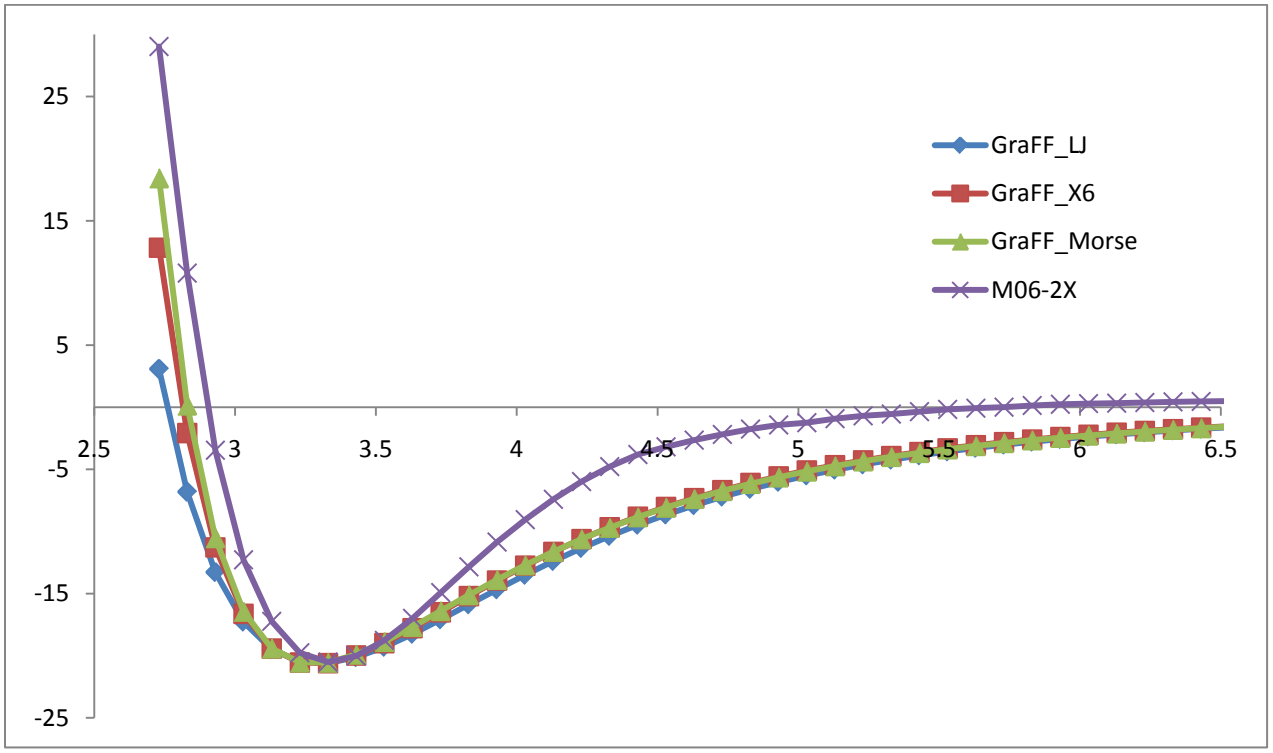


Figure 2. PD-X van der Waals curves for QM-FF (X6, Morse and LJ) compared to M06-2X. The optimization procedure minimized the differences for the 5 points around the minima, combined with those of the PD-Y structure (the curve for the PD-Y structure is similar and not shown).

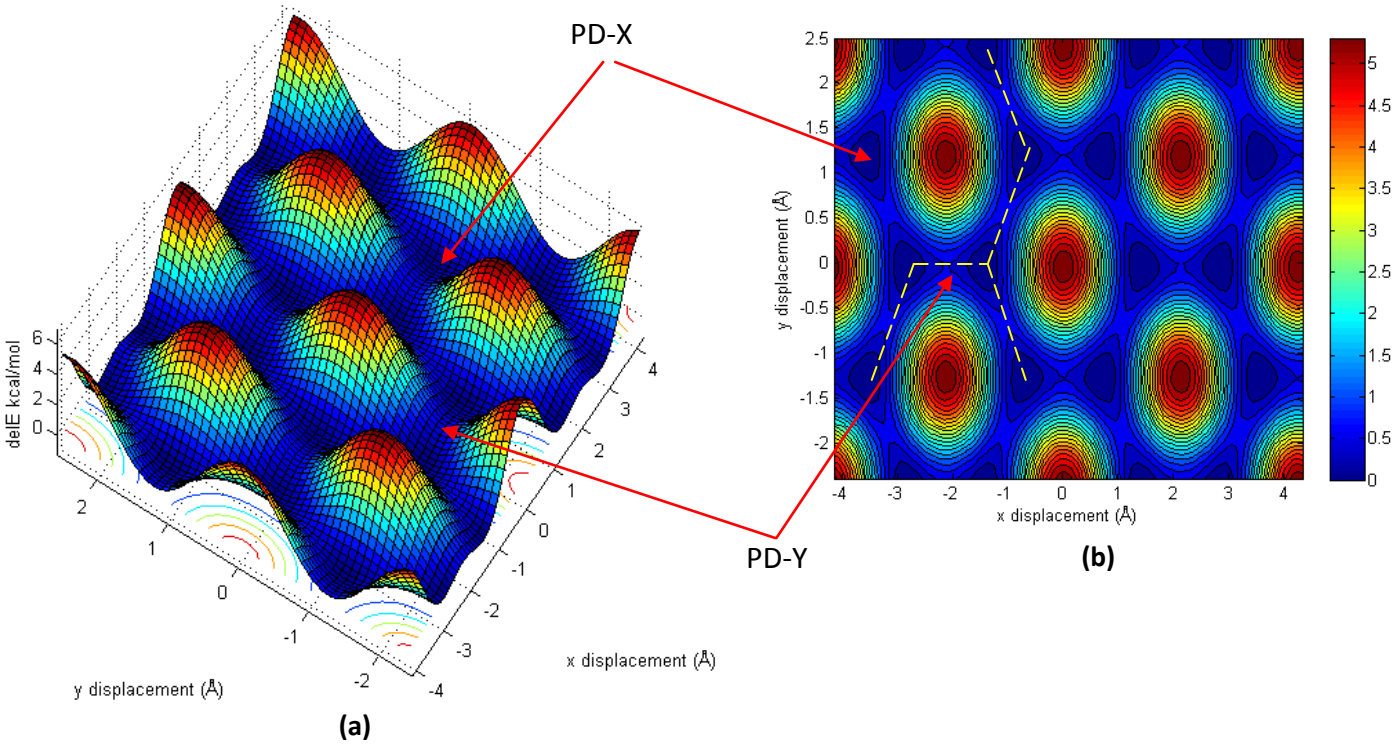


Figure 3. (a) Potential energy surface and **(b)** contour plot of sliding for the top layer of graphite, obtained using the QM-FF X6 potential. The unit cell has two layers of 800 carbons each. The low energy (graphite like) PD-X and PD-Y structures are indicated.

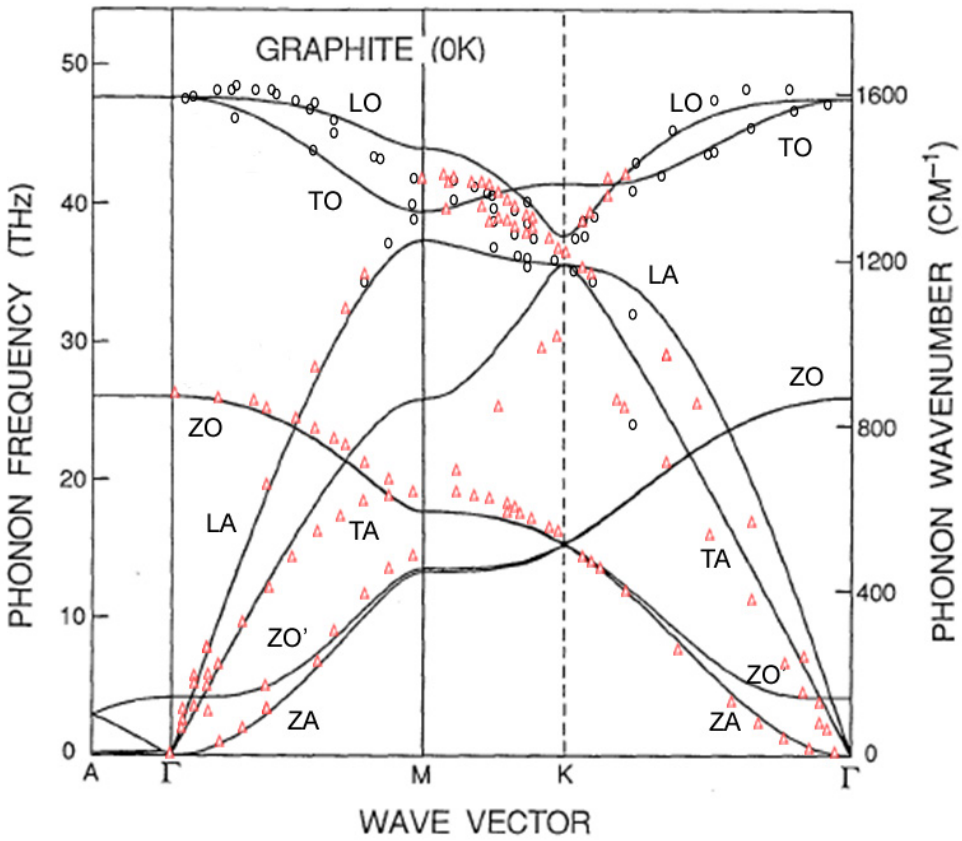


Figure 4. Phonon dispersion curve for all vibrational modes of (hexagonal) graphite at 0K using the X6 potential. Experimental results (300K) from inelastic x-ray scattering are shown: red triangles from reference[122], black circles from reference[112]

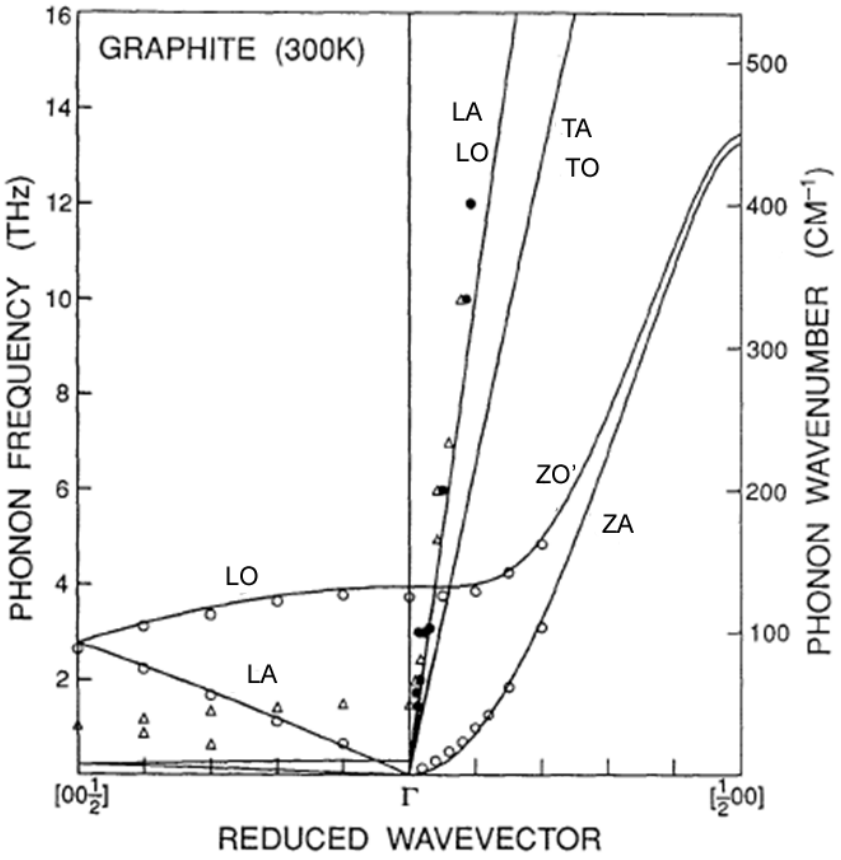


Figure 5. Phonon dispersion curves for the low frequency modes of (hexagonal) graphite at 300K. Experimental data (symbols) from reference[79-80,112,122].

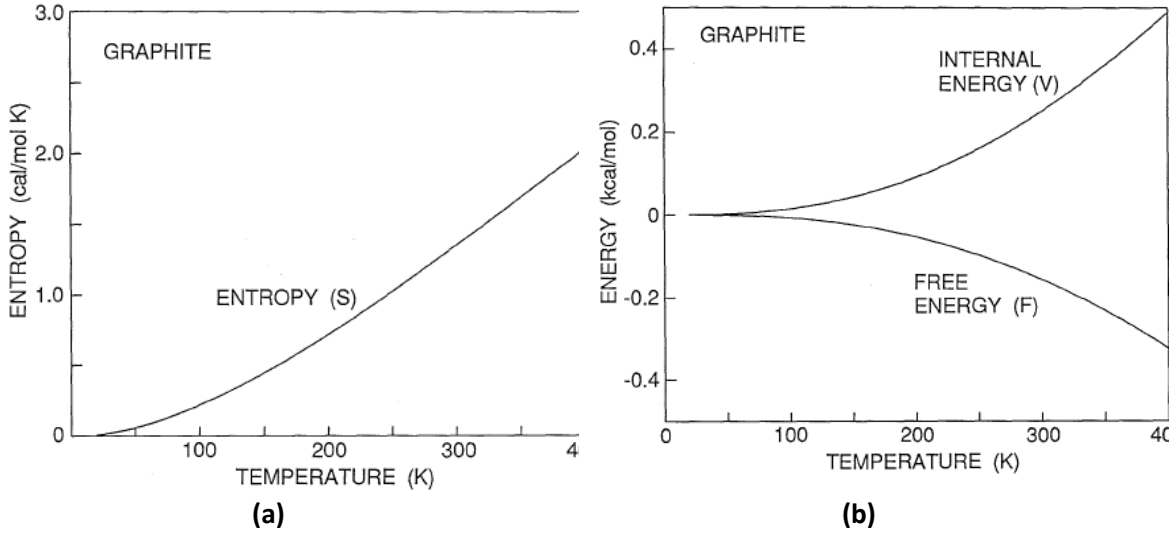


Figure 6. (a) Vibrational entropy (S) of graphite (b) Vibrational energy (V) and free energy (F) of graphite

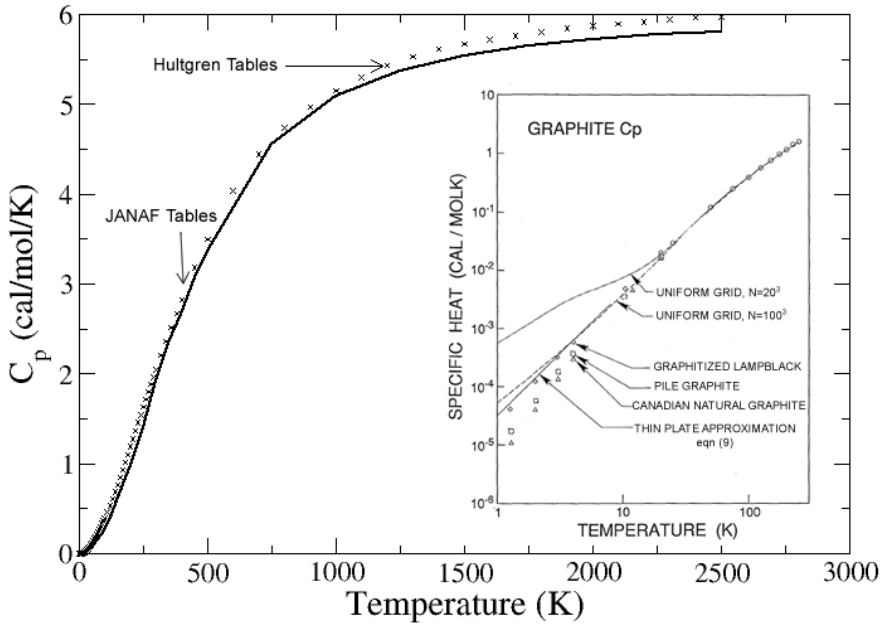
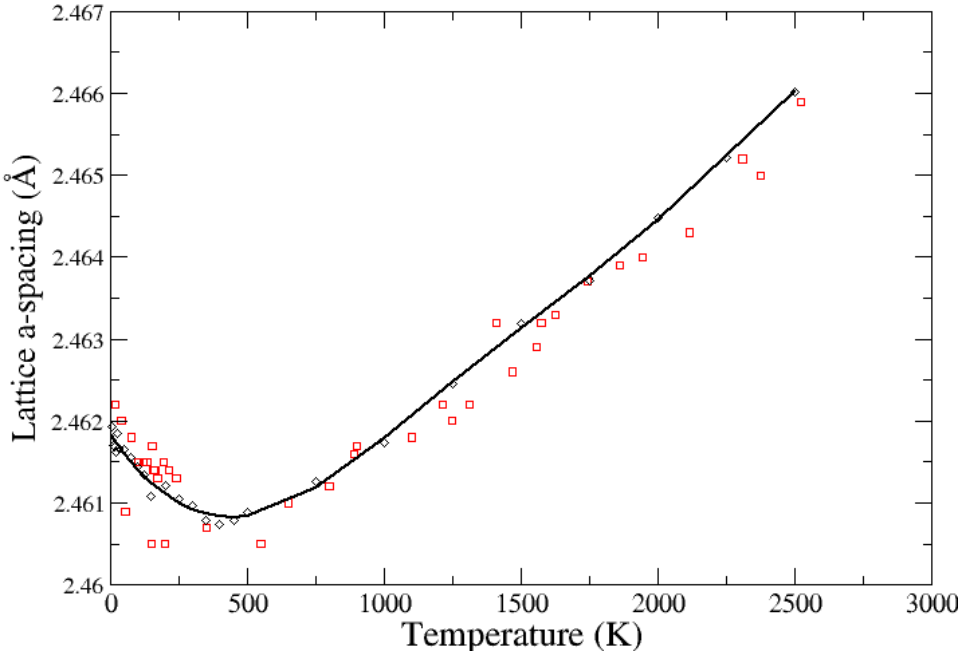
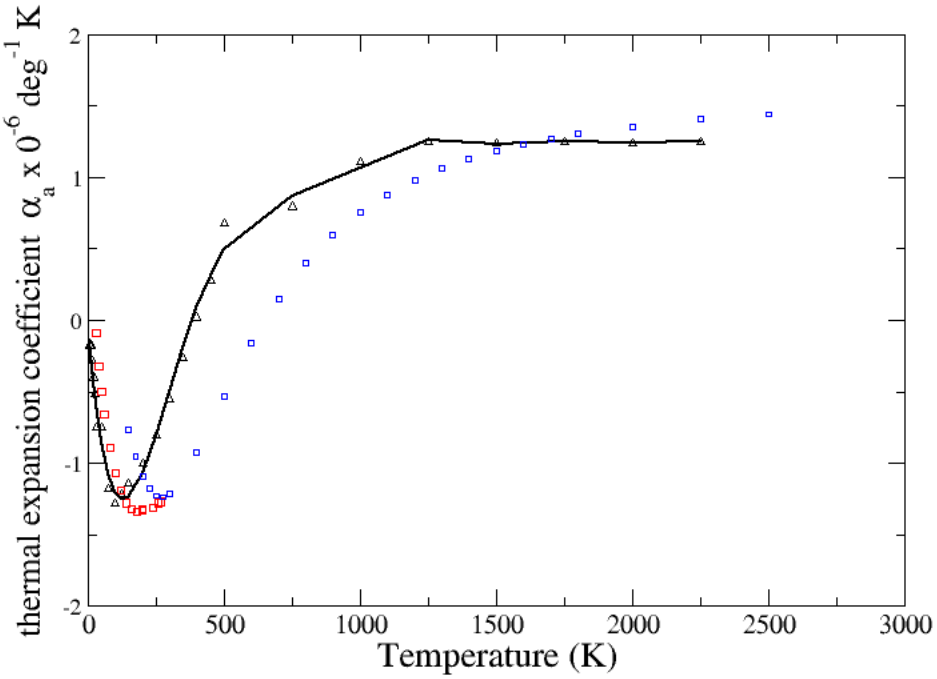


