Commun. Comput. Phys.
doi: 10.4208/cicp.210711.121011s

Vol. **13**, No. 1, pp. 223-237
January 2013

Examining Electrostatic Influences on Base-Flipping: A Comparison of TIP3P and GB Solvent Models

Allyn R. Brice and Brian N. Dominy*

Department of Chemistry, Clemson University, Clemson, South Carolina, 29634, USA.

Received 21 July 2011; Accepted (in revised version) 12 October 2011

Available online 12 June 2012

Abstract. Recently, it was demonstrated that implicit solvent models were capable of generating stable B-form DNA structures. Specifically, generalized Born (GB) implicit solvent models have improved regarding the solvation of conformational sampling of DNA [1, 2]. Here, we examine the performance of the GBSW and GBMV models in CHARMM for characterizing base flipping free energy profiles of undamaged and damaged DNA bases. Umbrella sampling of the base flipping process was performed for the bases cytosine, uracil and xanthine. The umbrella sampling simulations were carried-out with both explicit (TIP3P) and implicit (GB) solvent in order to establish the impact of the solvent model on base flipping. Overall, base flipping potential of mean force (PMF) profiles generated with GB solvent resulted in a greater free energy difference of flipping than profiles generated with TIP3P. One of the significant differences between implicit and explicit solvent models is the approximation of solute-solvent interactions in implicit solvent models. We calculated electrostatic interaction energies between explicit water molecules and the base targeted for flipping. These interaction energies were calculated over the base flipping reaction coordinate to illustrate the stabilizing effect of the explicit water molecules on the flipped-out state. It is known that nucleic base pair hydrogen bonds also influenced the free energy of flipping since these favorable interactions must be broken in order for a base to flip-out of the helix. The Watson-Crick base pair hydrogen bond fractions were calculated over the umbrella sampling simulation windows in order to determine the effect of base pair interactions on the base flipping free energy. It is shown that interaction energies between the flipping base and explicit water molecules are responsible for the lower base flipping free energy difference in the explicit solvent PMF profiles.

PACS: 87.15.K-, 87.15.-v

Key words: TIP3P, generalized Born, flipping, implicit solvation, potential of mean force, DNA, DNA damage.

*Corresponding author. *Email addresses:* abrice@ccny.cuny.edu (A. R. Brice), dominy@clemson.edu (B. N. Dominy)

1 Introduction

Base flipping is the process of a DNA base moving out of the base stack, breaking the Watson-Crick (WC) base pair hydrogen bonds, and being completely exposed in the solvent medium. The process is known to be energetically unfavorable since base pair interactions are stronger than base interactions with solvent [3,4]. However, base flipping has been shown to occur spontaneously [5], and in some cases enzymes utilize base flipping for catalysis [6]. For example, uracil DNA glycosylase enzymes target the exposed base, and stabilize the flipped-out state for the purpose of base excision repair [7,8].

Several studies have investigated the effects of the base flipping conformational transition on enzyme function [5,7,9]. Experimental and theoretical methods have both been used to study the base flipping conformational change. The imino proton exchange with solvent during the base flipping can be measured with NMR, and is a common technique for evaluating the transition experimentally [7]. These experiments yield base opening rates as well as the equilibrium constant ($K_{flip} = k_{op}/k_{clsd}$) between flipped-in and flipped-out state. Umbrella sampling [10,11] is a computational method that is commonly used to examine base flipping free energy differences. The method is used to construct a potential of mean force (PMF) with respect to a progress variable of some known path or reaction coordinate [10,11]. An umbrella biasing potential is applied to sample across the chosen reaction coordinate, from one end-point to the other. The reaction coordinate for the path between the flipped-out and flipped-in states has been the focus of several studies [12–14].

When molecular dynamics is used to describe conformational changes of proteins or nucleic acids, a suitable force field is critical [15,16]. Priyakumar *et al.* [17] tested the performance of three force fields (CHARMM27 [18], AMBER4.1 [19], and BMS [20]) for the construction of DNA base flipping PMF profiles. Profiles for the GC base pair were generated with umbrella sampling, using a center of mass (COM) pseudodihedral angle [12] as the reaction coordinate. The duplex dodecamer sequence $d(GTCAGCGCATGG)_2$ was used for the base flipping. Along with the umbrella sampling, the WC base pair interaction energies were calculated. The interaction energy calculated with CHARMM was 21.9 kcal/mol, which is similar to the literature value [21] for the GC base pair interaction energy. However, the AMBER (26.3 kcal/mol) and BMS (26.2 kcal/mol) force fields overestimated the experimental value for the GC base pair interaction energy [21]. Equilibrium constants for base flipping measured with NMR proton exchange [22] were compared with the free energy difference results from the force fields. The results indicated that free energies generated with CHARMM and AMBER were more similar to experimental values than those generated with BMS [17].

Along with finding an optimal force field, another challenge when modeling DNA conformational changes has been accurately representing the solvent environment, while also maintaining computational efficiency. The conformational equilibria of nucleic acids in particular are strongly influenced by the solvent environment [16,23], thus highlighting the importance of accurately modeling the solvent during free energy calculations.

Explicit solvent models accurately account for the solute-solvent interactions, however explicitly solvated systems can easily increase the system size by a factor of 10-20. Therefore, it is computationally very strenuous to simulate a large biomolecule over long timescales with an explicit solvent system. Reducing the number of atoms and the solute-solvent interactions in the system greatly improves the speed of these calculations, making accessible scientific questions involving larger solutes and longer timescales [24, 25].

