8-2011

A THEORETICAL INVESTIGATION EXAMINING DNA CONFORMATIONAL CHANGES AND THEIR EFFECTS ON GLYCOSYLASE FUNCTION

Allyn Brice
Clemson University, allyn009@yahoo.com

A THEORETICAL INVESTIGATION EXAMINING DNA CONFORMATIONAL CHANGES AND THEIR EFFECTS ON GLYCOSYLASE FUNCTION

A Dissertation
Presented to
the Graduate School of
Clemson University

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

---

by
Allyn Raymond Brice
August 2011

Accepted by:
Dr. Brian N. Dominy, Committee Chair
Dr. Steven J. Stuart
Dr. Dev P. Arya
Dr. Jason McNeill

ABSTRACT

Glycosylase enzymes initiate the process of base excision repair (BER) in order to prevent the irreversible modification of the genome. In the BER process a damaged DNA base is recognized, removed from the DNA sequence, and then the remaining abasic site is repaired. Glycosylase enzymes are responsible for the base recognition mechanism and catalysis of the base excision. One of the most studied glycosylase superfamilies is uracil DNA glycosylase (UDG). The UDG superfamily has demonstrated specificity for excising uracil, which is the deamination product of cytosine, from DNA sequences of prokaryotes and eukaryotes. Mismatch-specific uracil DNA glycosylase (MUG) is a member of the UDG superfamily, and interestingly has shown specificity for both uracil and xanthine bases.[1]

The following dissertation provides an anlaysis on the recognition mechanism of *E. coli* MUG for deaminated DNA bases. Glycosylase enzymes require the damaged base to be flipped out of the base stack, and into an active site for catalysis of the N-glycosidic cleavage. Typically, recognition of substrates by enzymes is characterized by binding affinities, but in the following work the binding of *E.Coli* MUG is broken down into contributions from the base flipping and enzyme binding equilibria.

Since DNA conformational changes play a large role in UDG systems, the robustness of molecular mechanics Poisson-Boltzmann surface area (MM/PBSA) free energy method was evaluated for a DNA conformational change. The A-form to B-form DNA conformational free energy differences were calculated using MM/PBSA, and compared with free energy differences determined with a more rigorous umbrella

sampling method. MM/PBSA calculations of the free energy difference between A-form and B-form DNA are shown to be in very close agreement with the PMF result determined using an umbrella sampling approach. The sensitivity to solvent model and force field used during conformational sampling was also established for the MM/PBSA free energies.

In order to determine the influence of base flipping conformational changes on the MUG recognition process, PMF profiles were generated for each of the damaged bases (uracil, xanthine, oxanine, inosine). Agreement was displayed between the base pair stability trends from the umbrella sampling, and the enzyme activities from experiment. Interaction energies and structural analyses were used to examine the MUG enzyme, which revealed regions of the active site critical for binding xanthine and uracil substrates. Site-directed mutagenesis experiments were performed on MUG to determine the role of specific amino acids in the recognition mechanism. Mutations were studied further through modeling and molecular dynamics (MD) simulations of the unbound and bound proteins.

DEDICATION

For my Mom and Dad

ACKNOWLEDGMENTS

I would like to thank my advisor Dr. Brian N. Dominy. Throughout my graduate work he has been an incredible mentor and teacher. I thank him for constantly helping me improve as a scientist.

I also want to thank my collaborators Dr. Weiguo Cao and Dr. Hyun-wook Lee for including me in their projects. Our collaboration together has generated some interesting studies on glycosylase recognition, and I'm just grateful to be making a research contribution in the field of DNA repair.

The time I have spent in Dr. Steve Stuart's class has been invaluable, and I thank him for teaching me the fundamentals of computational chemistry.

My family's love and support has given me the energy and motivation required to finish this work. They've always inspired me to move forward, but also encouraged me slow down and appreciate the greatness in every moment.