Figure 7. Specific heat of hexagonal graphite as computed with the QM-FF X6 potential. Low temperature (< 10K) results obtained from the thin-plate approximation, other results obtained from the uniform grid method. Experimental results from different sources are indicated, as reported in references[88,92].). Values for rhombohedral graphite is not plotted, since the lines would be essentially super-imposed on the hexagonal graphite lines.



(a)



(b)

Figure 8. (a) In-plate (a) lattice parameter of graphite as a function of temperature, calculated with the QM-FF X6 potential. The experimental results[115] (red squares) are compared the calculated values (black triangles). The solid black line is the least squares line by cubic spline regression. **(b)** Calculated in-plane coefficient of thermal expansion. The experimental results (red and blue squares) are shown, as reported in references[77,117]. The best fit of the data (black line) is used to compute the coefficient of thermal expansion in the *ab* and *c* directions.

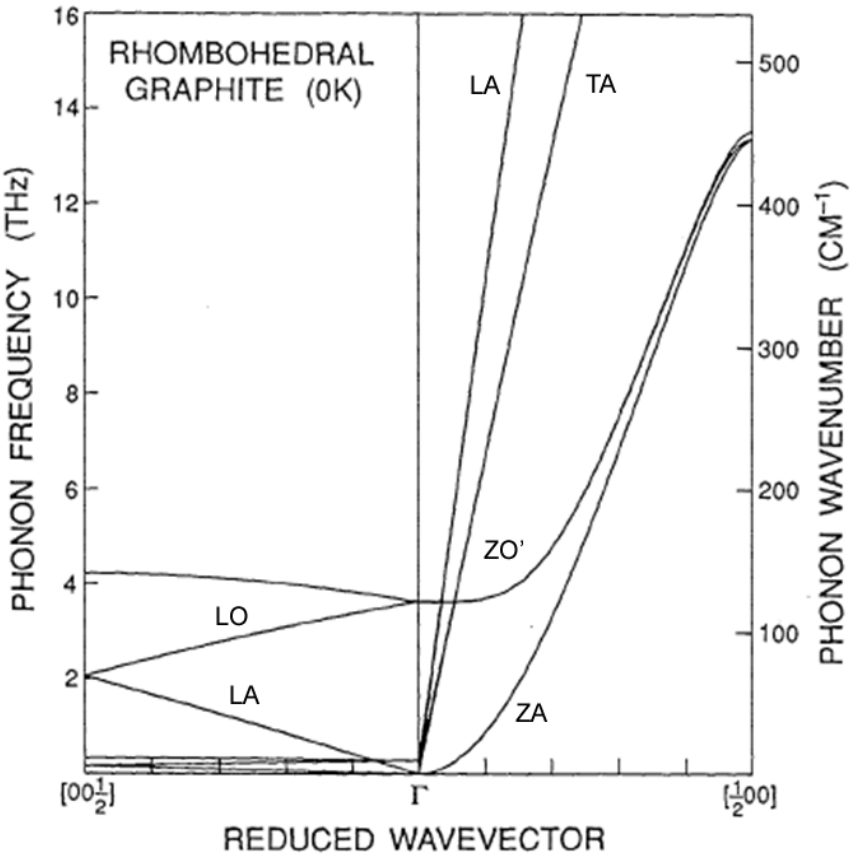


Figure 9. Phonon dispersion curves for the low frequency modes of rhombohedral graphite at 0K (using the X6 potential with ζ = 17.6)

SI. Supplementary Tables

Table S1. Experimental structural and elastic properties of Graphite.

	300K[a]	300K-> 0K[b]	0K
Lattice Parameters (Å)[c]			
a	2.4612	0.0072	2.4684
c	6.7090	-0.0369	6.6721
Elastic Constants (GPa)			
C_{11}	1060±20	66±8	1126±22
C_{12}	180±20	20±2.5	200±20
C_{33}	36.5±1	4.2±0.5	40.7±1.1
C_{44}	4.5±0.5	0.01±0.001	4.51±0.5
C_{13}	15±5	23.8±40	39.5±40
Young's Moduli (GPa)			
E_1	1020±30		
E_3	36.5±1.0		
Poisson Ratios			
μ_{21}	0.16±0.06		
μ_{31}	0.34±0.08		
μ_{13}	0.012±0.003		

[a] See Donohue[76] and Blakslee et al.[78]
[b] See Bailey and Yates[77] and Gauster and Fritz[86]
[c] 298K was used instead of 300K

Table S2. Charges (e^-) on the PD-X orientation of the coronene dimer at equilibrium, determined using three different charge schemes. The 6-311G(2df,2p) basis set was used. The columbic component of the interacting coronene units was determined using the ESP derived charges of the coronene monomer.

	Monomer			|ΔDimer[b]|		
Atom[a]	**Mulliken**	**ESP**	**CM4[c]**	**Mulliken**	**ESP**	**CM4[d]**
C1	0.02	-0.012	0.01	0.0075	0.0026	0.004
C2	-0.02	0.172	-0.01	0.0027	0.0030	0.004
C3	-0.16	-0.24	-0.07	0.0030	0.0046	0.007
H	0.16	0.16	0.07	0.0032	0.0031	0.003

[a] C1 – 6 carbons of the innermost benzene ring, C2 – 6 sp^3 carbons not bound to H atoms, C3 – remaining 12 outermost carbons bound to H atoms

[b]Average dimer charges of the PD-X configuration at the interaction minima (Δx = 1.76Å, Δz = 3.33Å)
[c]As reported in Ref.[75]
[d]Absolute value of change in dimer charge was reported

Table S3. Interaction Energies of the three coronene dimer deemed relevant to the graphite structure. The van der Waals parameters were obtained by fitting the bottom to the PD-X and PD-Y curves to X6 and LJ functions.

PD-X					PD-Y					Eclipsed				
dist	***ΔE***	***ΔE_cp***	***Coul***	***VDW***	***dist***	***ΔE***	***ΔE_cp***	***coul***	***VDW***	***dist***	***ΔE***	***ΔE_cp***	***coul***	***VDW***
2.73	23.85	31.35	2.72	28.63	2.73	34.60	42.12	3.80	38.32	3.06	12.25	19.56	5.98	13.58
2.83	5.94	13.07	2.68	10.39	2.83	13.19	20.35	3.63	16.72	3.16	-0.45	6.29	5.45	0.84
2.93	-8.05	-1.24	2.62	-3.86	2.93	-4.14	2.77	3.46	-0.69	3.26	-8.05	-1.84	4.98	-6.82
3.03	-16.69	-10.12	2.56	-12.68	3.03	-14.18	-7.61	3.30	-10.91	3.36	-12.49	-6.70	4.57	-11.27
3.13	-21.41	-15.14	2.48	-17.63	3.13	-20.08	-13.82	3.14	-16.96	3.46	-14.59	-9.16	4.21	-13.37
3.23	-23.71	-17.75	2.41	-20.16	3.23	-22.86	-16.89	2.99	-19.89	3.56	-15.33	-10.29	3.88	-14.17
3.33	-24.21	-18.58	2.33	-20.91	3.33	-23.70	-18.06	2.85	-20.91	3.66	-15.04	-10.29	3.59	-13.88
3.43	-23.44	-18.17	2.25	-20.42	3.43	-23.14	-17.88	2.71	-20.59	3.76	-13.87	-9.50	3.34	-12.83
3.53	-21.88	-16.99	2.17	-19.16	3.53	-21.74	-16.88	2.58	-19.46	3.86	-12.48	-8.41	3.10	-11.51
3.63	-19.85	-15.30	2.08	-17.39	3.63	-19.74	-15.24	2.46	-17.70	3.96	-10.84	-7.03	2.90	-9.92
3.73	-17.56	-13.32	2.01	-15.32	3.73	-17.59	-13.40	2.34	-15.74	4.06	-9.30	-5.75	2.71	-8.46
3.83	-15.27	-11.32	1.93	-13.25	3.83	-15.38	-11.45	2.23	-13.68	4.16	-7.71	-4.43	2.54	-6.96
3.93	-13.13	-9.40	1.85	-11.25	3.93	-13.26	-9.52	2.12	-11.64	4.26	-6.57	-3.42	2.38	-5.80
4.03	-11.22	-7.67	1.78	-9.45	4.03	-11.34	-7.81	2.02	-9.83	4.36	-5.79	-2.72	2.24	-4.96
4.13	-9.48	-6.10	1.71	-7.81	4.13	-9.37	-6.09	1.93	-8.02	4.46	-5.06	-2.10	2.11	-4.20
4.23	-7.91	-4.74	1.64	-6.38	4.23	-7.87	-4.84	1.84	-6.67	4.56	-4.25	-1.50	1.99	-3.49
4.33	-6.56	-3.61	1.57	-5.18	4.33	-6.52	-3.69	1.75	-5.45	4.66	-3.29	-0.80	1.88	-2.68
4.43	-5.46	-2.70	1.51	-4.21	4.43	-5.45	-2.81	1.67	-4.49	4.76	-2.95	-0.61	1.77	-2.38
4.53	-4.74	-2.14	1.45	-3.59	4.53	-4.70	-2.18	1.60	-3.78	4.86	-2.62	-0.37	1.68	-2.05
4.6	-4.13	-1.66	1.3	-3.05	4.6	-4.12	-1.71	1.5	-3.24	4.9	-2.44	-0.26	1.5	-1.85

3			9		3			3		6			9	
4.73	-3.60	-1.26	1.33	-2.59	4.73	-3.56	-1.27	1.46	-2.73	5.06	-2.21	-0.17	1.51	-1.68
4.83	-3.11	-0.88	1.28	-2.16	4.83	-3.06	-0.91	1.39	-2.30	5.16	-1.80	0.12	1.44	-1.31
4.93	-2.73	-0.60	1.23	-1.83	4.93	-2.63	-0.58	1.33	-1.91	5.26	-1.57	0.27	1.37	-1.10
5.03	-2.49	-0.46	1.18	-1.64	5.03	-2.43	-0.49	1.28	-1.77	5.36	-1.54	0.19	1.30	-1.11
5.13	-2.11	-0.19	1.13	-1.32	5.13	-2.05	-0.23	1.22	-1.45	5.46	-1.11	0.47	1.24	-0.77
5.23	-1.81	0.00	1.09	-1.09	5.23	-1.74	-0.03	1.17	-1.20	5.56	-0.89	0.56	1.18	-0.62
5.33	-1.60	0.09	1.04	-0.95	5.33	-1.52	0.08	1.12	-1.04	5.66	-0.74	0.58	1.13	-0.54
5.43	-1.33	0.25	1.00	-0.75	5.43	-1.30	0.18	1.07	-0.89	5.76	-0.63	0.59	1.08	-0.49
5.53	-1.07	0.39	0.96	-0.57	5.53	-1.02	0.35	1.03	-0.68	5.86	-0.36	0.73	1.03	-0.30
5.63	-0.90	0.45	0.93	-0.48	5.63	-0.85	0.41	0.99	-0.58	5.96	-0.21	0.78	0.98	-0.20
5.73	-0.74	0.50	0.89	-0.39	5.73	-0.68	0.46	0.95	-0.48	6.06	-0.13	0.75	0.94	-0.19
5.83	-0.54	0.59	0.85	-0.26	5.83	-0.51	0.53	0.91	-0.38	6.16	-0.06	0.75	0.90	-0.15
5.93	-0.37	0.66	0.82	-0.17	5.93	-0.33	0.60	0.87	-0.27	6.26	0.08	0.78	0.86	-0.08
6.13	-0.15	0.69	0.76	-0.07	6.13	-0.11	0.63	0.80	-0.17	6.46	0.26	0.80	0.79	0.01
6.33	0.05	0.74	0.70	0.03	6.33	0.09	0.67	0.74	-0.07	6.66	0.36	0.78	0.73	0.05
6.53	0.19	0.74	0.65	0.09	6.53	0.25	0.69	0.69	0.01	6.86	0.45	0.77	0.67	0.10
6.73	0.30	0.74	0.60	0.13	6.73	0.36	0.68	0.64	0.05	7.06	0.50	0.74	0.62	0.12
6.83	0.34	0.73	0.58	0.14	6.83	0.40	0.68	0.61	0.06					
6.93	0.51	0.85	0.56	0.29	6.93	0.42	0.65	0.59	0.06					
7.03	0.51	0.82	0.54	0.28	7.03	0.52	0.72	0.57	0.15					

Table S4. Mechanical properties at 0K for (hexagonal) Graphite using the QM-FF X6 potential with $\zeta = 17.6$

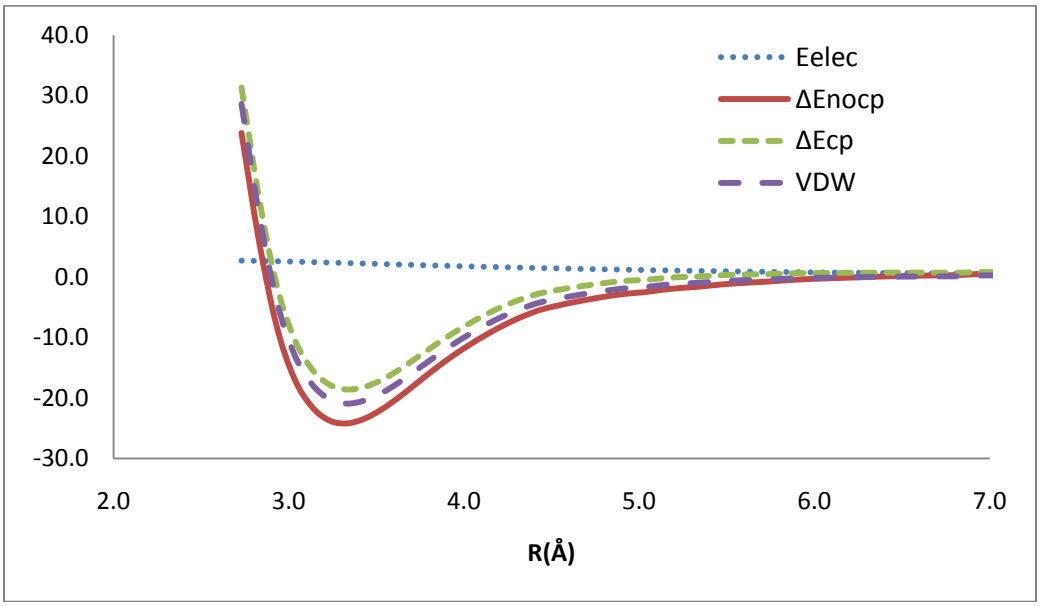
C_{ij} Elastic Constant Matrix (Stiffness) GPa

$$\begin{pmatrix} 1056.2 & 180.9 & 0.455 & 0 & 0 & 0 \\ & 1056.2 & 0.455 & 0 & 0 & 0 \\ & & 40.68 & 0 & 0 & 0 \\ & & & 0.442 & 0 & 0 \\ & & & & 0.442 & 0 \\ & & & & & 437.65 \end{pmatrix}$$