Implicit solvent models offer an alternative by representing the solvent as a function of the solute configuration. Many implicit solvent models have been developed [26–28]. Typically, the free energy of solvation ($\Delta G_{solvation}$) is broken down into the polar and nonpolar contributions (Eq. (1.1)):

$$\Delta G_{solvation} = \Delta G_{pol} + \Delta G_{nonpol}, \tag{1.1}$$

$$\Delta G_{nonpol} = \sum_i \gamma_i A_i. \tag{1.2}$$

The nonpolar contribution (Eq. (1.2)) is the cost of creating a cavity within the solvent moderated by the change in the vdw attraction between the solute and solvent, which is approximately proportional to the surface area (A_i) of the solute [26, 29]. When studying DNA, the polar solvation term is the dominant contribution to solvation due to the highly negative DNA backbone [30]. One class of implicit solvent models represents the solvent medium as a dielectric continuum in order to calculate the electrostatic free energy of solvation (ΔG_{pol}).

$$\nabla[\varepsilon(r)\nabla\varphi(r)] = -4\pi\rho(r). \tag{1.3}$$

The Poisson equation (Eq. (1.3)) [31] is commonly solved numerically by a finite difference method [32, 33], which can be computationally expensive when implemented in molecular dynamics or Monte Carlo simulations. When the influence of ionic strength is factored in, Eq. (1.3) becomes the Poisson-Boltzmann (PB) equation [31].

The PB equation yields an electrostatic potential $\varphi(r)$, where $\varepsilon(r)$ is the distance dependent dielectric, and $\rho(r)$ is the charge density of the biomolecule. The Poisson equation can be solved analytically, however those solutions are typically restricted to simple geometric shapes [31, 34].

$$\alpha_i = -\frac{1}{2}\left(\frac{1}{\varepsilon_p} - \frac{1}{\varepsilon_w}\right)\frac{1}{G^i_{pol}}. \tag{1.4}$$

The Born equation (Eq. 1.4)) [35] is the solvation of a single ion in a dielectric medium, where G_{pol} is the electrostatic free energy of solvation, ε_p is the low dielectric medium of the solute, and ε_w is the high dielectric medium of the solvent. The effective Born radius (α) roughly corresponds to the distance between atom i and the solvent boundary. An extension of the Born equation is the generalized Born (GB) equation (Eq. (1.5)), where the empirical factor F may range from 2 to 10, while 4 is the most common value [36]. The inverse Debye length (κ), which is proportional to the square root of the electrolyte ionic strength ($(I)^{1/2}$), represents the salt effects [37]. There are several analytic generalized Born (GB) models that are optimized to reproduce Poisson solvation energies using rapidly solved parameterized equations [26–28, 38, 39]. The effective distance to the

solvent-solute boundary, or the Born radius, is typically determined from the solute volume. In the CHARMM package, the GBMV (Generalized Born using Molecular Volume) model integrates over overlapping van der Waals spheres to represent the molecular volume [39]. Solvation energies calculated using the GBMV algorithm were found to exhibit an error of less than 1% relative to the values calculated using the PB equation [24]. A similar method for calculating the GB equation is used in the GBSW (Generalized Born with simple SWitching) model [38]. However, in GBSW the molecular surface calculation is replaced with a smooth van der Waals surface, making it less expensive than the GBMV model. Their speed and remarkable accuracy have made the use of these models very popular, particularly in accompanying molecular dynamics (MD) or Monte Carlo (MC) conformational sampling protocols.

$$\Delta G_{pol} = -\frac{1}{2}\left(\frac{1}{\varepsilon_p} - \frac{-e^{\kappa f_{GB}}}{\varepsilon_w}\right)\sum_{i,j}\frac{q_i q_j}{f_{GB}}, \tag{1.5}$$

$$f_{GB} = \left[r_{i,j}^2 + \alpha_i\alpha_j \exp(-r_{i,j}^2/F\alpha_i\alpha_j)\right]^{1/2}. \tag{1.6}$$

An accurate description of the solvent dielectric boundary is dependent on the atomic radii. The solvent dielectric boundary, is critical in generalized Born calculations for the accurate evaluation of the Born radii. In a recent study [2], two sets of atomic radii were compared, the atomic van der Waals radii and the atomic radii developed by Banavali and Roux (BR) [40]. Molecular dynamics simulations of a DNA dodecamer were performed with a generalized Born and TIP3P solvent model. These comparisons were analyzed with several DNA helical properties over the corresponding DNA trajectories. Molecular dynamics simulations generated with the implicit solvent displayed stable B-form DNA structures relative to the explicit solvent. These results agreed with previous studies that observed stable B-form simulations using a generalized Born solvent [1, 41, 42]. Both sets of atomic radii performed well in generating stable DNA conformations.

The choice of solvent model is crucial when modeling DNA structures. In the current study, we demonstrate the effects of solvent model on the base flipping thermodynamics of undamaged and damaged DNA bases. For the purpose of examining the performance of the GB model, the GB and TIP3P solvent models were compared during umbrella sampling calculations of the base flipping process. Additionally, the influence of the interior dielectric constant on the base flipping free energy difference was evaluated. The duplex dodecamer DNA sequence d(GTCAGCGCATGG)$_2$ was used for easy comparison to the PMF profiles from Priyakumar *et al.* [2]. The natural base pair GC was analyzed along with the damaged DNA bases uracil and xanthine.

2 Methods

A summary of the simulation and umbrella sampling procedures are provided below. Trajectory analysis methods are also described.