TABLE OF CONTENTS

Page

Table of Contents (Continued)

Page

LIST OF TABLES

List of Tables (Continued)

Table Page

LIST OF FIGURES

List of Figures (Continued)

Figure Page

List of Figures (Continued)

Figure Page

List of Figures (Continued)

Figure Page

List of Figures (Continued)

Figure Page

List of Figures (Continued)

Figure Page

List of Figures (Continued)

Figure Page

List of Figures (Continued)

Figure Page

List of Figures (Continued)

Figure Page

List of Figures (Continued)

Figure Page

List of Figures (Continued)

Figure Page

List of Figures (Continued)

Figure Page

CHAPTER ONE

INTRODUCTION

Enzyme specificity is important for governing biological reaction pathways. Recognition is the ability of an enzyme to identify the preferred substrate for sequential catalysis. Most commonly, the active site or binding pocket within an enzyme dictates the specificity for a given substrate. The free energy of binding for enzyme-substrate association distinguishes specific from non-specific substrates.

In order to provide a complete picture of enzyme recognition, a Michaelis-Menten formulation will be used.[4] The Michaelis-Menten equation is applied to simple reaction schemes where an enzyme binds a substrate, and then accelerates a reaction forming a product in the process. Equation 1.1 displays the reaction scheme by which Michaelis-Menten kinetics follow. The first step is the formation of the enzyme-substrate (ES) complex, which is described by the Michaelis-Menten constant (K_M). The forward and reverse rate constants for the substrate binding are given by k_1 and k_{-1}. After the ES complex is formed, the enzyme catalyzes the formation of the products and the complex dissociates.

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_{cat}} E + P \qquad \text{Eq 1.1}$$

$$\frac{d[P]}{dt} = \frac{V_{max}[S]}{K_M + [S]} \qquad \text{Eq. 1.2}$$

The Michaelis-Menten equation (equation 1.2) calculates the overall rate of the catalysis, where V_{max} is the rate of the enzyme at maximum efficiency, and K_M can be defined as the concentration of substrate at half of V_{max}. These equations apply to simple enzyme systems when the steady-state approximation is true. Steady-state approximation assumes the concentration of an intermediate complex is unchanging over time, or within the time of catalysis. In general, binding affinities (K_D) are used to describe the recognition of an enzyme for a substrate. A pre-equilibrium assumption can be applied to simplify the rate equation. Since K_M is $(k_{-1} + k_{cat})/k_1$, and assuming the substrate dissociation is much faster than the enzyme catalysis ($k_{-1} >> k_{cat}$), K_M can be approximated as the ES dissociation constant (K_D).

$$\frac{d[P]}{dt} = k_{cat}[E]_0 \qquad \text{Eq 1.3}$$

$$\frac{d[P]}{dt} = \frac{k_{cat}}{K_M}[E]_0[S]_0 \qquad \text{Eq 1.4}$$

The impact of recognition on enzyme rate efficiency is demonstrated through equations 1.3 and 1.4. In the case of a very specific enzyme, when K_M is small, and the substrate and enzyme interact strongly (relatively low K_D), the enzyme-catalyzed rate is approximated as equation 1.3. For a specific enzyme-substrate complex, this implies the enzyme is working near or at peak efficiency (V_{max}). When the enzyme is promiscuous, and binds more than one substrate, the K_M will be larger, and the substrate(s)-enzyme

interactions will be weaker (relatively high K_D). Therefore, the enzyme-catalyzed rate is reduced to equation 1.4. In the case of promiscuous enzymes, the influence of K_M ($\approx K_D$) is observed in equation 1.4.

In molecular recognition studies, the free energy of binding is generated experimentally or theoretically in order to characterize the K_M described above. Experimental measurements of thermodynamic properties yield sufficient evidence for molecular recognition.[5] Computational methods can be employed to calculate free energy differences, and also have the benefit of analyzing specific electrostatic interactions within the active site of the enzyme. A common method for estimating the binding free energy is molecular mechanics Poisson-Boltzmann surface area (MM/PBSA).[6, 7] In chapter two, the approximate free energy calculation MM/PBSA is evaluated against the more rigorous umbrella sampling method. The free energy differences were calculated over a DNA conformational change using two different sequences of DNA. The effect of solvent model on the free energy calculation was also determined. Implicit and explicit solvent models were compared for the free energy differences of the DNA conformational change. Taking the comparison further, Feig *et al.* demonstrated that a generalized Born solvent model can be used in simulations of protein-DNA complexes. Structural properties of explicit solvent simulations were compared with those from simulations that used the implicit model.[8]