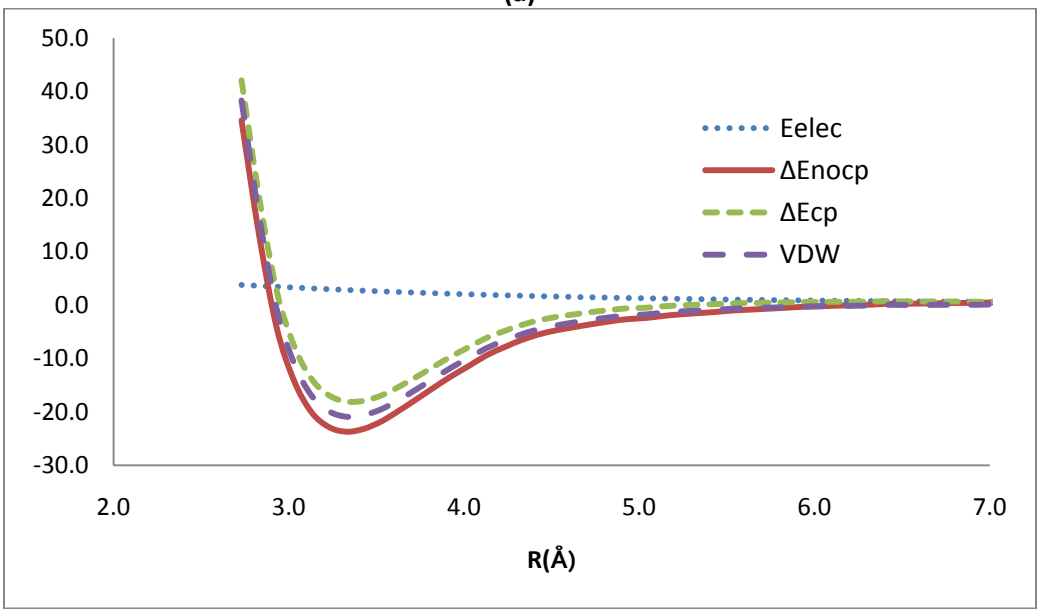
S_{ij} Compliance Matrix GPa^{-1}

$$\begin{pmatrix} 9.754 \times 10^{-4} & -1.670 \times 10^{-4} & 9.04 \times 10^{-6} & 0 & 0 & 0 \\ & 9.75 \times 10^{-4} & 9.04 \times 10^{-6} & 0 & 0 & 0 \\ & & 2.46 \times 10^{-2} & 0 & 0 & 0 \\ & & & 2.26 & -3.92 \times 10^{-4} & 0 \\ & & & & 2.26 & 0 \\ & & & & & 2.29 \times 10^{-2} \end{pmatrix}$$

SII. Supplementary Figures



(a)



(b)

Figure S1. Comparison between M06-2X and QM-FF (X6) for **(a)**PD-X and **(b)** PD-Y coronene dimer. The total interaction energy (ΔEnocp) and the counterpoised corrected interaction energy (ΔEcp) curves are shown. The atom centered electrostatic contributions (Eelec) are removed using the ESP charge scheme. We obtain the Dv and Rv for the QM-FF X6 potential by fitting the five points around the PD-X minima. The ζ curvature parameter is obtained by obtaining the correct energy splitting with the PD-Y curve.

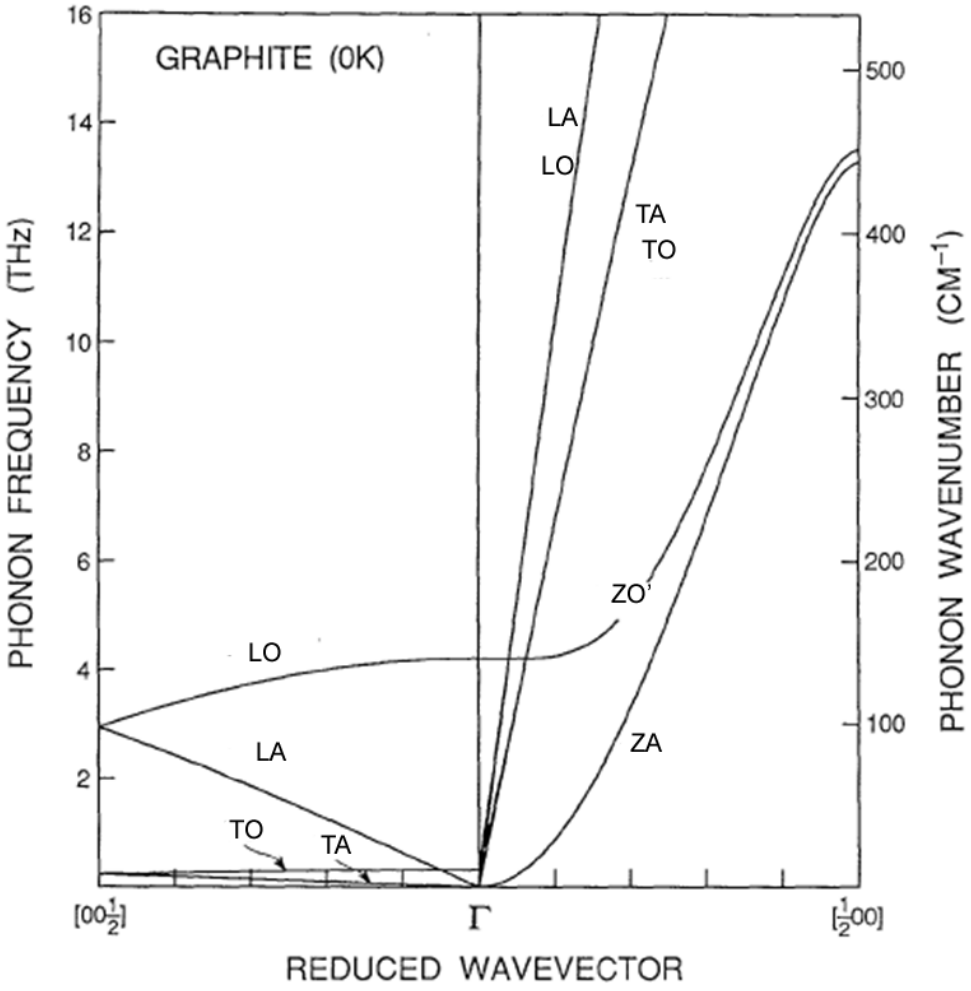


Figure S2. Phonon dispersion curve for low frequency modes of (hexagonal) graphite at 0K (using the X6 potential with ζ = 17.6)

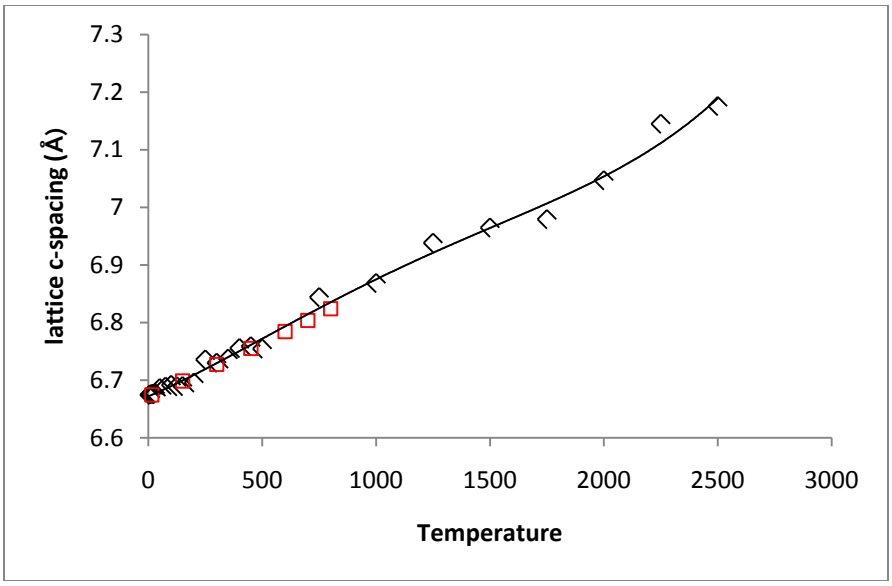


Figure S3. Out of plane (c) lattice parameter of graphite as a function of temperature, calculated with the QM-FF X6 potential. The experimental results[115] (red squares) are compared the calculated values (black triangles). The solid black line is the least squares line by cubic spline regression.

(1) Qian, D.; Wagner, G. J.; Liu, W. K.; Yu, M.-F.; Ruoff, R. S. *Applied Mechanics Reviews* **2002**, *55*, 495.
(2) Dunlap, B. I.; Boettger, J. C. *Journal of Physics B-Atomic Molecular and Optical Physics* **1996**, *29*, 4907.
(3) Kahaly, M. U.; Waghmare, U. V. *J. Nanosci. Nanotechnol.* **2007**, *7*, 1787.
(4) Mounet, N.; Marzari, N. *Physical Review B* **2005**, *71*.
(5) Wirtz, L.; Rubio, A. *Solid State Communications* **2004**, *131*, 141.
(6) Nagamiya, T.; Komatsu, K. *J. Chem. Phys.* **1954**, *22*, 1457.
(7) Komatsu, K. *Journal of the Physical Society of Japan* **1955**, *10*, 346.
(8) Komatsu, K.; Nagamiya, T. *Journal of the Physical Society of Japan* **1951**, *6*, 438.
(9) Liu, A.; Stuart, S. J. *J. Comput. Chem.* **2008**, *29*, 601.
(10) Los, J. H.; Ghiringhelli, L. M.; Meijer, E. J.; Fasolino, A. *Physical Review B* **2005**, *72*.
(11) Tewary, V. K.; Yang, B. *Physical Review B* **2009**, *79*.
(12) Yanagisawa, H.; Tanaka, T.; Ishida, Y.; Matsue, M.; Rokuta, E.; Otani, S.; Oshima, C. *Surface and Interface Analysis* **2005**, *37*, 133.
(13) Zimmermann, J.; Pavone, P.; Cuniberti, G. *Physical Review B* **2008**, *78*.
(14) Brenner, D. W. *Physical Review B* **1990**, *42*, 9458.
(15) Tersoff, J. *Phys. Rev. Lett.* **1988**, *61*, 2879.

(16) Stillinger, F. H.; Weber, T. A. *Physical Review B* **1985**, *31*, 5262.

(17) Aizawa, T.; Souda, R.; Otani, S.; Ishizawa, Y.; Oshima, C. *Physical Review B* **1990**, *42*, 11469.

(18) Perebeinos, V.; Tersoff, J. *Physical Review B (Condensed Matter and Materials Physics)* **2009**, *79*, 241409.

(19) Collignon, B.; Hoang, P. N. M.; Picaud, S.; Liotard, D.; Rayez, M. T.; Rayez, J. C. *Journal of Molecular Structure: THEOCHEM* **2006**, *772*, 1.

(20) Guo, Y. J.; Karasawa, N.; Goddard, W. A. *Nature* **1991**, *351*, 464.

(21) Hummer, G.; Rasaiah, J. C.; Noworyta, J. P. *Nature* **2001**, *414*, 188.

(22) Chen, R. J.; Bangsaruntip, S.; Drouvalakis, K. A.; Kam, N. W. S.; Shim, M.; Li, Y. M.; Kim, W.; Utz, P. J.; Dai, H. J. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 4984.

(23) Meyer, E. A.; Castellano, R. K.; Diederich, F. *Angew. Chem.-Int. Edit.* **2003**, *42*, 1210.

(24) Karasawa, N.; Dasgupta, S.; Goddard, W. A. *J. Phys. Chem.* **1991**, *95*, 2260.

(25) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E.; Wang, J.; Ross, W. S.; Simmerling, C.; Darden, T.; Merz, K. M.; Stanton, R. V.; al., e.; 7 ed.; University of California, San Francisco.: 1999.

(26) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817.

(27) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157.

(28) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *Journal of the American Chemical Society* **1995**, *117*, 5179.

(29) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.

(30) Mackerell, A. D.; Wiorkiewiczkuczera, J.; Karplus, M. *Journal of the American Chemical Society* **1995**, *117*, 11946.

(31) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.

(32) Jorgensen, W. L.; Severance, D. L. *Journal of the American Chemical Society* **1990**, *112*, 4768.

(33) Allinger, N. L. *Journal of the American Chemical Society* **1977**, *99*, 8127.

(34) Allinger, N. L.; Lii, J. H. *J. Comput. Chem.* **1987**, *8*, 1146.

(35) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. *J. Phys. Chem.* **1990**, *94*, 8897.

(36) Goddard, W. A. *Handbook of nanoscience, engineering, and technology*; CRC Press: Boca Raton, 2003.

(37) Wu, J.; Pisula, W.; Mullen, K. *Chemical Reviews* **2007**, *107*, 718.

(38) Stankovich, S.; Dikin, D. A.; Dommett, G. H. B.; Kohlhaas, K. M.; Zimney, E. J.; Stach, E. A.; Piner, R. D.; Nguyen, S. T.; Ruoff, R. S. *Nature* **2006**, *442*, 282.

(39) Venema, L. *Nature* **2007**, *446*, 36.

(40) Son, Y.-W.; Cohen, M. L.; Louie, S. G. *Nature* **2006**, *444*, 347.
(41) Barone, V.; Hod, O.; Scuseria, G. E. *Nano Lett.* **2006**, *6*, 2748.
(42) Girifalco, L. A.; Hodak, M.; Lee, R. S. *Physical Review B* **2000**, *62*, 13104.
(43) Ulbricht, H.; Moos, G.; Hertel, T. *Phys. Rev. Lett.* **2003**, *90*, 095501.
(44) Fetzer, J. C. *Polycyclic Aromatic Compounds* **2007**, *27*, 143.
(45) Marzec, A. *Carbon* **2000**, *38*, 1863.
(46) Falconer, J. L.; Madix, R. J. *Journal of Catalysis* **1977**, *48*, 262.
(47) Fichthorn, K. A.; Miron, R. A. *Phys. Rev. Lett.* **2002**, *89*.
(48) Gameson, I.; Rayment, T. *Chem. Phys. Lett.* **1986**, *123*, 150.
(49) Mercedes Rubio; Enrique Ortí; José Sánchez-Marín *Int. J. Quantum Chem.* **1996**, *57*, 567.
(50) Ruuska, H.; Pakkanen, T. A. *The Journal of Physical Chemistry B* **2001**, *105*, 9541.
(51) Rapacioli, M.; Calvo, F.; Spiegelman, F.; Joblin, C.; Wales, D. J. *The Journal of Physical Chemistry A* **2005**, *109*, 2487.
(52) Conney, A. H. *Cancer Research* **1982**, *42*, 4875.
(53) Schmidt, M. W. I.; Noack, A. G. *Global Biogeochemical Cycles* **2000**, *14*, 777.
(54) Pignatello, J. J.; Xing, B. S. *Environmental Science & Technology* **1996**, *30*, 1.
(55) Mastral, A. M.; Callen, M. S. *Environmental Science & Technology* **2000**, *34*, 3051.
(56) Langhoff, S. R. *J. Phys. Chem.* **1996**, *100*, 2819.
(57) Allamandola, L. J.; Tielens, A.; Barker, J. R. *Astrophysical Journal Supplement Series* **1989**, *71*, 733.
(58) Sagan, C. E.; Lippinco.Er; Dayhoff, M. O.; Eck, R. V. *Nature* **1967**, *213*, 373.
(59) Zacharia, R.; Ulbricht, H.; Hertel, T. *Physical Review B* **2004**, *69*, 7.
(60) Lee, N. K.; Kim, S. K. *The Journal of Chemical Physics* **2005**, *122*, 031102.
(61) Møller, C.; Plesset, M. S. *Physical Review* **1934**, *46*, 618.
(62) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M. *The Journal of Chemical Physics* **2004**, *120*, 647.
(63) Zhao, Y.; Truhlar, D. G. *Journal of Chemical Theory and Computation* **2006**, *3*, 289.
(64) Stefan Grimme *J. Comput. Chem.* **2004**, *25*, 1463.
(65) Hohenberg, P.; Kohn, W. *Physical Review B* **1964**, *136*, B864.
(66) Kohn, W.; Sham, L. J. *Physical Review* **1965**, *140*, 1133.
(67) Rydberg, H.; Dion, M.; Jacobson, N.; Schroder, E.; Hyldgaard, P.; Simak, S. I.; Langreth, D. C.; Lundqvist, B. I. *Phys. Rev. Lett.* **2003**, *91*.
(68) Sato, T.; Tsuneda, T.; Hirao, K. *The Journal of Chemical Physics* **2007**, *126*, 234114.
(69) Kohn, W.; Meir, Y.; Makarov, D. E. *Phys. Rev. Lett.* **1998**, *80*, 4153.
(70) Zhang, Y.; Xu, X.; Goddard, W. A. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 4963.

(71) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *Journal of Chemical Theory and Computation* **2006**, *2*, 364.

(72) Zhao, Y.; Truhlar, D. G. *Journal of Chemical Theory and Computation* **2005**, *1*, 415.

(73) Zhao, Y.; Truhlar, D. G. *J. Chem. Phys.* **2006**, *125*.

(74) Zhao, Y.; Truhlar, D. *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)* **2008**, *120*, 215.

(75) Zhao, Y.; Truhlar, D. G. *Journal of Physical Chemistry C* **2008**, *112*, 4061.