2.1 Starting structures

The CHARMM c32b1 molecular mechanics package and the CHARMM27 [18, 43] all-atom nucleic acid force field were used in all molecular dynamics simulations. The starting coordinates of the dodecamer sequence d(GTCAGFGCATGG)$_2$ were generated within CHARMM. The base to be flipped-out of the helix is represented as "F". This sequence was chosen because it has been used in many base flipping studies previously [14, 17] and provides an easy comparison for the PMF profiles. The program 3DNA [44] was used to create a canonical B-form DNA structure of the sequence d(GTCAGCGCATGG)$_2$. The base targeted for flipping (F) in each of the DNA models was systematically modeled as cytosine, uracil or xanthine, while the base complementary to F was modeled as one of the natural DNA bases (guanine, adenine, cytosine and thymine).

2.2 Base flipping umbrella sampling

Base flipping potentials of mean force (PMF) were constructed from these 9 starting structures of B-form DNA, following the methods of Priyakumar *et al.* [17]. Umbrella sampling was performed to calculate the PMF associated with base flipping using an explicit and implicit solvent. The pseudo-dihedral angle used in Priyakumar *et al.* was applied as a reaction coordinate [17]. The pseudo-dihedral angle was defined through the centers of mass (COM) [12] corresponding to a) the base pair on the 3′ side of the flipping base b) the sugar of the base on the 3′ side of the flipping base c) the sugar of the flipping base and d) the flipping base. The PMF profiles were created by incrementing the pseudo-dihedral angle 5° in each simulation window for 0° - 360° (72 windows). A pseudo-dihedral angle of 0° - 30° is defined as the base-paired state and an angle of 190° is defined as the flipped-out state. Minimizations of 100 steps were performed with the adopted basis Newton-Rapheson, and using the miscellaneous mean field potential (MMFP) module in the charmm package, the pseudo-dihedral angle was varied to create the starting structures. A force constant of 10,000 kcal/mol/rad^2 was used for the MMFP restraint. Starting structures were varied $\pm 5^\circ$ from the final structures of the previous minimization. A molecular dynamics simulation was performed for each window of the reaction coordinate using the starting structures. A harmonic umbrella potential

$$w_i(x) = k_i(x - x_i)^2 \tag{2.1}$$

was used to restrain the pseudo-dihedral angle with a force constant (k_i) of 1000 kcal/mol/rad^2. Harmonic restraints (force constant of 2 kcal/mol/rad^2) were applied to the four terminal bases to keep them from fraying, and the covalent hydrogen bond distances were constrained by SHAKE. The nonbonded cutoffs were 14Å, with a switching function from 10Å to 12Å. The GBMV2 module [39] was used as the implicit solvent system since it was determined by Feig *et al.* [45] to closely reproduce PB solvation energies. For GBMV, we used a β value of -20, and a water probe radius of 1.4Å. β is an adjustable parameter controlling the width of the switching function of GBSW module. The inverse

Debye length (κ) [37] was set to 0.129Å^{-1}, which corresponds to the physiological salt concentration (0.15M). Nonpolar contributions (Eq. (1.2)) to the solvation free energy are accounted for here as the product of the solvent accessible surface area (A) and the surface tension (γ) [26, 29]. The surface tension was set to 0.03kcal/mol/Å^2 consistent with prior studies to calculate the nonpolar solvation energy [38]. Systems were heated from 200K to 300K in increments of 1K every 2ps for a total of 200ps. Langevin dynamics were used, with an integration timestep of 1fs, to construct a canonical ensemble (NVT). The GBSW [38] solvent model was used to test the influence of the interior dielectric constant (ε_p). Umbrella sampling was performed for the GC base pair and the damaged base pairs using the procedure above, and the GBSW solvent model. The dielectric constant was increased from 1.0 to 2.0 for the GBSW solvent model, and the nonbond interactions to generate PMF profiles for GC and the damaged bases. In the TIP3P solvated systems a water box was created, which resulted in the solvent extending 13Å beyond the longest DNA axis, and 24Å beyond the perpendicular axis. Systems were heated from 200K to 300K in increments of 1K every 2ps for a total of 200ps. A Langevin barostat was used with an integration timestep of 2fs, to construct an isothermal-isobaric ensemble for equilibration (NPT). A canonical ensemble (NVT) was then created with the Andersen thermostat for the 1ns production phase. Long distance electrostatic interactions were accounted for using a particle-mesh Ewald summation [46]. In order to achieve a neutral system necessary for the efficient calculation of long-range electrostatics using Ewald summation, 22 sodium ions were added. The pseudo-dihedral values were recorded throughout all of the trajectories, and used to calculate a probability distribution. The weighted histogram analysis method (WHAM) was used to create unbiased PMF profiles [47]. The interaction energies between explicit solvent and flipping base (F) were calculated using INTER module in the CHARMM package. Energies were calculated over umbrella sampling windows of the GC base pair flipping. Explicit waters within 5Å of the flipping base were included in the calculation. Interaction energies were also calculated between the flipping base and its complementary base. Hydrogen bond fractions were determined for the base pair hydrogen bonds by using the QUICK module in CHARMM to calculate the bond distances and angles.

3 Results and discussion

The following results are organized to clearly describe the effects of a solvent model on the base flipping process, and also breakdown the contributions to these effects in detail. Firstly, the PMF profiles, generated in implicit and explicit solvent, for the natural base pair GC are shown. Secondly, the profiles, generated in implicit and explicit solvent, for the damaged base pairs of uracil and xanthine are reported. Profiles with an adjusted interior dielectric constant are also provided. As support, the WC base pair hydrogen bond fractions are provided, as well as the interaction energies between explicit waters with the flipping base.