(76) Donohue, J. *The Structures of the Elements*; Robert E. Krieger Publishing Co.: Malabar, Florida, 1974.

(77) Bailey, A. C.; Yates, B. *Journal of Applied Physics* **1970**, *41*, 5088.

(78) Blakslee, O. L. *Journal of Applied Physics* **1970**, *41*, 3373.

(79) Dolling, G.; Brockhouse, B. N. *Physical Review* **1962**, *128*, 1120.

(80) Nicklow, R.; Smith, H. G.; Wakabaya.N *Physical Review B* **1972**, *5*, 4951.

(81) Baker, C.; Kelly, A. *Philosophical Magazine* **1964**, *9*, 927.

(82) Ayasse, J. B.; Ayache, C.; Jager, B.; Bonjour, E.; Spain, I. L. *Solid State Communications* **1979**, *29*, 659.

(83) Komatsu, K. *Journal of Physics and Chemistry of Solids* **1958**, *6*, 380.

(84) Komatsu, K. *Journal of Physics and Chemistry of Solids* **1964**, *25*, 707.

(85) Bowman, J. C.; Krumhansl, J. A. *Journal of Physics and Chemistry of Solids* **1958**, *6*, 367.

(86) Gauster, W. B.; Fritz, I. J. *Journal of Applied Physics* **1974**, *45*, 3309.

(87) Nemanich, R. J.; Lucovsky, G.; Solin, S. A. *Solid State Communications* **1977**, *23*, 117.

(88) Nihira, T.; Iwata, T. *Physical Review B* **2003**, *68*.

(89) Pierson, H. O.; Knovel; Noyes Publications: Park Ridge, N.J., 1993.

(90) Desorbo, W.; Tyler, W. W. *J. Chem. Phys.* **1953**, *21*, 1660.

(91) Prophet, H.; Stull, D. R.; United States. National Bureau of, S. *JANAF thermochemical tables*; USGPO: Washington,D.C., 1971.

(92) Hultgren, R. R.; American Society for, M. *Selected values of the thermodynamic properties of the elements*; American Society for Metals: Metals Park, Ohio, 1973.

(93) Nelson, J. B.; Riley, D. P. *Proceedings of the Physical Society* **1945**, *57*, 477.

(94) Schrödinger: New York, NY, 2007.

(95) Hehre, W. J. R., L.; Schleyer, P. v. R.; Pople, J. A.; *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986.

(96) Boys, S. F.; Bernardi, F. *Molecular Physics: An International Journal at the Interface Between Chemistry and Physics* **1970**, *19*, 553

(97) Schwenke, D. W.; Truhlar, D. G. *The Journal of Chemical Physics* **1985**, *82*, 2418.

(98) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833.

(99) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **2005**, *113*, 133.

(100) Powell, M. J. D. *Mathematical Programming* **1977**, *12*, 241.

(101) Dasgupta, S.; Goddard, W. A. *J. Chem. Phys.* **1989**, *90*, 7207.

(102) Karasawa, N.; Goddard, W. A. *J. Phys. Chem.* **1989**, *93*, 7320.
(103) Komatsu, K. *Journal of the Physical Society of Japan* **1955**, *10*, 346.
(104) Plimpton, S. *J. Comput. Phys.* **1995**, *117*, 1.
(105) Hoover, W. G. *Physical Review A* **1986**, *34*, 2499.
(106) Hoover, W. G. *Physical Review A* **1985**, *31*, 1695.
(107) Nose, S. *J. Chem. Phys.* **1984**, *81*, 511.
(108) Nose, S. *Molecular Physics* **1984**, *52*, 255.
(109) Grimme, S.; Muck-Lichtenfeld, C.; Antony, J. *The Journal of Physical Chemistry C* **2007**, *111*, 11199.
(110) Lazzeri, M.; Calandra, M.; Mauri, F. *Physical Review B* **2003**, *68*.
(111) Oshima, C.; Aizawa, T.; Souda, R.; Ishizawa, Y.; Sumiyoshi, Y. *Solid State Communications* **1988**, *65*, 1601.
(112) Maultzsch, J.; Reich, S.; Thomsen, C.; Requardt, H.; Ordejon, P. *Phys. Rev. Lett.* **2004**, *92*.
(113) Billmeyer, F. W. *Journal of Applied Physics* **1957**, *28*, 1114.
(114) Williams, D. E.; Cox, S. R. *Acta Crystallographica Section B-Structural Science* **1984**, *40*, 404.
(115) Kellett, E. A.; Richards, B. P. *Journal of Nuclear Materials* **1964**, *12*, 184.
(116) Riley, D. P. *Proceedings of the Physical Society of London* **1945**, *57*, 486.
(117) Morgan, W. C. *Carbon* **1972**, *10*, 73.
(118) Aljishi, R.; Dresselhaus, G. *Physical Review B* **1982**, *26*, 4514.
(119) Fasolino, A.; Los, J. H.; Katsnelson, M. I. *Nat Mater* **2007**, *6*, 858.
(120) Tuinstra, F.; Koenig, J. L. *J. Chem. Phys.* **1970**, *53*, 1126.
(121) Siebentritt, S.; Pues, R.; Rieder, K. H.; Shikin, A. M. *Physical Review B* **1997**, *55*, 7927.
(122) Mohr, M.; Maultzsch, J.; Dobardzic, E.; Reich, S.; Milosevic, I.; Damnjanovic, M.; Bosak, A.; Krisch, M.; Thomsen, C. *Physical Review B (Condensed Matter and Materials Physics)* **2007**, *76*, 035439.

Chapter 5. The QM-FF force field for Graphite – water interactions

The interaction of water molecules with graphitic surfaces have received much attention lately[1-5], with most attention being paid to confinement in nanotubes[6-9]. In these systems, the water molecules at the surface show dramatically different microscopic structure and dynamical properties than the bulk. Molecular dynamics, relying on smooth analytic potentials, have been the preferred method of theoretical investigations.

The strength of the interaction is therefore crucial. Werder and coworkers[5] showed that the contact angle of water molecules graphitic surface is a dramatic function of interaction strength. In fact, it was demonstrated that for binding energies < 1.45 kcal/mol, the water behaved as a super-hydrophobic fluid (contact angle > 110^{o}), medium binding energies (between 1.45 and 1.95 kcal/mol), the graphite surface is hydrophobic, and for strong interactions (> 2.0 kcal/mol) the surface becomes *hydrophilic*.

Experimentally, the contact angle of water on graphite has also received considerable attentions. Values of 86 – 80^{o} has been obtained by several authors, using the using the tilting plate[10], rising meniscus[11] and captive bubble methods[12]. A lower value of 42 ± 7^{o} has been obtained by Schrader, also using the xx method. Scanning force microscopy recently produces a value of 30^{o} under high humidity, although the author pointed out that contamination of the surface could be responsible for the low value.

We are interested in developing two-body, analytic potentials based on accurate quantum mechanical (QM) calculations. To this end, Feller and Jordan[2] have performed calculations on the binding of water molecules to polycyclic aromatic hydrocarbons[13] (PAH) using Second-order Møller-Plesset perturbation theory[14]. They obtained an extrapolated graphite-water binding energy of -5.0 ± 0.1 kcal/mol, a value that would lead to super *hydrophilic* behavior on graphite. Nevertheless, Pertsin and Grunze[3] developed a two-body potential from these results, showing that the thermodynamics and structure of the water-graphite interfacial region are extremely sensitive to the range and orientation dependence of the model potential. Interestingly, Sudiarta and Geldart[15] used MP2 calculations on hydrogen and fluorine terminated PAHs and determined the binding energy of water to be between 2.49 and 2.32 kcal/mol.

Werder and co-workers[5] used the results of 28 simulations and potentials to show that a binding energy of 1.5 kcal/mol is required to reproduce the experimental contact angle of 86^{o}. The oxygen-carbon and hydrogen-carbon distances were not optimized in their calculations however. We will show that these distances couple directly with the binding energy to determine the nature of the interaction, as was discussed recently by Pertsin and Grunze[3].

We use the recently developed M06-2X[16] hybrid-meta exchange-correlation GGA functional, based on Density Functional Theory (DFT) [17-18], to investigate the binding of water molecules on various symmetric PAH molecules. The M06-2X[16] functional has been shown to be accurate at describing weakly interacting, dispersion dominated and hydrogen bonded systems. Two different binding modes of water on the surface of

these PAH molecules are investigated and we obtain a binding energy of 2.46 kcal/mol for the alpha mode (both protons pointing towards the PAH surface) and 2.16 kcal/mol for the beta mode (protons pointing away from the surface). We match the binding energies and the oxygen and hydrogen distances to a LJ forcefield for use in standard MD/Monte Carlo codes. This forcefield is used to investigate the contact angle on water on graphite, where we obtain a contact angle of 75^{o}, within range of the experimental values of $86^{o} - 40^{o}$.

The remainder of the paper is organized as follows. In Section II, we provide the computational details used in this study. Section III presents the results and comparison to available theory and experiments. We also discuss the implications of these findings. We offer a concluding statement in Section IV.

II. Computational Methods

a) Binding modes on water on PAH molecules

We approximate the binding of water molecule on graphite by observing the binding on symmetric, successively larger PAH molecules (**Figure 1**):

- Benzene - C_6H_6
- Coronene – $C_{24}H_{12}$
- Circumcoronene – $C_{54}H_{18}$
- Di-Circumcoronene - $C_{96}H_{24}$
- Tri-Circumcoronene - $C_{150}H_{30}$

We obtain the binding energy in graphite by extrapolating the binding energies to the infinite case, using a decaying exponential function.

On each of these PAH molecules, we investigate three binding modes (**Figure 2**):

- Alpha - H atoms pointing into ring
- Beta – H atoms pointing away from ring

Table 1: Binding energies of water on PAH surfaces for QM and QM-FF. The atom centered charge are obtained from the ESP and Mulliken schemes and used to produce the electrostatic component of the binding energies. The binding energy of water on graphite **C∞** is obtained by fitting the energies to decaying exponential functions

	BE_CP	**Fit**		**ESP**	**Mulliken**	
			ΔE(coul)	ΔE(vdwl)	ΔE(coul)	ΔE(vdwl)
			ALPHA			
benzene	-3.91	-3.95	-1.97	-1.94	-3.70	-0.21
coronene	-3.66	-3.65	-0.89	-2.77	-6.35	2.69
circumcoronene	-3.20	-2.93	-0.49	-2.71	-5.25	2.05
di-circumcoronene	-2.99	-2.72	-0.40	-2.59	-4.65	1.66
tri-circumcoronene	-2.61	-2.62	-0.13	-2.77	-5.46	2.56
C∞		**-2.46**				
			BETA			
Benzene	0.64	0.64	3.00	-2.36	2.64	-2.00
Coronene	-1.29	-1.35	0.80	-2.09	1.13	-2.42
Circumcoronene	-1.60	-1.75	0.37	-1.97	-0.62	-0.98
di-circumcoronene	-2.20	-1.92	0.22	-2.42	-0.65	-1.55
tri-circumcoronene	-2.03	-2.00	-0.02	-2.25	-0.49	-1.78
C∞		**-2.16**				

- Gamma – 1 H atom perpendicular to ring (hydrogen bonding)

b) Electronic Structure Calculations

The M06-2X[16] flavor of DFT is a new generation hybrid meta-GGA exchange-correlation functional with 54% HF exchange that has be shown[19] to describe weakly bound and hydrogen bonded complexes accurately near equilibrium. Our calculations used the M06-2X implementation in the Jaguar7.0[20] QM code with the with the augmented polarized triple-ζ 6-311+G(2df,2p)[21]. We used ultra-fine, rectangular grids for the self consistent field calculations, with a convergence criterion of $5x10^{-6}$ Hartree for the energy convergence and $5x10^{-7}$ rms for the density matrix. We removed bias due to basis set superposition error (BSSE) by performing counterpoise corrections (cp)[22-23]. We take the cp binding energy to be the best energy in our systems.

Table 2: Parameters for the QM-FF water-carbon forcefield. Parameters for other popular water models are also presents, obtained from geometric combination rules with our GraFF-LJ carbon forcefield: D_c = 0.0692 kcal/mol, R_c = 3.805Å. We also compare the calculated binding energy (ΔE) and center of mass displacement of the water to the surface (ΔR). We find that none of other potential to be as accurate as QM-FF in reproducing the QM results.

	D_o	R_o	D_h	R_h	$D_{O\text{-}C}$	$R_{O\text{-}C}$	$D_{H\text{-}C}$	$R_{H\text{-}C}$	ΔE (kcal/mol)		ΔR (Å)	
									Alpha	**Beta**	**Alpha**	**Beta**
QM									**-2.72**	**-2.42**	**2.88**	**3.25**
SPC-EW	0.1553	3.5537	0.00	0.0	0.1123	3.6791	0.0000	1.9025	**-1.64**	**-0.08**	**2.89**	**3.25**
F3C	0.1848	3.5532	0.01	0.9	0.1029	3.6794	0.0083	2.3525	**-1.98**	**-0.17**	**2.87**	**3.25**
TIP3P	0.1020	3.5365	0.00	0.0	0.0834	3.6708	0.0000	1.9025	**-1.33**	**-0.12**	**2.88**	**3.26**
TIP4P	0.1628	3.5519	0.00	0.0	0.1054	3.6784	0.0000	1.9025	**-1.68**	**-0.09**	**2.89**	**3.26**
QM-FF	0.1848	3.5532	0.01	0.9	0.2060	3.4400	0.0166	2.1250	-2.72	-2.42	2.87	3.25

The initial position of the water molecule on the surface of each PAH molecule was determined by a molecular mechanics, energy minimization calculation, using the UFF[24] forcefield, and an energy and force convergence criterion of 5 x 10^{-6} kcal/mol. Full geometry optimization was performed for all the water/PAH systems, with the exception of tri-circumcoronene, due to the system size. For this system, the water position was obtained from the optimized di-circumcoronene system and single-point energy calculations performed.

We evaluate the atom-centered point charges of our system using the electronic potential and Mulliken[25] population models. From these we calculated the coulombic component of the binding energy, and the corresponding true van der waals energies. The results of these calculations are presented in **Figure 3** and Table 1.

c) The QMFF-CW forcefield

We use the cp binding energies and distances of the water molecules on the PAH surface to obtain the C-O and C-H LJ parameters for van der Waals potentials (QMFF-CW)

$$E = D_v[\rho^{-12} - 2\rho^{-6}] \quad (1)$$

where

$$\rho = \frac{R}{R_v} \quad (2)$$

is the scaled distance, R_v is the well radius (distance between the atoms at equilibrium). These parameters are optimized by a scheme that minimizes the difference between the QM and FF energies of water molecules on an infinite graphene sheet, using the Powell conjugate gradient method[26] and molecular mechanics energies obtained from the LAMMPS[27] MD engine. The parameters for QMFF-CW are shown in Table 2.

d) MD Simulation Details

The contact angle of water on graphene was performed using. We represent the C-C interactions from our previously GraFF-LJ[28] forcefield, which has been shown to reproduce the specific heat of C60 and C70 crystals. The forcefield has van der Waals parameters of D_c = 0.0692 kcal/mol, R_c = 3.805Å, and many-body valence terms aimed at reproducing the phonon states of bulk graphite.

We chose the flexible, 3 site F3C water model of Levitt and coworkers[29] which reproduces many of the thermodynamic and bulk properties of water, most important of which are the radial distribution function, heat of vaporization and cohesive energy at STP. As published, the force constants for the OH bonds (500 kcal/mol) and angles (120 kcal/mol) do not reproduce the vibrational spectrum of water. We believe these values were used to facilitate an MD integration timestep of 2 fs. In all our calculations, we instead use the following force constants that do reproduce the vibrational spectrum,

and use a MD integration timestep of 1fs: bond stretch = 1105 kcal/mol, angle bend = 100.5 kcal/mol.

A full list of the parameters used in our calculations is presented in Table 2. Here, we minimize the energy of a single water molecule over a 2-D infinite graphene sheet, with cell dimensions of 74.65Åx73.38Åx106.55Å (1972 carbon atoms). The system is finite in the z dimension, to remove any artifacts of interactions across cells in this dimension. The calculations were performed using the LAMMPS[27] MD engine and the conjugate gradient minimization method. The van der Waals interactions were evaluated exactly using Ewald summations [30] with a convergence tolerance of 0.001 and a real space cutoff of 10Å. For this system, this gave a Ewald factor of 0.2856. The long range electrostatics were evaluated with the Particle-Particle Particle Mesh Ewald (PPPM) method[31] with an interpolation order of 4 and an tolerance of 10^{-4}.

Also in table 2, we compare QMFF-CW to several water models: SPC-EW[32], TIP3/TIP4[33] and F3C[29]. For these calculations, we obtained C-O and C-H interaction terms using geometric combination rules: $R_{AB} = \sqrt{R_{AA}R_{BB}}, D_{AB} = \sqrt{D_{AA}D_{BB}}$. We also include results from Bojan and Steele[34] (BS), Gordillo and Marti[35] (GM), Werder and co-workers[5] (WW) and Pertsin and Grunze[3] (PG), all of which have explicit C-O and H-C interaction terms.

e) **Contact Angle Measurements and Thermodynamics analysis**

We measure the contact angle of a rectangular water drop (14.5x14.0x3.1 $Å^3$) of 60000 water molecules on an infinite 2D periodic graphene layer (22.2x21.3x100 $Å^3$, 19440 carbon atoms) as shown in **Figure 4**. We use a procedure similar to that of

Werder and coworkers[5], where we assume an ellipsoid and calculate the mass weighted center of the water molecules (Z_c) on the surface of the graphite by the center of the axes of the principle moment of inertia. At discrete distances along the z-axis, the extreme radial position of the water molecules in that slab is calculated. The corresponding profile is then fit to a circle, of radius r_0 and center Z_0 displaced from Z_c. The contact angle is then taken as the tangent angle of the Z_c plane makes with the circle (**Figure 5**).

Young's modified equation related the microscopic θ and macroscopic θ_∞ contact angles[40]. For the microscopic system

$$\gamma_{SV} = \gamma_{SL} + \gamma_{LV} cos\theta + \frac{\tau}{r_0} \qquad (3)$$

where

γ_{SV}, γ_{SL} and γ_{LV} are the solid-vapor, solid-liquid and liquid-vapor surface tensions respectively

τ is the line tension

θ is the contact angle and

r_0 is the radius of the droplet defined previously

The macroscopic relation is obtained as $r_0 \rightarrow \infty$, so that

$$cos\theta = cos\theta_\infty - \frac{1}{\gamma_{LV}} \frac{\tau}{r_0} \qquad (4)$$

All MD simulations were performed using the LAMMPS MD engine, with an equilibration procedure similar to our previous simulations[36-38]. We calculated the contact angle at 300K and 350K, with the graphene unit "frozen" (position restrained by a 500 kcal/mol harmonic spring) and flexible, for a total of 4 simulations. We calculate the entropy and zero-point energy corrections to the enthalpy using the 2PT method of Lin, Blanco and Goddard[39]. The results of these calculations are presented in Table 3 and Figures xx and xx.

III. Results and Discussion

a) Comparison of alpha and beta binding modes

The parameters for our carbon-water force field are listed in table 2. We obtain a binding energy of 2.42 kcal/mol for the alpha orientation and 2.12 kcal/mol for the beta orientation (Table 1). The measured coulombic interactions in the tri-circumcoronene system was negligible (-0.13 and -0.02 kcal/mol respectively) compared to benzene (-1.93 kcal/mol and 3.00 kcal/mol respectively). We also see a systematic decrease in the magnitude of this coulombic interaction as the PAH size is increased. This result is particularly comforting since it is assumed that the water molecules at the surface will not affect any appreciable polarization on the underlying carbon atoms. It also validates our model of using increasingly large PAH molecules and extrapolating to the infinite case to obtain the London dispersion interaction energy.

Interestingly, the beta binding mode is unstable for benzene (+0.64 kcal/mol), an effect that is entirely accounted for by the large, positive coulombic interactions (+3.00 kcal/mol). We find the oxygen to be located 3.02Å over in the center of the ring and a ESP charges of +/-0.17 e^- for the hydrogens and carbons respectively. The oxygen has a charge of -0.50 e^-, giving a net unfavorable binding energy, since the oxygen-carbon distances are shorter than the oxygen-hydrogen distances.

The situation is reversed in the alpha case, where the positively charged hydrogens have a net favorable coulombic energy, due to the closer, negatively charged

carbons. Indeed, the van der Waals interactions increase with the increasing system size in both cases, as expected.

b) QMFF-cw forcefield parameters

IV. Conclusions

V. Figures

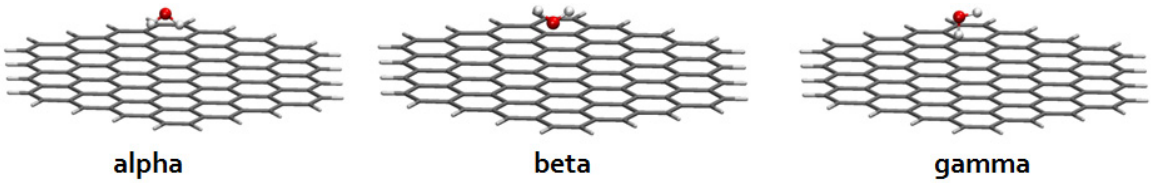


Figure 0

Figure 1 Binding modes of water on Circumcoronene. In this case, we find the alpha mode to be the most stable, with a counterpoise corrected binding energy of 3.20 kcal/mol. The beta mode is the least stable with a cp binding energy of 1.6 kcal/mol.

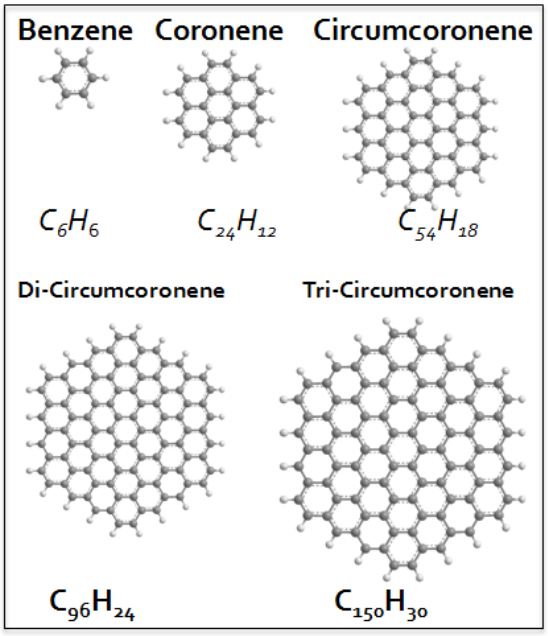


Figure 2 Five successively large, symmetric PAH molecules used in this study

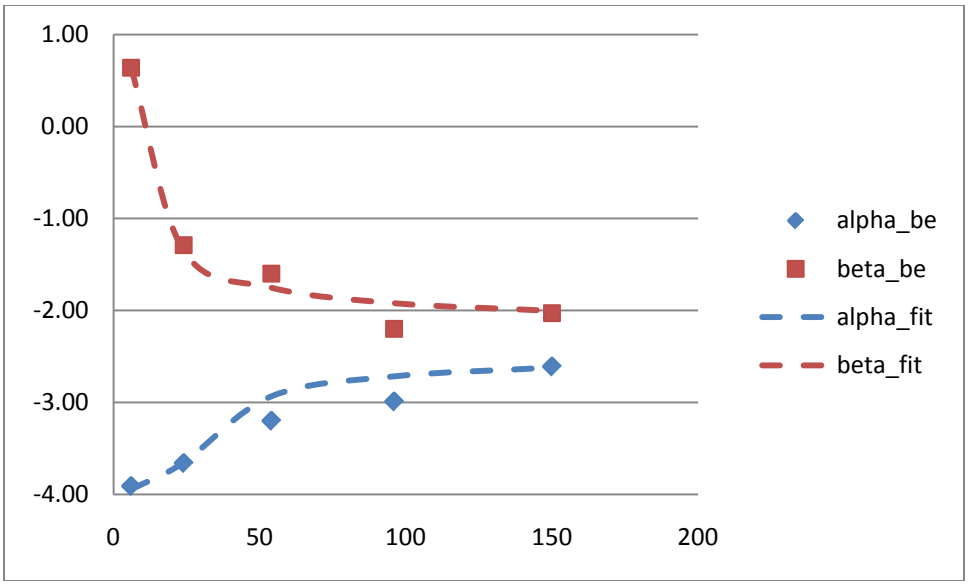


Figure 3 Ordinate: Binding energy (kcal/mol) of water molecules (alpha and beta modes). Abscissa: #C atoms of 5 PAH molecules in this study. By fitting the energies to decaying exponentials and extrapolating to infinity, we obtain the finding energy of water on graphite: 2.46 kcal/mol and 2.16 kcal/mol for the alpha and beta binding modes respectively.

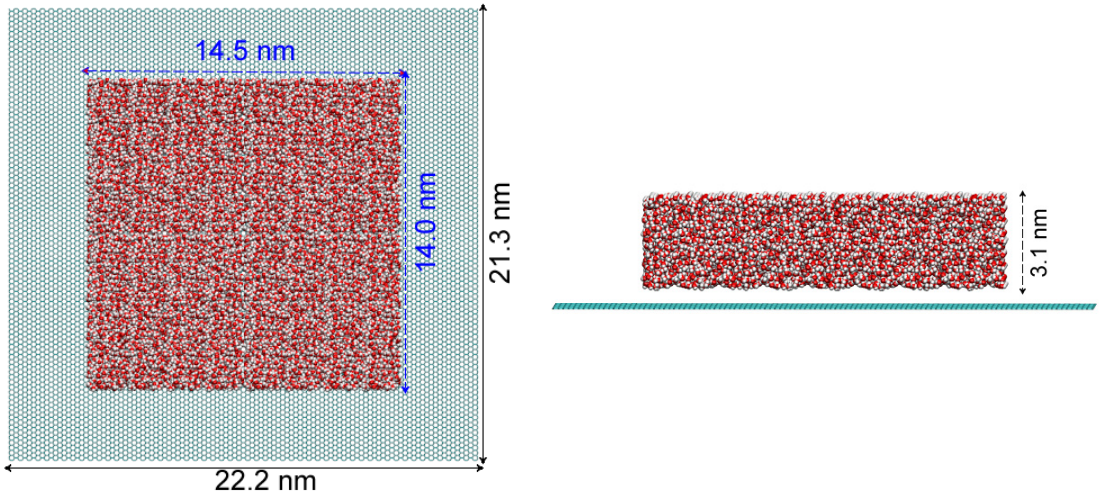


Figure 4 Starting structure for contact angle measurement. (a) Top view (b) Side view

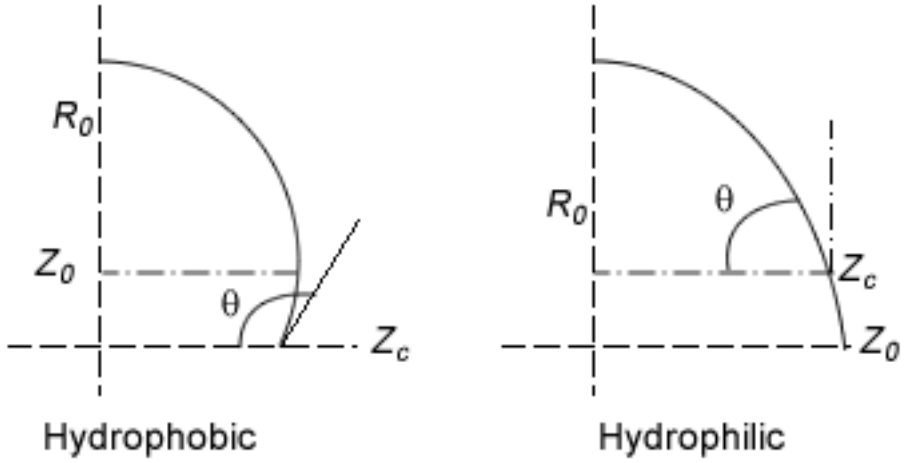


Figure 5 Schematic of procedure to measure contact angle. The center of the moment of intertia (Z_0) is obtained and the angle θ the fitted circle makes with the graphene surface (Z_c) is obtained. As shown, hydrophobic surfaces have θ > 90^o, while hydrophilic surfaces have θ < 90^o.

	Rigid Graphene		Flexible Graphene	
	Top	Side	Top	Side
300K				
315K				

V. References

(1) Dujardin, E.; Ebbesen, T. W.; Krishnan, A.; Treacy, M. M. J. *Adv. Mater.* **1998**, *10*, 1472.

(2) Feller, D.; Jordan, K. D. *Journal of Physical Chemistry A* **2000**, *104*, 9971.

(3) Pertsin, A.; Grunze, M. *J. Phys. Chem. B* **2004**, *108*, 1357.

(4) Birkett, G. R.; Do, D. D. *Journal of Physical Chemistry C* **2007**, *111*, 5735.

(5) Werder, T.; Walther, J. H.; Jaffe, R. L.; Halicioglu, T.; Koumoutsakos, P. *J. Phys. Chem. B* **2003**, *107*, 1345.

(6) Cicero, G.; Grossman, J. C.; Schwegler, E.; Gygi, F.; Galli, G. *Journal of the American Chemical Society* **2008**, *130*, 1871.

(7) Gordillo, M. C.; Nagy, G.; Marti, J. *J. Chem. Phys.* **2005**, *123*.

(8) Marti, J.; Nagy, G.; Guardia, E.; Gordillo, M. C. *J. Phys. Chem. B* **2006**, *110*, 23987.

(9) Marti, J.; Nagy, G.; Gordillo, M. C.; Guardia, E. *J. Chem. Phys.* **2006**, *124*.

(10) Fowkes, F. M.; Harkins, W. D. *Journal of the American Chemical Society* **1940**, *62*, 3377.

(11) Morcos, I. *J. Chem. Phys.* **1972**, *57*, 1801.

(12) Tadros, M. E.; Hu, P.; Adamson, A. W. *Journal of Colloid and Interface Science* **1974**, *49*, 184.

(13) Fetzer, J. C. *Polycyclic Aromatic Compounds* **2007**, *27*, 143.

(14) Møller, C.; Plesset, M. S. *Physical Review* **1934**, *46*, 618.

(15) Sudiarta, I. W.; Geldart, D. J. W. *Journal of Physical Chemistry A* **2006**, *110*, 10501.

(16) Zhao, Y.; Truhlar, D. *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)* **2008**, *120*, 215.

(17) Hohenberg, P.; Kohn, W. *Physical Review B* **1964**, *136*, B864.

(18) Kohn, W.; Sham, L. J. *Physical Review* **1965**, *140*, 1133.

(19) Zhao, Y.; Truhlar, D. G. *Journal of Physical Chemistry C* **2008**, *112*, 4061.

(20) Schrödinger: New York, NY, 2007.

(21) Hehre, W. J. R., L.; Schleyer, P. v. R.; Pople, J. A.; *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986.

(22) Boys, S. F.; Bernardi, F. *Molecular Physics: An International Journal at the Interface Between Chemistry and Physics* **1970**, *19*, 553

(23) Schwenke, D. W.; Truhlar, D. G. *The Journal of Chemical Physics* **1985**, *82*, 2418.

(24) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. *Journal of the American Chemical Society* **1992**, *114*, 10024.

(25) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833.

(26) Powell, M. J. D. *Mathematical Programming* **1977**, *12*, 241.

(27) Plimpton, S. *J. Comput. Phys.* **1995**, *117*, 1.

(28) Guo, Y. J.; Karasawa, N.; Goddard, W. A. *Nature* **1991**, *351*, 464.

(29) Levitt, M.; Hirshberg, M.; Sharon, R.; Laidig, K. E.; Daggett, V. *J. Phys. Chem. B* **1997**, *101*, 5051.

(30) Karasawa, N.; Goddard, W. A. *J. Phys. Chem.* **1989**, *93*, 7320.

(31) Plimpton, S. J.; Pollock, R.; Stevens, M. In *Proc of the Eighth SIAM Conference on Parallel Processing for Scientific Computing* Minneapolis, MN 1997.

(32) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269.

(33) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

(34) Bojan, M. J.; Steele, W. A. *Langmuir* **1987**, *3*, 1123.

(35) Gordillo, M. C.; Marti, J. *Chem. Phys. Lett.* **2000**, *329*, 341.

(36) Maiti, P. K.; Pascal, T. A.; Vaidehi, N.; Goddard, W. A. *NUCLEIC ACIDS RES* **2004**, *32*, 6047.

(37) Maiti, P. K.; Pascal, T. A.; Vaidehi, N.; Heo, J.; Goddard, W. A. *Biophys. J.* **2006**, *90*, 1463.

(38) Maiti, P. K. P., Tod A.; Vaidehi, Nagarajan; Goddard, William A. *J. Nanosci. Nanotechnol.* **2007**, *7*, 1712.

(39) Lin, S. T.; Blanco, M.; Goddard, W. A. *J. Chem. Phys.* **2003**, *119*, 11792.

(40) Wang, J. Y.; Betelu, S.; Law, B. M. *Phys. Rev. E* **2001**, *63*.

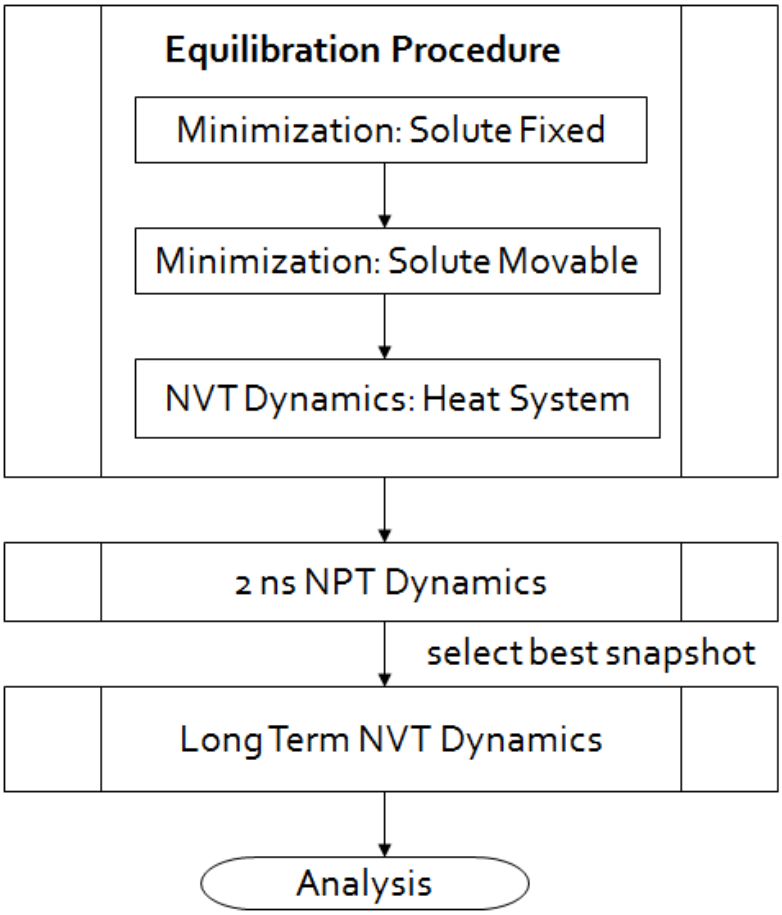
Chapter 6. Properties of water molecules at interfaces

[This manuscript for this paper is being prepared with William A. Goddard III]

APPENDIX I. THEORETICAL FRAMEWORK

A. General Simulation scheme

Periodic MD simulations require charge neutrality. Typically, this is achieved by adding salt to the system. It is our belief that a signigicant amount of time needs to be devoted to equilibrating the system, due to the motion of the ions. We've developed a general scheme for rapid equilibration of condensed phase systems. This scheme is based first on careful placement of the ions, according to the calculated electrostatic potential, followed by gradual heating of the system from 0K to the required temperature with a Langevin thermostat (see figure below). We find equilibration times of typically 2 ns for the systems under consideration, approximately an order of magnitude shorter than the standard approach of randomely placing the ions.