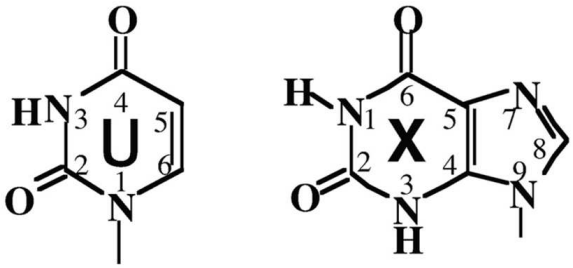

Figure 1: Structures of the deaminated bases uracil (U) and xanthine (X) [8].

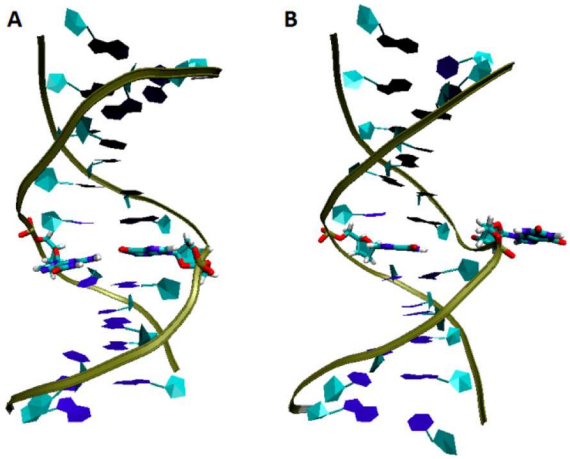

Figure 2: Base flipping structures for xanthine in TIP3P solvent. The above structures were averaged over the corresponding umbrella sampling window. A. Structure of the flipped-in state. B. Structure of the flipped-out state.

3.1 Free energy of base flipping

Comparing GB and TIP3P

Base flipping umbrella sampling of the natural base pair GC was performed for a convenient comparison to results from Priyakumar *et al.* [17]. The results from base flipping, where cytosine was the flipping base, are displayed in Fig. 3. Free energies are plotted along the pseudodihedral angle, which was employed as the reaction coordinate (described in methods) for umbrella sampling. Simulations of umbrella sampling windows were solvated with a GB solvent model (Fig. 3A black), and TIP3P solvent model (Fig. 3A red). From these results, it can be seen that generation of base flipping profiles yields divergent free energy differences when using GB or TIP3P as the solvent model.

The region where the base is outside of the stack (60° - 300°), displays a varying shape for the two profiles. The TIP3P profile is more linear, which is similar to the profile from Priyakumar *et al.*. Both of the PMF profiles have a flipped-in state at ∼10°, which is consistent with results from Priyakumar *et al.* [17]. The profile generated using GB resulted

in a 31.0kcal/mol free energy difference, while the profile generated using TIP3P resulted in a 19.4kcal/mol free energy difference. This indicates the explicit solvent model favors the flipped out state more than the implicit solvent model.

The PMF profile of base flipping was impacted by altering the interior dielectric constant of the GB solvent model. In continuum solvent models the solvent is represented as a high-dielectric medium, and the solute is treated as a low dielectric medium. Values for the lower dielectric constant (ε_p in Eq. (1.5)) are typically chosen in order to account for electronic polarizability of the solute molecule [31]. Previous studies have determined that an interior dielectric for biomolecules such as proteins and membranes, can be adjusted from 2-4 when using implicit solvent models [48, 49]. Fig. 3B shows the effect of a higher interior dielectric constant on the base flipping free energy difference. When the dielectric of 2.0 was employed, the free energy difference for base flipping was more similar to the explicit solvent results (Fig. 3A red). Electrostatic interaction energies between the base targeted for flipping and the complementary WC base became less favorable as the interior dielectric was increased. The base pair interaction energy was -20.56 ± 2.24 when 1.0 was used for the interior dielectric, then decreased by -11.1 ± 1.08 when the higher dielectric of 2.0 was used. Increasing the interior dielectric constant influences the solvation energy as well. Since the higher interior dielectric constant is more similar to the water dielectric constant, the movement of the base from the base-paired state to the complete solvent exposed state results in less of a solvation energy difference. The destabilization of the base pair interactions and the weakened solvation energy difference likely both contributed to the decrease in base flipping free energy after the interior dielectric adjustment.

Damaged DNA base flipping

In order to examine the base flipping equilibria of damaged bases, the free energy difference between the flipped-out and flipped-in states were represented with PMF profiles. In Fig. 4, the base flipping PMF profiles for the damaged bases uracil and xanthine with the four complementary DNA bases are shown. These profiles were generated with umbrella sampling, where GB and TIP3P solvent models were both used for solvation of the base flipping process. In general, when the implicit solvent model is used, the base flipping umbrella sampling produces a greater free energy difference (Figs. 4A and D). This implies the base undergoing flipping favors the flipped-out state in explicit solvent over implicit solvent. Since the implicit solvent model approximates solute-solvent interactions of explicit waters, the hydrogen bonds between the flipped-out base and explicit waters may be responsible for the discrepancy between the solvent models.