B. 2PT-FE-MD method for accurate thermodynamics

We use the 2PT Free Energy method,[20] to obtain the quantum corrections to the entropy, the zero-point energy, and specific heat to obtain the Helmholtz free energy (A),

$$H = E_0 - TS \qquad (1)$$

where E_0 is the quantum corrected enthalpy. The 2PT approach makes it practical to calculate the free energy of large, complex molecular systems, requiring only a 20ps dynamics trajectory.

From this 20 ps trajectory we calculate the velocity autocorrelation function (VAC) for each atom,

$$C(t) = \sum_{j=1}^{N} \sum_{k=1}^{3} m_j \left[\lim_{t\to\infty} \frac{1}{2\tau} \int_{-\tau}^{\tau} v_j^k(t' + t) v_j^k(t') dt' \right] \qquad (2)$$

where $v_j^k(t)$ is the *k-th* component of the velocity of atom *j* at time *t*. We then take the fast Fourier Transform (FFT) to obtain the power spectrum (vibrational density of states) (figure S5 of supplementary materials). We would like to apply quantum statistical mechanics to this function to obtain the partition function from which all thermodynamic quantities can be calculated:

$$\ln Q = \int_0^\infty dv \hat{C}(v)\, q_{HO}(v) \qquad (3)$$

where:

$\hat{C}(v)$ is the density of normal modes at frequency *v* and

$q_{ho}(v)$ is the quantum harmonic oscillator partition function: $q_{HO}^{Q}(v) = \frac{e^{-\beta h v/2}}{1-e^{-\beta h v/2}}$

The problem with this procedure is that $\hat{C}(v)$ is finite at ν =0, which leads an infinite contribution to entropy. The peak in $\hat{C}(v)$ at ν =0 is due to diffusional motion (obvious from inset in figure S6 of supplementary materials). Lin, Blanco, and Goddard (LBG) showed that this problem is to extract from the MD a model for the diffusion whose entropy could be calculated

accurately. LBG did this by partitioning the DoS into a component related to a "gas-like" diffusion term $\hat{C}(v)_{diff}$ (modeled as a hard sphere) and a "solid-like" vibrational term $\hat{C}(v)_{vib}$ (modeled as a vibrating crystal):

$$\hat{C}(v) = f\hat{C}(v)_{diff} + (1-f)\hat{C}(v)_{vib} \quad (4)$$

where *f* is the fraction of the 3N total modes corresponding to the fluid or diffusional parts of the dynamic system, such that at *v = 0,* $\hat{C}(v)_{vib}$ goes smoothly to 0 and

$$\hat{C}(0) = \hat{C}(v)_{diff}(0) = \frac{12mND}{kt} \quad (5)$$

where:

N is the total degrees of freedom of the system
m is the total mass and
D is the system-specific, self-diffusion constant obtained from the MD

LBG validated[1-2] that the 2PT model Lennard-Jones systems give free energies in excellent agreement with accurate Monte Carlo results for the entire phase diagram, including metastable regions[1]. In addition the 2PT model has been successful in calculating the entropy of water in different domains of PAMAM dendrimers[3], in determining various phases of dendrimer liquid crystals[4], and in calculating the relative stability of various aggregates[5]. Recently, Jana et al.[6] used 2PT to show that water molecules in both grooves of DNA, have significantly lower entropies than for bulk water.

A powerful feature of the 2PT method is that it is a sum over atoms, so the contributions can be grouped together to calculate the thermodynamic properties of different molecules independent of the rest of the system. This facilitates straightforward analysis of the contributions from different parts of the system. We utilize this feature to study the differential thermodynamics of water molecules in the two solvation shells surrounding the DNA, and in bulk.

i) DoS Decomposition

For molecular systems, we decompose the contribution to the velocity of each molecule (V_{tot}) into its translational, rotational and vibrational components. This allows us to determine the contribution of each atom or molecule to the thermodynamic properties calculated by the 2PT methods. The decomposition of the total velocity for each molecule is obtained as follows:

1. The COM translational contribution to the total velocity (V_{trans}) (for molecule *i* and total mass M_i)is obtained as the center of mass velocity of that molecule. For each of the *j* atoms in that molecule:

$$V_{trans}\,(i) = \frac{1}{M_i}\sum_j m_j V_j \qquad (6)$$

2. The rotational contribution (V_{rot}) is obtained by calculating the angular velocity (V_{ang}), treating the system as a classical rigid rotor:

$$V_{rot}\,(i) = V_{ang}\,(i) \times V_{tot}\,(i) \qquad (7a)$$

$$V_{ang}\,(i) = \mathrm{I}_i^{-1} \times V_{ang\_momentum}\,(i) \qquad (7b)$$

where: I_i^{-1} is the inverse of the moment of inertial tensor for molecule i,

$\times$ represents the vector cross product

and $V_{ang\_momentum}(i)$ is the angular momentum:

$$V_{ang\_momentum}\,(i) = \sum_j m_j\,(R_j \times V_j) \qquad (7c)$$

(here R_j is the position of atom j in molecule i)

3. Finally, the vibrational component (V_{vib}) to the velocity is taken as the remaining velocity after subtracting the first two contributions:

$$V_{vib}(i) = V_{tot}(i) - (V_{trans}(i) + V_{rot}(i)) \quad (8)$$

In these calculations the water molecules are kept rigid (using SHAKE constraints on all OH bonds and HOH angles). Thus the vibrational contributions for the water molecules are all zero.

ii) Enthalpy with Zero-Point Energy corrections

We partition the total potential energy into a per-atom enthalpy (including both non-bond and valence contributions). In this analysis the solute atoms were "movable" while the water and ions were "fixed". For interactions involving only movable atoms half of each pairwise energy was associated with each atom, while three-body angle terms and four-body inversion terms were associated with the central atom, and four body torsion terms were associated half with each central atom. For movable-fixed interactions, the full interactions were associated with the movable atoms.

The internal Energy (*E*) is:

$$E = V_0 + T\beta^{-1}\left(\frac{\partial lnQ}{\partial T}\right)_{N,V} = V_0 + \beta^{-1}\int_0^{\infty} dv\hat{C}(v)W_E(v) \quad (9a)$$

where V_0 is the reference energy of the system: the potential energy of the system for which all the vibrations are in their lowest (zero) vibrational level. Here:

$$W_E^Q(v) = \frac{\beta hv}{2} + \frac{\beta hv}{e^{\beta hv} - 1} \quad (9b)$$

We write $\int W_E^Q(v) = W_E^{diff}(v) + W_E^{HS}(v)$,and rewrite 10b as:

$$V_0 = E^{MD} - \beta^{-1}\int_0^{\infty} dv\hat{C}(v)W_E^Q(v) = E^{MD} - \frac{3N}{\beta}(1 - 0.5f) \quad (9c)$$

where V_0 includes all zero point energy effects and is taken as the reference enthalpy for our free energy calculations. The components of the enthalpy per water molecule in the various hydration shells are presented in table 8.

iii) Entropy

For molecular systems, with a distribution of vibrational modes $\hat{C}(v)$, the entropy (*S*) and Helmholtz free energy (*A*) can be written as a function of the partition function in eqn. 6:

$$S = k\,lnQ + \beta^{-1}\left(\frac{\partial lnQ}{\partial T}\right)_{N,V} = k\int_0^\infty dv\hat{C}(v)W_S(v) \tag{10a}$$

$$A = V_0 - \beta^{-1}lnQ = V_0 + \beta^{-1}\int_0^\infty dv\hat{C}(v)W_A(v) \tag{10b}$$

with quantum weighting functions

$$W_S^Q(v) = \frac{\beta hv}{e^{\beta hv}-1} - ln\left[1-e^{\beta hv}\right] \tag{11a}$$

$$W_A^Q(v) = ln\frac{1-e^{\beta hv}}{e^{\beta hv/2}} \tag{11b}$$

IV. References

(1) Lin, S. T.; Jang, S. S.; Cagin, T.; Goddard, W. A. *J. Phys. Chem. B* **2004**, *108*, 10041.

(2) Lin, S. T.; Blanco, M.; Goddard, W. A. *J. Chem. Phys.* **2003**, *119*, 11792.

(3) Lin, S. T.; Maiti, P. K.; Goddard, W. A. *J. Phys. Chem. B* **2005**, *109*, 8663.

(4) Li, Y. Y.; Lin, S. T.; Goddard, W. A. *Journal of the American Chemical Society* **2004**, *126*, 1872.

(5) Jang, S. S.; Lin, S. T.; Maiti, P. K.; Blanco, M.; Goddard, W. A.; Shuler, P.; Tang, Y. C. *J. Phys. Chem. B* **2004**, *108*, 12130.

(6) Jana, B.; Pal, S.; Maiti, P. K.; Lin, S. T.; Hynes, J. T.; Bagchi, B. *J. Phys. Chem. B* **2006**, *110*, 19611.

C. Obtaining Interaction Free Energy of DNA from First Principles

The stability of the DNA structures has been actively investigated since Watson and Crick[1-2] proposed the double helix as the molecular structure: where A is hydrogen bonded with T and G is hydrogen bonded with C. A number of qualitative factors have been identified as contributing to this thermodynamic stability[3-5]: base stacking interactions (van da Waals), Watson-Crick hydrogen bond interactions and interactions with water molecules and metal ions (electrostatics). Different levels of water and ion concentration lead to the three distinct forms of DNA found in nature: A-DNA, B-DNA and Z-DNA. B-DNA is the predominant form, commonly occurring in well solvated system and the presence of sodium and magnesium ions at relatively low concentrations (0.1 M).

Quantitative measures of DNA duplex stabilities have been attempted in order to explain everything from binding in protein-DNA[6-7] complexes to DNA melting[8-9]. Future applications such as DNA computers, DNA based nano-sensing arrays, and DNA as a component in molecular electronics demand a fundamental and quantitative understanding of DNA structure and the factors affecting it. To this end, a number of experiments have been performed, the most famous results of which was pioneered by the SantaLucia group[4,9-12]. Using NMR spectroscopy, they determined the sequence and solution dependence of the DNA double helix; the data was show to closely fit the Nearest-Neighbor (NN) interaction model. This model assumes that the stability of a DNA duplex can be depends on the identity and orientation of neighboring base pairs. This approximation therefore reduces the free energy to a base-pair to its interaction with the neighboring bases. Ten different nearest-neighbor interactions are therefore possible in any Watson-Crick DNA duplex structure.

Using the same molecular dynamics procedure as the 3-way junction, 13 atomistic B-DNA double helices are simulated. These helices were of random sequence and variable length. The 10 NN

parameters was determined by single value decomposition (SVD) of the per-atom enthalpies of the entire set. The energetic obtained is similar to those obtained experimentally, namely: GC > CG > GT ~ AA > GA > CA > GG > AT > TA > CT. Further, there is a >99% correlation between the predicted helical enthalpies from our parameters and that of SantaLucia. More importantly, our parameters allow us to give a priori estimations of the simulation enthalpy of a random sequence of B-DNA to a confidence level 0.5%.

i) Methods

Each of the 13 double helices (see table 1) were created using the Namot2 nucleic acid builder using standard B-DNA parameters (33.3^{o} backbone twist, 3.4Å helical rise). The same simulation procedure as before was followed, except the structures were simulated for 5ns of NPT dynamics. Snapshot of the structures from the last 1ns (every 10ps) were obtained and the energy per atom calculated using MPSim. This per atom energy is the interaction of that atom with every other in the system. This calculation involved holding the solvent molecules fixed while allowing the solute to be movable. This had the effect of removing the solvent-solvent interactions from the calculated energies. The electrostatics were calculated using a direct method: for each pair of atoms the electrostatic contribution to energy (E_{ij}) is: $E_{ij} = \frac{q_i q_j}{\epsilon r_{ij}} \quad r < r_e$, q_i, q_j is the charge on atom *i* and *j*, r_{ij} is the distance between the atoms and ϵ is the dielectric constant (set to be 1). The atoms' energies were grouped by nucleotide (the solvent and ions were discarded) and average over all the snapshots of that particular simuation.

The first chain of each dna double helix was iteratively traversed and two kinds of analysis performed:

Single Base: Base$_n$ Energy $E_j = InternalEnergy_i + NonBond_i$

Nearest Neighbor: Energy $E_j = InternalEnergy_i + \frac{1}{4}(NonBond_{i-1} + 2 * NonBond_i + NonBond_{i+1})$

5' End: Energy $E_j = InternalEnergy_{5'} + \frac{1}{4}NonBond_{i+1} + \frac{3}{4}NonBond_{5'}$

3' End: $E_j = InternalEnergy_{3'} + \frac{1}{4}NonBond_{i-1} + \frac{3}{4}NonBond_{3'}$

where InternalEnergy is the sum of the valence terms (bond, angles, torsions and inversions) and NonBond Energy is the sum of the van der Waals and Columbic terms. Note that the sum of individual energy term is the total solute energy.

A linear system $A^{n\times m}x^{n\times 1} = B^{m\times 1}$ is generated: where x is an n x 1 matrix of the parameters (a_i) in table 2. A is an n x m binary matrix where

$a_{ij} = \begin{cases} 1 \; if \; element \; a_i \; was \; used \; to \; calculate \; energy \; E_j \\ 0 \; otherwise \end{cases}$. Finally, B is a m x 1 matrix of energies E_j.

Here 1 < j < m and m = 412 (total number of base-pairs of all 13 double helices) and 1 < I < n and n = 22 (total number parameters – table 2). This linear system is over-determined, so Single Value Decomposition was used to obtained a least squares fitting of the parameters. All matrix operations were computed in Mathematica. The parameters obtained are listed in table 2.

# BP	RI	SEQUENCE
5	1.00	GATGG
10	0.67	CTTCTCGGGA
18	1.00	GACATGTTACAGTAAATT
20	0.95	ACCATGCAGATCACGTGTTA
25	0.92	GACGCGGGTATTCTATGGGTTGCAG
26	0.72	GTCATCGTAGCTACGCGTATCGACGT
30	0.86	TGTACTCAAGATCACGCGGAGCTGTGTTCA
35	0.85	CGTGTCGATAGCTTAGTGCAGGCCTCTTTTTTTGC
38	0.97	GATACCGGGCGGCAACAGAAGATTACCTCTAAGGTCGC
40	0.79	CCATATGAAAGTAAGCCGGTAGAATAGTAACATGAGGACA
50	0.98	GAGGATGATTGCCTCTAAGTCTTAACGCGATGCGCACGTTAGAAAGGTGG

55	0.80	ACCAACGGTAGGGGGGATAGCATGGATGGAATGGCGTTTGGGGCGTGGTTACATT
60	0.81	AGACCTCTCGCCGCGACTTGATCAATGAGGAGCGGCGTATCCTTGATTATCGGAAATC GA

Table 1. Sequence of the 13 B-DNA double helices

a) **Selection of double helix sequences**

Due to limited computer resources, only a finite number of DNA double helices can be simulated. Many simulations, of random sequences, would ensure unbiased weighting and further refine the obtained parameters. To this end, a "Randomness Index (RI)" was created to exclude sequences that contained repeated blocks. The sequences were chosen so that they had a RI of at least 0.66. In calculating the RI, every pair of bonded nucleotides were compared and uniform distribution was assumed. Hence:

$$\Pr[Doublet\ in\ sequence] = {}^{1}/_{16}$$

$$E[\#\ Doublet\ in\ sequence] = \#\ basepairs * \Pr[Doublet]\ (rounded\ up\ to\ nearest\ integer)$$

$$Penalty[Doublet] = \begin{cases} Count[Doublet] - E[Doublet] & if\ Count[Doublet] > E[Doublet] \\ 0 & otherwise \end{cases}$$

$$RI = 1 - \left(\sum_{n=1}^{16} Penalty[Doublet_n] \Big/ total\ basepairs \right)$$

Parameter (a_i)	ΔH (kcal/mol)	# Points	ΔH (kcal/mol)	# Points
	Nearest Neighbor		**Single Base**	
T	-101.260	103	-217.631	103
A	-101.051	103	-217.674	103
C	-99.692	83	-252.615	83
G	-98.522	123	-251.712	123
GC/CG	-76.677	52	-	
CG/GC	-76.317	62	-	
GG/CC	-76.157	90	-	
CT/GA	-67.534	88	-	
GT/CA	-65.737	98	-	
GA/CT	-65.623	108	-	
CA/GT	-65.404	94	-	
TA/AT	-60.179	54	-	
AA/TT	-57.951	90	-	
AT/TA	-57.760	62	-	
5'G	-28.056	6	69.852	6
3'G	-24.396	3	70.942	3
5'C	-23.929	3	71.679	3
3'C	-21.262	2	72.995	2

5'T	-10.672	1	69.541	1
5'A	-8.526	3	70.264	3
3'A	-7.492	5	71.412	5
3'T	-7.378	3	70.524	3

Table 2. Parameters obtained using the “Nearest Neighbor” Analysis

ii) Results

a) Parameters from SVD

Single Base Parameters

The matrix A, used in the SVD for the Single Base method consists of a 1 for the base being observed, with zeros everywhere else in most cases. This is only different at the ends, where an additional 1 would represent one of the 8 possible end conditions (5’A, 5’T, 5’C, 5’G, 3’A, 3’T, 3’C, 3’G). Hence, the solution of the linear system amounts to a least squares fitting of the parameters (table 1).