Variations in the effect of solvent are observed for the damaged bases after a detailed comparison of the GB and TIP3P PMF profiles. The PMF profiles of uracil are influenced by the GB solvent similarly to the GC profile. The uracil profiles generated using GB solvent (Fig. 4A) are greater in energy than those generated using TIP3P (Fig. 4B), but the arrangement of the profiles remains constant regardless solvent model. In uracil, only the flipped-out state appears to be affected by the difference in solvent models. Therefore, it

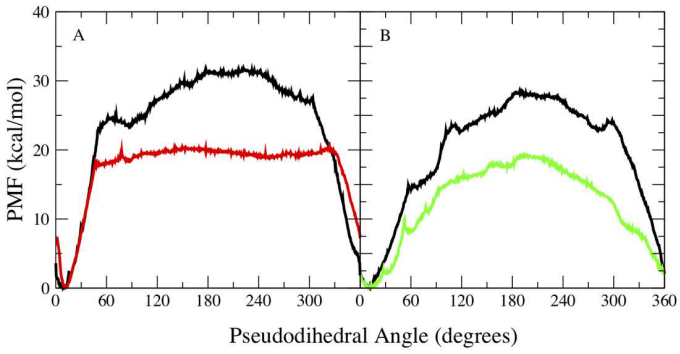

Figure 3: Potentials of mean force (PMF) of base flipping for GC base pairs along the pseudodihedral angle coordinate. Watson-Crick base pairing is approximately 10° - 30° pseudodihedral angle and the flipped out state is approximately 190°. A. Umbrella sampling performed with GBMV implicit solvent (black) with interior dielectric of 1.0 and TIP3P explicit solvent (red) B. Umbrella sampling performed with GBSW implicit solvent, using an interior dielectric of 1.0 (black) and 2.0 (green).

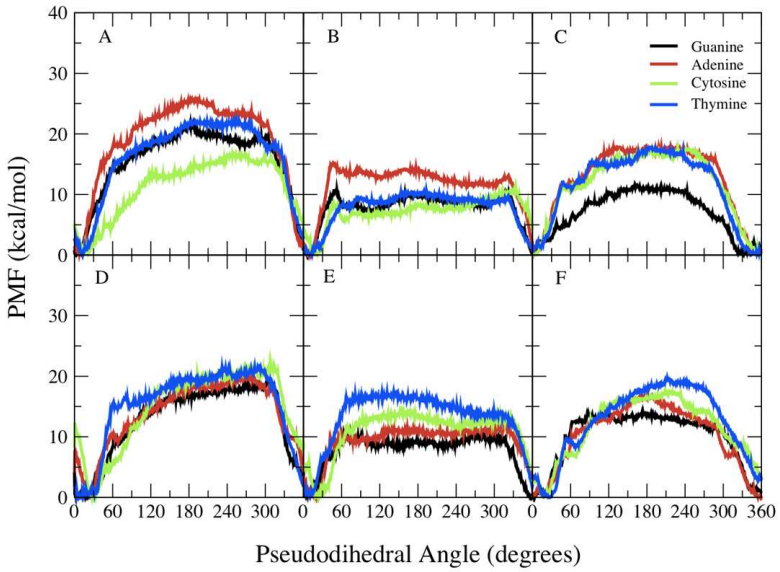

Figure 4: Potentials of mean force (PMF) of uracil- containing (A,B,C) and xanthine- containing (D,E,F) base pairs along the pseudodihedral angle coordinate. Watson-Crick base pairing is approximately 10°-30° pseudodihedral angle and the flipped out state is approximately 190°. A. Base flipping PMF profiles generated with GBMV solvent model and $\varepsilon_p = 1.0$. B. Base flipping PMF profiles generated with TIP3P solvent model. C. Base flipping PMF profiles generated with GBSW solvent model and $\varepsilon_p = 2.0$. D. Base flipping PMF profiles generated with GBMV solvent model and $\varepsilon_p = 1.0$. E. Base flipping PMF profiles generated with TIP3P solvent model. F. Base flipping PMF profiles generated with GBSW solvent model and $\varepsilon_p = 2.0$.

was hypothesized that discrepancies in base flipping free energy differences between the solvent models were a result of solute-solvent interactions. In the PMF profiles generated for xanthine, the GB solvent model (Fig. 4D) yields greater base flipping free energy

differences than the TIP3P solvent model (Fig. 4E). Also, the base flipping profiles for xanthine are narrowly distributed when implicit solvent is used, and with explicit solvent, xanthine base flipping profiles are more broadly spread. Solvent models may affect not only the flipped-out state, but also the flipped-in state. The number of hydrogen bonds in the flipped-in state influences the stability of the base pair, and in effect the base flipping free energy difference. The difference between GB and TIP3P xanthine PMF profiles was hypothesized to be caused not only by the interactions between solvent molecules and the flipped-out base, but also the interactions between the WC base pairs.

Adjusting the interior dielectric constant of the implicit solvent model during umbrella sampling of the damaged bases improved the agreement with explicit solvent PMF profiles. Since the GC base pair displayed improved results with a higher interior dielectric constant (Fig 3B), umbrella sampling of the base flipping for the damaged bases was performed with a raised dielectric. The interior dielectric constant was increased from 1.0 to 2.0 in the GB solvent model and non-bonded interactions. Uracil and xanthine PMF profiles with increased dielectric constants are shown in Figs. 4C and 4F. In the PMF profiles with the higher dielectric, a lower base flipping free energy difference is observed for uracil and xanthine, bringing them closer to the TIP3P results. However, when the interior dielectric is raised, the rank order of the uracil base flipping free energy differences does not agree with previous GB or TIP3P solvent profiles. In the GB simulations with increased dielectric constant, the DNA helix structure is distorted, which results in the discrepancy between the orders of the free energy differences. While the increased interior dielectric constant provides base flipping free energy differences that are quantitatively similar to explicit solvent profiles, the free energy differences are destabilized at the cost of the DNA backbone structure.

3.2 Interacting with the flipped-out base

Since base flipping requires the flipping base to disrupt the favorable base pair interactions when it leaves the base stack, the difference in base pair interactions may influence the flipping free energies. In Priyakumar *et al.* the interaction energies between the WC base pairs were calculated to show differences in the force fields (CHARMM, AMBER, BMS) for the base pair interaction. Here, interaction energies were calculated between the two bases in the WC base pair region (0° - 30°) for GB and TIP3P of GC umbrella sampling windows. These energies both agreed with interaction energies reported in Priyakumar *et al.* ($\sim$20kcal/mol) for the CHARMM27 force field, and with the experimental GC interaction energies used in the parameterization of the CHARMM27 force field [17, 18]. Therefore the discrepancies observed in the base flipping PMF profiles can be attributed to solvent-solute interactions in the flipped-out state.

In order to confirm the favorable solute-solvent interactions with explicit waters, interaction energies were calculated between the flipping base and TIP3P water molecules through the umbrella sampling of the base flipping. All electrostatic interactions between water molecules and the base undergoing flipping were included within a 5Å

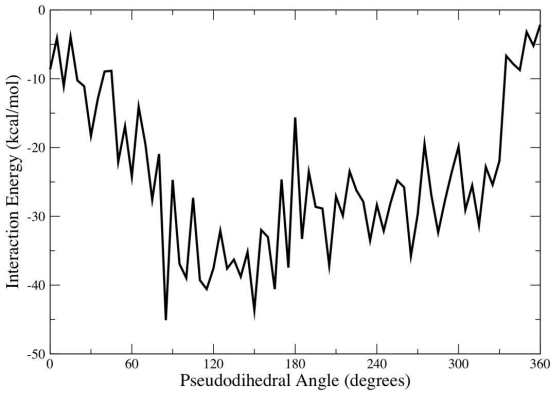

Figure 5: Interaction energies between cytosine base as a target of flipping and TIP3P explicit water molecules (within 5Å of base). Energies averaged over 500ps of production trajectory for each pseudodihedral simulation window.

cutoff. It can be seen in Fig. 5 that interactions between the base undergoing flipping and water molecules are the most favorable in the flipped out state ($\sim$190°). Priyakumar *et al.* demonstrated that the solvent accessible surface area is greatest for the flipping base from 60° to 330°, which is the region of the most favorable interactions between the TIP3P and flipping base [17]. Interaction energies show approximately a 30 kcal/mol difference between the flipped-in and flipped-out state. This interaction energy in the flipped-out state significantly stabilizes the flipped-out state of the explicit solvent simulations. The implicit solvent model attempts to represent solute-solvent interactions, but has been shown to underestimate these interactions relative to explicit solvent interactions [2, 24, 26].

3.3 Watson Crick hydrogen bonds

A more detailed analysis of the WC base pairing for uracil and xanthine were performed in order to understand the differences in the xanthine PMF profiles. During the base flipping process, base pair hydrogen bonds must be broken. Therefore, WC base pair hydrogen bonds have a significant influence on the base flipping free energy difference. Hydrogen bond fractions of the WC base pairs were calculated for the base pair (flipped-in region) windows, and are displayed in Table 1. These percentages show that in GB solvent uracil and xanthine base pairs are very stable and maintained throughout the trajectory. However, in TIP3P the GX and AX base pairs form fewer hydrogen bonds than their corresponding trajectories generated with GB solvent. The xanthine PMF profiles exhibited similar variations in the AX and GX profiles. In the TIP3P profiles, the AX and GX have a lower free energy difference than CX and TX, which is due to weaker base pair interactions. In the GB solvent PMF profiles, the xanthine base pairs display relatively the same base flipping free energy difference, which can be attributed to the forming of similar WC base pair interactions.

Table 1: Uracil and Xanthine H-bond fractions.

Hbond Fractions for Implicit and Explicit solvent WC Base Pair Simulations.					
		Complementary Base			
		Guanine	Adenine	Cytosine	Thymine
TIP3P Solvent					
	Uracil	96%	96%	91%	84%
	Xanthine	48%	50%	97%	98%
GB Solvent					
	Uracil	98%	99%	97%	99%
	Xanthine	86%	99%	95%	95%

4 Conclusions

Umbrella sampling was performed over the base flipping pathways of natural and damaged DNA base pairs. The influence from electrostatic interactions, and more specifically solvent interactions, was determined by utilizing two solvent models during the umbrella sampling simulations. When explicit solvent was used, umbrella sampling of the GC base pair qualitatively agreed with results from Priyakumar *et al.* [17]. However, the PMF profiles of the GC and uracil base pairs showed the WC flipped-in state was overstabilized in the implicit solvent model. It was hypothesized that differences in solute-solvent interactions in the flipped-out state were responsible for the discrepancies between the solvent models for the GC and uracil PMF profiles. We confirmed this by calculation of interaction energies over the base flipping coordinate, between the flipping base and explicit waters. It was demonstrated that the flipping base forms favorable interactions with the solvent in the flipped-out state. These interactions are not represented in the GB solvent model, and therefore the flipped-out state is less stable than when explicit solvent is used, as demonstrated by the PMF profiles. The damaged DNA base xanthine produced a unique trend for the solvent effect on PMF profiles. The PMF profiles generated with GB solvent were narrowly distributed, while the profiles from TIP3P were more broadly distributed. This was attributed to the difference in hydrogen bonding in the WC flipped-in state. Hydrogen bond fractions of the WC flipped-in state showed that the GB solvent model produced similar interaction patterns for the four xanthine base pairs, while TIP3P resulted in fewer hydrogen bonds for the AX and GX base pairs. The differences in flipped-in interactions led to the differences in PMF profile distribution. Xanthine most likely displayed this difference because of the greater number of hydrogen bond acceptors and donors. The GB solvent models have demonstrated their ability of representing stable structures of the B-form DNA [1, 2, 42]. However, the PMF profiles from the current GBMV solvent model with a dielectric of unity, did not compare well to the explicitly solvated systems. Evidence was shown that adjusting the interior dielectric constant (ε_p) lowered the free energy difference of base flipping for the GC base pair, in effect making it more similar to explicit solvent free energy difference. PMF profiles of the damaged bases also displayed improved agreement with explicit solvent