“Nearest Neighbor” Parameters

The matrix A for this method of analysis is more dense than that of the base-only analysis, comprising of 1’s for the current base (representing the internal energy) as well as a 1 for any of the 10 nearest neighbor pairs (representing the non-bond energy) on either side of the base.

For each of the nearest neighbor pairs (see table 2), the code xx/xx is used and corresponds to the Base-Base(+1) sequence for strand 1 coupled and the corresponding Base-Base(-1) sequence on strand 2.

b) Comparison of Calculated Nearest Neighbor Parameters to SantaLucia

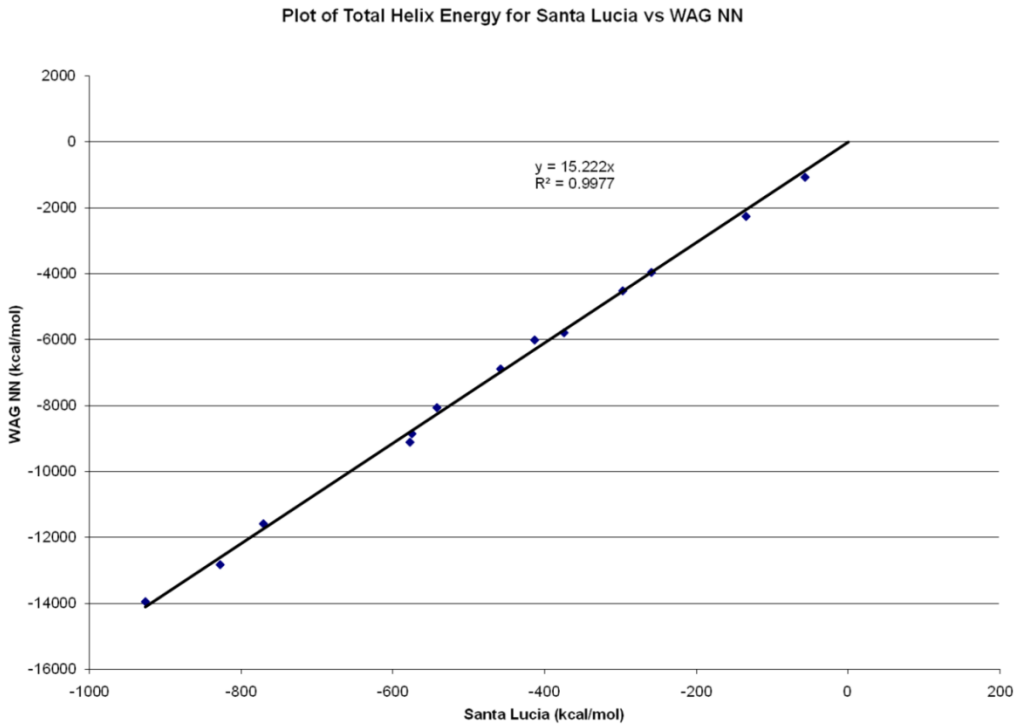


Figure 8 - Comparioson of SantaLucia and WAG NN parameters

Using thermal melting data of different DNA double helices, Santa Lucia et. al. were able to extract nearest neighbor parameters to predict duplex stabilities. Their data was obtained from melting of 26 oligonucleotides ranging in length from 4 to 16 base-pairs.

The observed trend from their study was:

GC > CG > GT ~ AA > GA > CA > GG > AT > TA > CT.

As seen in the table 2, the following trend is observed:

GC ~ CG ~ GG > CT > GT ~ GA ~ CA > TA > AA ~AT.

Using the Santa Lucia Numbers, the total predicted energy for the various double helices were computed and plotted vs. the Nearest Neighbor predicted enthalpy (figure 8).

c) Error Comparison between Nearest Neighbor and Base-Only

The parameters determined by the SVD analysis above were used to predict the total energy of the double helix system used in determining the parameters. This was done to provide some measure of the error resulting from the least squared fitting (figure 9).

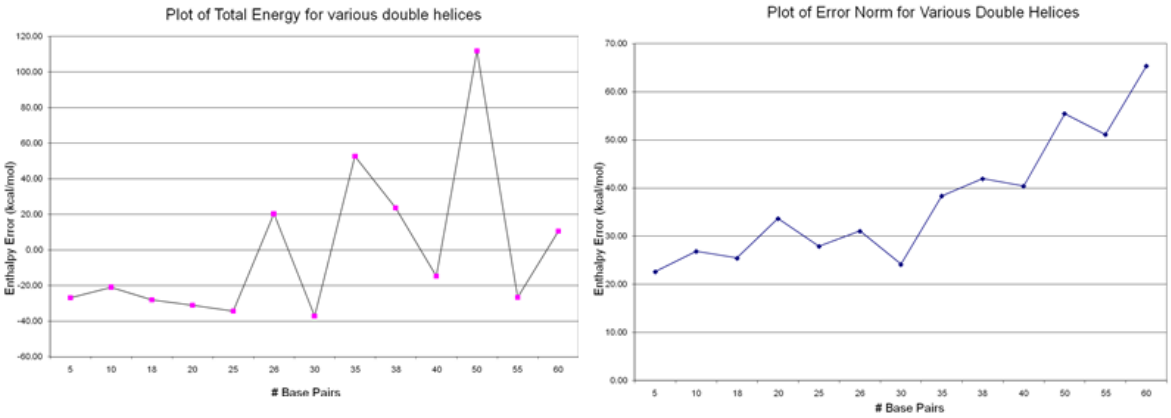


Figure 9 - Plots of Total Error (left) and Error Norm (right) for 13 DNA Helices

d) Prediction of Random double helix energy

The above analysis is a worthwhile measure of the accuracy of the parameters. The point of determining these parameters however was to accurately predict the enthalpy of a DNA double helix of unknown length and sequence. This is done for two case: a 45bp and a 15bp DNA double helix of random sequence. The error in calculating each base is shown. As shown in the following table, the errors in predicting the total energy of the double helices were less that 1%.

		Predicted E (kcal/mol)	Sim E (kcal/mol)	Diff (kcal/mol)	% Diff
40bp	**NN**	-10538.47	-10449.76	-88.71	0.85%
	Single Base	-10538.47	-10479.68	-58.79	0.56%
15bp	**NN**	-3359.78	-3378.23	18.45	0.55%
	Single Base	-3362.89	-3359.78	3.11	0.09%

Table 3 – Comparison of accuracy in predicted total energy of random helix

iii. Discussion

a) Accuracy of Simulation

The simulation procedure used above is the standard procedure observed while simulating DNA molecules in explicit water and counter-ions. The total DNA energy and temperature of each system was monitored in order to ensure convergence. On average, the fluctuations in the total energy of the DNA in each simulation was less that 0.05%, with thermal fluctuations of the order 300.0 K ± 3 K, well within the accepted limits of convergence.

b) Adequacy of Basis set

Using the randomness index, we were able to ensure that the parameters obtained were suitable to sample nearly every possible base-base stacking conformation. This is necessary in order to ensure that any nearest neighbor effects are properly quantified. As seen on table 3, each nearest neighbor pair is sufficiently well sampled, with over 50 data points used for the SVD least squares fitting. One potential source of error stems from the relative few points used to determine the end corrections. In one case, only 1 data point was used. An expanded basis set could therefore provide for even more refined parameters.

The total energy of each base-pair (or nearest neighbor pair) contains not only the energy of the pair, but also includes the energy of its interactions with the rest of the systems. This includes all the other atoms in the DNA, as well as the solvent and counter ion interactions. Hence, this method of determining the total energy of a particular base is akin to a mean field analysis. All interactions, including “solvation” effects, are inherently included in this analysis.

Finally, the total Error Norm (a measure of variance) is less that 0.5% of the total energy, which is well within the precision range of the original simulations.

iv) Conclusions

Comparison of the predicted energies from nearest neighbor analysis vs. the predicted energies from Santa Lucia gave an excellent correlation of > 99%. The differences in the observe trends is a function of differences in the conditions for the simulations and the experiment (different salt concentrations etc.). It is important to note that both sets of parameters are optimized to reproduce the total energy of the system. Hence the high correlation between the numbers ascertains that our simulation numbers are capable of describing the energetics of real systems, based on the proven success of the Santa Lucia parameters.

As seen in figure 9, there is almost a linear increase in the error norm as the length of the helix increases. This is indicative of some long range (electrostatic) effects not properly captured by either measure. One possible explanation is the natural bending of a DNA double helix as it increases in length. This is thought to be a mechanism of reliving strain in the helix. Hence, the longer helices would be more stable than predicted by this analysis. Further, the error made in predicting stability would increase as the helices get longer.

V References

(1) Watson, J. D.; Crick, F. H. C. *Cold Spring Harbor Symposia on Quantitative Biology* **1953**, *18*, 123.

(2) Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 737.

(3) Antao, V. P.; Lai, S. Y.; Tinoco, I. *Nucleic Acids Res.* **1991**, *19*, 5901.

(4) SantaLucia, J.; Allawi, H. T.; Seneviratne, A. *Biochemistry* **1996**, *35*, 3555.

(5) Sugimoto, N.; Nakano, S.; Katoh, M.; Matsumura, A.; Nakamuta, H.; Ohmichi, T.; Yoneyama, M.; Sasaki, M. *Biochemistry* **1995**, *34*, 11211.

(6) Blier, P. R.; Griffith, A. J.; Craft, J.; Hardin, J. A. *Journal of Biological Chemistry* **1993**, *268*, 7594.

(7) Thompson, K. S.; Vinson, C. R.; Freire, E. *Biochemistry* **1993**, *32*, 5491.

(8) Dauxois, T.; Peyrard, M.; Bishop, A. R. *Phys. Rev. E* **1993**, *47*, 684.

(9) SantaLucia, J. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 1460.

(10) Allawi, H. T.; SantaLucia, J. *Biochemistry* **1998**, *37*, 9435.

(11) Peyret, N.; Seneviratne, P. A.; Allawi, H. T.; SantaLucia, J. *Biochemistry* **1999**, *38*, 3468.

(12) Xia, T. B.; SantaLucia, J.; Burkard, M. E.; Kierzek, R.; Schroeder, S. J.; Jiao, X. Q.; Cox, C.; Turner, D. H. *Biochemistry* **1998**, *37*, 14719.

APPENDIX II. MESO-SCALE DNA DYNAMICS

Atomistic level simulations, while powerful and generally accurate, are fundamentally limited in their usefulness because of their very nature: the atomistic representation of the bio-molecules. Even on the most powerful of computers, current equilibrium simulations of a bio-polymer such as DNA are limited to the nano-second regime of total simulation time, a timescale that is far too small to observe many of the interesting structural transitions that occur.

Efficient codes, employing parallel computing and fast methods of computing the non-bonded interactions present in these molecules are to be credited for the increase in simulation time recently. However, in order for realistic simulations to be performed in a reasonable amount of time, there is a need to have an alternate description of these molecules, a description that is preferably based on atomistic simulations, just as most force fields for atomistic simulations are based on quantum mechanical information.

Solvent molecules can be included in molecular dynamics simulations for short DNA duplexes, but at a high computational cost. There is therefore a need to not only to simplify the representation of the solvent molecules, but also decrease the number of degrees of freedom in the model for DNA itself. This simplified model of DNA (meso-scale model) will inherently not be as detailed as the atomistic model. However, the reduction in the degrees of freedom will permit the monitoring of the simulation dynamics on a much larger timescale.

i) **Description of Meso-scale Model**

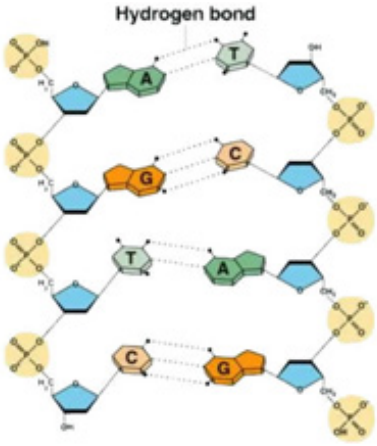


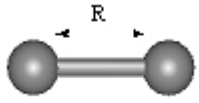
Figure 16: Meso-DNA schematic

Our meso-scale model of DNA is based on a backbone-base structure, where each nucleotide is represented by three pseudo-atoms, with 2 or 3 pseudo-beads. For each nucleotide, one pseudo-atom represents the phosphate group and another represents the sugar group of the sugar-phosphate backbone. There is then a pseudo-atom representing each of the different nucleosides (Adenine, Cytosine, Guanine and Thymine). Additionally, Guanine and Cytosine have 3 mass- less pseudo-beads, representing the 3 hydrogen bonding sites. Similarly, Thymine and Adenine have 2 pseudo-beads (see figure 16). These pseudo-beads move as rigid bodies with their parent pseudo-atom. This removes the need to define bond and angle terms for them.

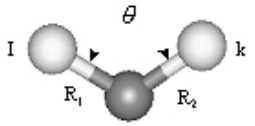
Statistics of the potential of mean force (between sets of atoms representing each bead) is computed from various full atomistic DNA simulations. These distributions are then fitted to standard potentials in order to obtain the interaction parameters for every bead (see figure 2). The resulting force-field uniquely describes the interaction of each bead with every other bead in system, who total energy is given by:

$E_{Total} = E_{Bonds} + E_{Angles} + E_{Torsions} + E_{non\,Bond} + E_{hBond}$, *where*

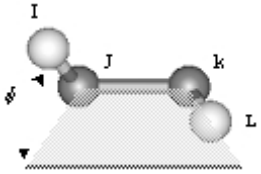
- $E_{Bonds} = \frac{1}{2}\kappa(r - r^{eq})^2$ (Harmonic)



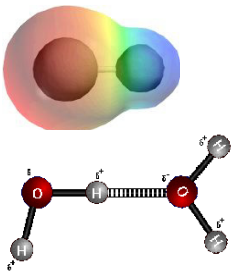
- $E_{Angles} = \frac{1}{2}\beta(cos[\theta] - cos[\theta^{eq}])^2$ (Harmonic)



- $E_{Torsions} = \frac{1}{2}\beta(1 + cos[n\phi - \phi^{eq}])$ (Harmonic)



- $E_{nonBond} = D_0 \left[e^{-\alpha\left(-1+{}^{r_{ij}}/_{r_0}\right)} - 2\left(e^{-0.5\alpha\left(-1+{}^{r_{ij}}/_{r_0}\right)} \right) \right]$ (Morse)



- $E_{hBond} = D_0 \left[5\left(\frac{r_0}{r}\right)^{12} - 6\left(\frac{r_0}{r}\right)^{6} \right] cos^4\theta$ (Dreiding)

ii) Parameter Development

The snapshot used when computing the nearest neighbor parameters are analyzed (1049 structures). For each snapshot, atoms comprising each pseudo-atom were determined. The center of mass of these atoms were then computed. Finally, the corresponding pseudo-atoms were placed at this center of mass.

a) Valence Parameters

Each snapshot was iteratively traversed and the 2 body bond lengths, 3 body angles and 4 body dihedrals calculated from the positions of the pseudo-atoms. These were tabulated and histograms of the distribution of these valence term, for each unique combination of pseudo-atoms, were generated (see figure 17). These distributions were shown to be a single-model Gaussian distribution for the bonds and angles, and multi-modal for dihedrals. The only exception occurred with considering the bond distribution between the Phosphate (PHO) and Sugar (SUG) pseudo-atoms which was bi-modal. This anomaly is explained by the natural difference by considering that starting at the 5'end, the SUG –> PHO bond length (4.4 Å) is longer than the next PHO –> SUG bond length (4.0 Å). When constructing the meso-model from the atomistic representation, if the backbone atoms are alternatively labeled as PHO-SUG-PHP-SUS-PHO, this anomaly is effectively resolved and the PHO-SUG, PHO-SUS, PHP-SUG and PHP-SUS distributions are uni-modal.

The parameters were obtained by performing a least squares fitting of a Gaussian curve of the desired functional form. The fitting function is an exponential with exponent equal to the relevant

potential/2RT. For the bonds, the fitting function was therefore $y = a_0\left(e^{-\frac{\frac{1}{2}\kappa(r-r^{eq})^2}{2RT}}\right)$. To start, an initial guess was given for each of the parameters was given, with new parameters obtained by observing the quality of the fit as the trial parameters were given small perturbations. This process was continued until the quality of the fit feel within a certain tolerance range, or the number of iterations were exhausted. Table 5 gives the parameters obtained for the various bond-stretch terms.

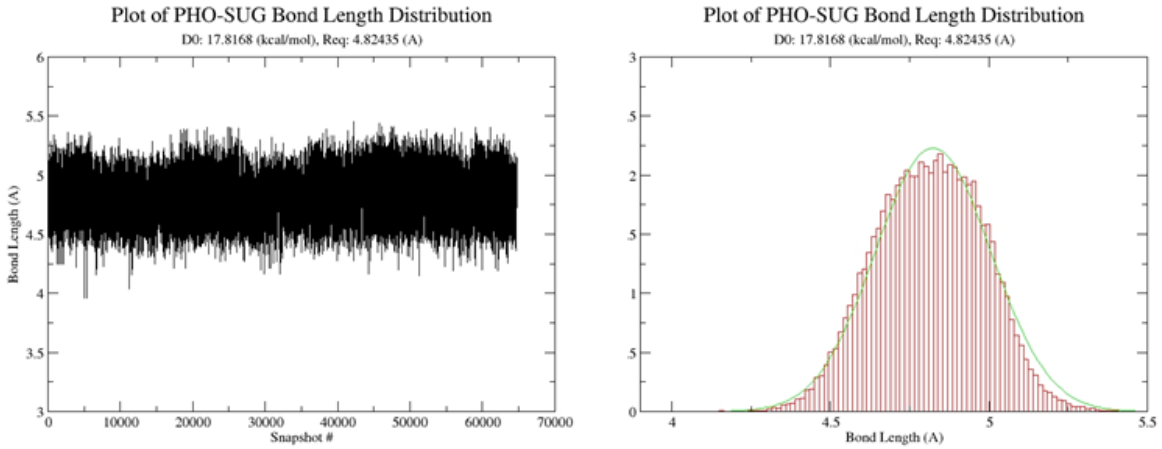


Figure 17: Plot of PHO-SUG Bond Distance per Snapshot (left) and Histogram (right)

	k_0 ($kcal\ mol^{-1}Å^{-2}$)	r_0 ($Å$)
SUG-THY	167.848	4.2136
SUG-ADE	155.974	4.6175
SUG-GUA	143.419	4.7412
SUG-CYT	193.31	4.1101
SUG-PHO	46.0059	4.0053
PHP-SUG	18.2605	4.4283

Table 5. Bond Parameters for pseudo-beads

b) Non bond Parameters

Two copies of the atoms comprising each pseudo-atom were obtained, displaced by 4 Å in the z-direction and individually immersed in a pre-equilibrated box of TIP3 water. The charges were maintained, hence the only set of atoms with a net charge were the 5 atoms comprising the phosphate (PHO/PHP) bead, with a net -1 charge per set of atoms. If necessary the system was neutralized with Na+ ions and subjected to the same simulation protocol as before. 100 ps of NPT dynamics was performed, with snapshots of the last 20 ps (every 2 ps)saved. The 10 snapshots were then superimposed on each other and the average MD solution structure obtained as previously discussed.

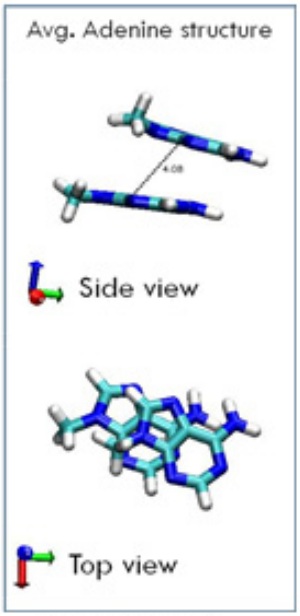


Figure 18: Ade Dimer

A constrained MD simulation was then ran, where the center of mass of each of the 2 sets of atoms are constrained by a harmonic potential. Every time step, the distance between the two center of masses is checked, and if they are not equal to a specified distance, a restoring spring of magnitude $-k\Delta x$ is applied to all of the atoms in the second set. Here, Δx is the differential distance from equilibrium and k is a force constant of 500 kcal mol^{-1} Å^{-1}. The center of mass of the two sets of atoms are constrained, starting at 2.0 Å, in 0.1 Å increments, until 10 Å. After each increment, 40 ps of NPT dynamics is performed to equilibrate the structure at the restraint, followed by a further 20 ps of NPT dynamics, during which time the non bond energy of both sets of atoms is calculated and tabulated.

The average energy per center of mass separation was then computed and plotted vs. the center of mass (figure 19). This function was then shifted so that it's plateau was at 0, and fitted to a Morse potential $E_{Morse} = D_0 \left[e^{-\alpha\left(-1+{}^{r_{ij}}/_{r_0}\right)} - 2\left(e^{-0.5\alpha\left(-1+{}^{r_{ij}}/_{r_0}\right)} \right) \right]$ using the least squares fitting

procedure previously discussed. This fitting is the effective potential of mean force between both sets of atoms, in the presence of water. This potential has both electrostatic and van der Waals contributions, which is critical since none of the pseudo-atoms are charged.

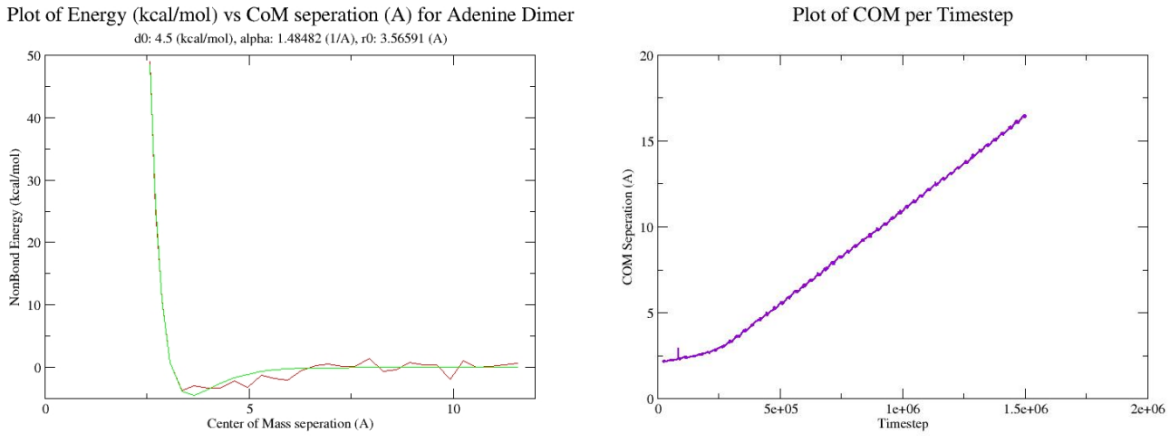


Figure 19: Energy profile for ADE Dimer (left) and COM vs Timestep profile (right)

ii) Meso-scale simulation of B-DNA dodecamer

a) Simulation Protocol

Bond stretches are the highest frequency modes in any MD simulation. The largest time step of any MD simulation is inversely proportional to the highest frequency modes, which is related to the force constant by $f = \frac{1}{2\pi}\sqrt{\frac{k}{\mu}}$ (assuming harmonic bonds), whereμ is the reduced mass of the parameter atoms being considered. A rule of thumb is that the largest time step possible should be 1/(6*f). The pseudo-atoms of this meso-scale force field are much heavier than those of regular atoms and largest bond force constants is 1/5 that of the typical largest atomistic force constant. Hence, during our meso-scale simulation, a much larger time step can be used. In these simulations, a time step of 10 fs is used, compared to 2 fs for a typical atomistic simulation. Further, the meso-scale system contains much less atoms than an atomistic one (each nucleotide is composed of 3 pseudo-atoms, compared to about 40 for an atomistic nucleotide). Finally, in a typical atomistic simulation, about 60% of the CPU

time is spent computing electrostatics. The meso-scale model avoid this by having neutral pseudo-atoms.

The simulation procedure is similar to before. Whereas before, 2fs time steps were used, now 10fs time steps are used and the system is simulated for a lot more steps due to it's reduced size and complexity of calculation. A 1.4Å crystal structure of the B-DNA dodecamer (355D) was downloaded, solvated in a TIP3 water box and neutralize. This fully atomistic structure was simulated using the usual procedure for 2ns of NPT dynamics, with snapshots of the last 500ps (every 5ps) saved for thermodynamic and structural analysis. The average MD structure was calculated as before and the ions were removed. It was then converted into its Mesoscale description, with the waters being modeled as inflated van der Waals sphere. This water model is the same as the M3B model of Valeria Molinero, and was parameterized to reproduce the density, diffusivity and cohesive energy of experimental bulk water. The meso-scale dodecamer is then simulated as before except with a larger time step and for $2x10^6$ steps, representing 2 micro-seconds of total simulation time. Snapshots of the system during the last 50ns of simulation were saved and used for thermodynamic and structural analysis. The average MD structure during this last 50ns was calculated and the atomistic level description reconstructed from this average structure. This reconstructed atomistic structure was then minimized and simulated for 1000 steps of NPT dynamics (figure 20).

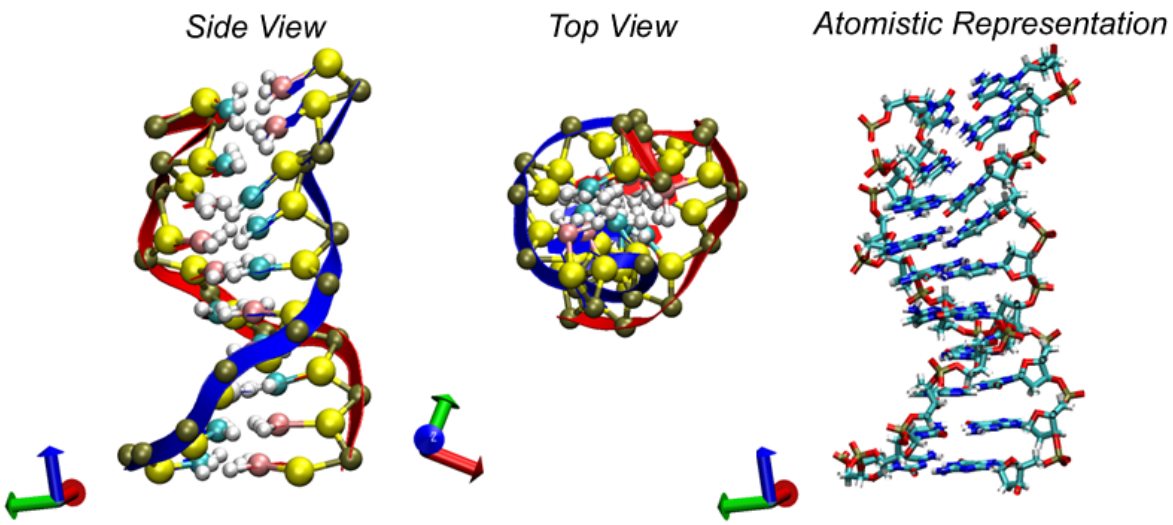


Figure 20: Pictures of Meso dodecamer and its atomistic reconstructed model

b) Comparison of Meso-scale and Atomistic Dodecamer simulations

The CRMS between the average MD meso dodecamer, the average MD atomistic dodecamer and the crystal structure is shown in table 6. From this data, it is evident that the simulation structures are quite different from the crystal structure, as well as each other. This is further illustrated in figure 21, where the average MD structures of both the meso (the atomistic reconstructed version) and the atomistic structures are compared. Observe the smooth backbone profile in the case of the atomistic dna, contrasted with the irregular profile in the meso-scale. The overlay of the two structure illustrated the sharp differences in the backbone torsions of the atomistic (blue) and the meso scale (red) structures. Notice there is considerably more distortion in the meso structure. In particular there are several backbone kinks in the structure. As shown in table 6, the backbone parameters for both sets of simulation are within the acceptable range of B-DNA, except for the backbone twist, where the average md structures are under twisted. This is a known issue with the AMBER force field. There is also significant deviation in the backbone twist for the meso dodecamer, which is a consequence of the aforementioned kinks in the structure.

	CRMS			Helical Parameters					
	Meso	Crystal	Atoms	Shift	Slide	Rise	Tilt	Roll	Twist
Meso	-	5.583	4.727	-0.23 (0.87)	-0.08 (0.7)	3.92 (1.28)	-1-12 (5.79)	5.7 (14.99)	19.22 (28.06)
Atoms	4.727	3.128	-	0.03 (0.28)	-0.25 (0.30)	3.5 (0.32)	0.48 (3.35)	8.74 (4.63)	27.69 (2.47)

Table 6: Structural Comparison of the Meso-scale and Atomistic dodecamer average structures. Deviations are in brackets

From these helical parameters, it is clear that further optimization of the meso-scale force field need to occur. In particular, the Phosphate-Sugar-Phosphate backbone twist angle and the Phosphate-Sugar-Phosphate-Sugar backbone dihedral angles need to be optimized to prevent the under-twisting of the helix seen in the dodecamer simulation. The helical rise in the Meso dodecamer is outside the accepted range for B-DNA (3.4 +/- 0.2), which when combined with the under twisting of the helix, point to unwinding of the helix. This is further indication of the need to optimize the backbone dihedrals.

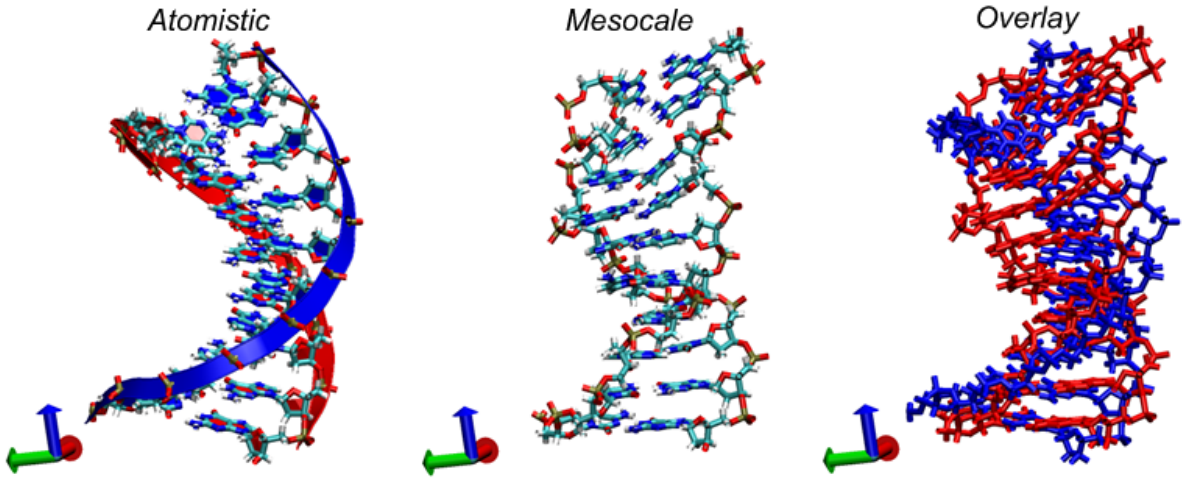


Figure 21: Comparison of average MD structures between Meso-scale and atomistic

b) Thermodynamic properties of Meso dodecamer

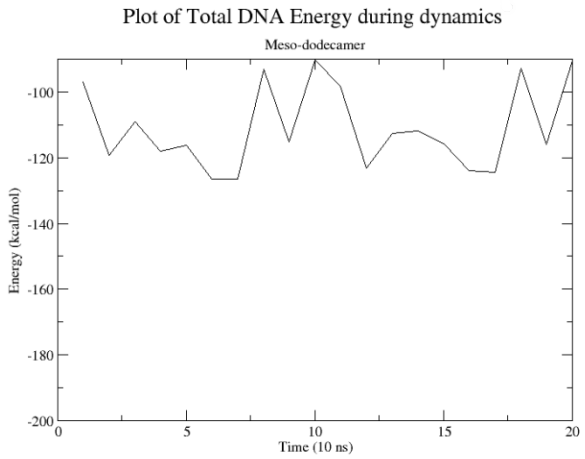


Figure 22: Total Energy of Meso Dodecamer only during MD. Solvent interactions are included

The total energy of the meso dodecamer was determined as described above. Each snapshot was minimized prior to per atom energy computation. Care was taken to neglect the solvent – solvent interaction, hence the energies obtained were the interaction of the dna with the solvent and itself only. The results are shown in figure 22. The total dna energy was then compared to the dna energy of the atomistic dna simulation. The meso dna energy differed from the average atomistic dna energy by 13.24% (+/- 3.5%) during the last 500 ns of the 2 micro-second run. This difference points suggest that further optimization of the non-bond parameters can be done.

One idea is to develop explicit water-meso dna off-diagonal van der Waals terms. With these present simulations, the off-diagonal terms are obtained from geometric combination rules. Explicit off diagonal terms would allow for the tuning of the meso dna – water interaction, without altering the water – water interactions, which have been optimized to reproduce the cohesive energy of bulk water. It should be pointed out though that since our model does not have charge – charge interactions, the agreement with atomistic simulation is somewhat remarkable.

Another source of error in this simulation is the neglect of ions. In fact, the sodium ions in these simulation were treated as regular waters in their interaction with the DNA. This is clearly inadequate as the ion effects in DNA are well known both experimentally and computationally. Interaction parameters unique to the different ions are currently being developed.

iii) Conclusions

The large fluctuations in the total energy of the system is attributed to dramatic conformational changes that occur, as the meso dodecamer is better able to sample its energy landscape than a typical atomistic simulation. This is primarily due to the length of the meso dna simulation, being 3 orders of magnitude longer than the atomistic level simulation. Further, timing tests indicates that the atomistic level simulation too twice as long to complete 2 ns than the meso scale did to complete 2 ms (150 vs 69 CPU hours). This is a remarkable results, as the meso scale simulation represents a x2000 times speedup. This speedup make it now feasible to study DNA system in the micro-second timescale. Further optimization of the force field is required and currently being pursued.