when the interior dielectric was raised, but the ranking of the free energies of base flipping calculated using different basepairings did not agree with previous GB and TIP3P solvent results.

Acknowledgments

We thank Dr. Weiguo Cao and Dr. Hyun-wook Lee for their helpful discussions. This work has been supported by National Science Foundation Career Award MCB-0953783 (to B.N.D.).

References

[1] V. Tsui, D.A., Case, Molecular dynamics simulations of nucleic acids with a generalized born solvation model, Journal of the American Chemical Society, 122 (2000) 2489-2498.
[2] J. Chocholousova, M. Feig, Implicit solvent simulations of DNA and DNA-protein complexes: Agreement with explicit solvent vs experiment, Journal of Physical Chemistry B, 110 (2006) 17240-17251.
[3] J.T. Stivers, Extrahelical damaged base recognition by DNA glycosylase enzymes, Chemistry A European Journal, 14 (2008) 786-793.
[4] J.T. Stivers, Site-specific DNA damage recognition by enzyme-induced base flipping, Progress in Nucleic Acid Research and Molecular Biology, 44 (2004) 37-65.
[5] K. Snoussi, J.L. Leroy, Imino proton exchange and base-pair kinetics in RNA duplexes, Biochemistry, 40 (2001) 8898-8904.
[6] P.J. Berti, J.A.B. McCann, Toward a detailed understanding of base excision repair enzymes: Transition state and mechanistic analyses of N-glycoside hydrolysis and N-glycoside transfer, Chemical Reviews, 106 (2006) 506-555.
[7] J.B. Parker, M.A. Bianchet, D.J. Krosky, J.I. Friedman, L.M. Amzel, J.T. Stivers, Enzymatic capture of an extrahelical thymine in the search for uracil in DNA, Nature, 449 (2007) 433-438.
[8] H.W. Lee, A.R. Brice, C.B. Wright, B.N. Dominy, W. Cao, Identification of escherichia coli mismatch-specific uracil DNA glycosylase as a robust xanthine DNA glycosylase, Journal of Biological Chemistry, 285 (2010) 41483-41490.
[9] U.D. Priyakumar, A.D. MacKerell, Computational approaches for investigating base flipping in oligonucleotides, Chemical Reviews, 106 (2006) 489-505.
[10] G.M. Torrie, J.P. Valleau, Monte-Carlo study of a phase-separating liquid-mixture by umbrella sampling, Journal of Chemical Physics, 66 (1977) 1402-1408.
[11] P. Kollman, Free-energy calculations: Applications to chemical and biochemical phenomena, Chemical Reviews, 93 (1993) 2395-2417.
[12] N. Banavali, A.D. Mackerell Jr., Free energy and structural pathways of base flipping in a DNA GCGC containing sequence, Journal of Molecular Biology, 319 (2002) 141-160.
[13] K. Song, A.J. Campbell, C. Bergonzo, C. Santos, A.P. Grollman, C. Simmerling, An improved reaction coordinate for nucleic acid base flipping studies, Journal of Chemical Theory and Computation, 5 (2009) 3105-3113.
[14] U.D. Priyakumar, A.D. Mackerell Jr., Computational approaches for investigating base flipping in oligonucleotides, Chemical Reviews, 106 (2006) 489-505.

[15] T.E. Cheatham, P.A. Kollman, Molecular dynamics simulation of nucleic acids, Annual Review of Physical Chemistry, 51 (2000) 435-471.
[16] J. Norberg, L. Nilsson, Molecular dynamics applied to nucleic acids, Accounts of Chemical Research, 35 (2002) 465-472.
[17] U.D. Priyakumar, A.D. Mackerell Jr., Base flipping in a GCGC Containing DNA Dodecamer: A comparative study of the performance of the nucleic acid force fields, CHARMM, AMBER, and BMS, Journal of Chemical Theory and Computation, 2 (2006) 187-200.
[18] A.D. MacKerell, N.K. Banavali, All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution, Journal of Computational Chemistry, 21 (2000) 105-120.
[19] T.E. Cheatham, P. Cieplak, P.A. Kollman, A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat, Journal of Biomolecular Structure and Dynamics, 16 (1999) 845-862.
[20] D.R. Langley, Molecular dynamic simulations of environment and sequence dependent DNA conformations: The development of the BMS nucleic acid force field and comparison with experimental results, Journal of Biomolecular Structure and Dynamics, 16 (1998) 487-509.
[21] N. Foloppe, A.D. Mackerell Jr., All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data, Journal of Computational Chemistry, 21 (1999) 86-104.
[22] U. Dornberger, M. Leijon, H. Fritzsche, High base pair opening rates in tracts of GC base pairs, Journal of Biological Chemistry, 274 (1999) 6957-6962.
[23] T.E. Cheatham, M.F. Crowley, P.A. Kollman, A molecular level picture of the stabilization of A-DNA in mixed ethanol-water solutions, Proceedings of the National Academy of Sciences of the United States of America, 94 (1997) 9626-9630.
[24] M. Feig, J. Chocholousova, S. Tanizaki, Extending the horizon: Towards the efficient modeling of large biomolecular complexes in atomic detail, Theoretical Chemistry Accounts, 116 (2006) 194-205.
[25] Z.A. Sands, C.A. Laughton, Molecular dynamics simulations of DNA using the generalized born solvation model: Quantitative comparisons with explicit solvation results, Journal of Physical Chemistry B, 108 (2004) 10113-10119.
[26] M. Feig, C.L. Brooks, Recent advances in the development and application of implicit solvent models in biomolecule simulations, Current Opinion in Structural Biology, 14 (2004) 217-224.
[27] B.N. Dominy, C.L. Brooks, Development of a generalized born model parametrization for proteins and nucleic acids, Journal of Physical Chemistry B, 103 (1999) 3765-3773.
[28] B. Roux, T. Simonson, Implicit solvent models, Biophysical Chemistry, 78 (1999) 1-20.
[29] T. Ooi, M. Obatake, G. Nemethy, H.A. Scheraga, Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides, Proceedings of the National Academy of Sciences of The United States of America, 84 (1987) 3083-3090.
[30] K.J. McConnell, D.L. Beveridge, DNA Structure: What's in charge?, Journal of Molecular Biology, 304 (2000) 803-820.
[31] K.A. Sharp, B. Honig, Electrostatic interactions in macromolecules: Theory and applications, Annual Review of Biophysics and Biophysical Chemistry, 19 (1990) 301-332.
[32] A.H. Boschitsch, M.O. Fenley, H.X. Zhou, Fast boundary element method for the linear Poisson-Boltzmann equation, Journal of Physical Chemistry B, 106 (2002) 2741-2754.
[33] M. Holst, N. Baker, F. Wang, Adaptive multilevel finite element solution of the Poisson-

Boltzmann equation I. Algorithms and examples, Journal of Computational Chemistry, 21 (2000) 1319-1342.
[34] B.H. Zimm, M. Lebret, Counterion condensation and system dimensionality, Journal of Biomolecular Structure and Dynamics, 1 (1983) 461-471.
[35] M. Born, Volumen und hydratationswärme der ionen, Z. Phys., 1 (1920) 45-48.
[36] W.C. Still, A. Tempczyk, R.C. Hawley, T. Hendrickson, Semianalytical treatment of solvation for molecular mechanics and dynamics, Journal of the American Chemical Society, 112 (1990) 6127-6129.
[37] J. Srinivasan, M.W. Trevathan, P. Beroza, D.A. Case, Applications of a pairwise generalized born model to proteins and nucleic acids: Inclusion of salt effect, Theoretical Chemistry Accounts, 101 (1999) 426-434.
[38] W. Im, M.S. Lee, C.L. Brooks, Generalized born model with a simple smoothing function, Journal of Computational Chemistry, 24 (2003) 1691-1702.
[39] M.S. Lee, M. Feig, F.R. Salsbury, C.L. Brooks, New analytic approximation to the standard molecular volume definition and its application to generalized born calculations, Journal of Computational Chemistry, 24 (2003) 1348-1356.
[40] N. Banavali, B. Roux, Atomic radii for continuum electrostatic calculations on nucleic acids, Journal of Physical Chemistry B, 106 (2002) 11026-11035.
[41] J. Srinivasan, T.E. Cheatham, P. Cieplak, P.A. Kollman, D.A. Case, Continuum Solvent studies of the stability of DNA, RNA, and phosphoramidate - DNA helices, Journal of the American Chemical Society, 120 (1998) 9401-9409.
[42] A.R. Brice, B.N. Dominy, Analyzing the robustness of the MM/PBSA free energy calculation method: Application to DNA conformational transitions, Journal of Computational Chemistry, 32 (2011) 1431-1440.
[43] A.D. MacKerell, D. Bashford, M. Bellott, R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus, All-atom empirical potential for molecular modeling and dynamics studies of proteins, Journal of Physical Chemistry B, 102 (1998) 3586-3616.
[44] X.J. Lu, W.K. Olson, 3DNA: A software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures, Nucleic Acids Research, 31 (2003) 5108-5121.
[45] J. Chocholousova, M. Feig, Balancing an accurate representation of the molecular surface in generalized born formalisms with integrator stability in molecular dynamics simulations, Journal of Computational Chemistry, 27 (2005) 719-729.
[46] T. Darden, D. York, L.J. Pedersen, Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems, Journal of Chemical Physics, 98 (1993) 10089.
[47] S. Kumar, D. Bouzida, R.H. Swendsen, P.A. Kollman, J.M. Rosenberg, The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method, Journal of Computational Chemistry, 13 (1992) 1011-1021.
[48] M. Feig, A. Onufriev, M.S. Lee, W. Im, D.A. Case, C.L. Brooks, Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures, Journal of Computational Chemistry, 25 (2003) 265-284.
[49] T. Simonson, Macromolecular electrostatics: Continuum models and their growing pains, Current Opinion in Structural Biology, 11 (2001) 243-252.