DNA binding proteins are essential for many processes, which include the recognition of specific DNA sequences. Theoretical methods have been applied to protein-DNA complexes, in order to determine mechanisms of specificity with molecular

detail. Nilsson and Mackerell provide a review of recent atomic simulation studies on protein-DNA complexes.[9] In enzyme reactions, where DNA is the substrate, careful consideration must be taken due to both enzymes and DNA helices being large dynamic molecules. Conformational equilibria of both the enzyme and DNA may influence the recognition and rate of catalysis.

The conformational diversity of the DNA helix is necessary for DNA to participate in several biological processes.[10] Many studies have been dedicated to investigating changes in DNA conformation.[11-13] Major and minor groove size vary with DNA helix transitions between conformations such as A-form, B-form, and Z-form. The compact A-form conformation has a much deeper major groove and a shallower minor groove[10] than the more elongated B-form conformation. Conformation of the DNA strongly depends on the electrostatic forces from the solvent environment, which is mostly due to the electrostatic repulsions of the phosphate backbone.[14] Conformational equilibria in DNA can be shifted simply by altering the salt concentration. Not only is the conformation of DNA dependent on the interactions from the phosphate backbone, but also the base pair sequence. It is well established that AT-rich sequences more B-like in conformation, and GC-rich sequences are more A-like in conformation.[15, 16]

The ability of enzymes to recognize specific sequences of duplex DNA has been the focus of several studies.[17, 18] Since each base pair has a distinct hydrogen bonding pattern in the major groove, in the past sequence specificity of DNA duplexes was demonstrated to be produced from hydrogen bonds with the base pairs in the major

groove.[19] It is not well known how the dynamic quality of both DNA and enzyme structures influenced the sequence recognition.[18] In a recent study by Rohs *et al*. [20] all of the protein-DNA complexes available in the RCSB protein database (1,031 complexes) were analyzed for correlations between minor groove width and DNA sequence. It was discovered that narrowing of the minor groove (<5Å width) provided specificity for DNA sequences. The AT-rich sequences resulted in narrower minor grooves, while GC-rich sequences resulted in larger width minor grooves. Poisson-Boltzmann[21] calculations were used to show the AT tracts in the narrower minor grooves had more negative potentials, which had a higher specificity for the amino acids at the DNA-protein interface. Particularly, arginine was found to occupy the narrower minor grooves 28% more than the other amino acids. It was concluded that most DNA binding enzymes specifically interact with the minor groove of the DNA duplex, and most of these minor groove interactions are with arginine. While most DNA binding enzymes utilize these minor groove interactions for specific sequence recognition, there are other enzymes that use them as non-specific interactions for searching or scanning the DNA.[22]

In 1964, a class of enzymes was discovered that evolved to prevent modifications of the coding information of cellular DNA.[23, 24] The base excision repair (BER) process is critical since even minor changes to an organism's DNA can be damaging to the whole organism. DNA base pairs are continuously subjected to exogenous and endogenous agents that result in deamination, oxidation, and alkylation. However,

genome integrity is maintained by BER with great efficiency. The BER process includes the recognition and removal of the damaged base, then the repair of the abasic site.

There are previous reviews that describe the chemistry of BER.[25, 26] A glycosylase enzyme initiates the process of BER through the hydrolysis of the N-glycosidic bond, which cleaves the damaged base. The resulting abasic site is repaired by an endonuclease. When these two steps are performed by separate enzymes, they are called monofunctional, however some DNA repair enzymes are capable of performing both steps of BER, and are referred to as bifunctional. It is not clear how glycosylase enzymes are capable of recognizing extremely rare damaged bases while searching the landscape of all the natural DNA bases of the genome.[27]

Glycosylase enzymes are unique from other DNA-binding enzymes due to the fact that they require the damaged base to be flipped out of the helix in order to cleave it from the DNA.[28] The recognition mechanism has not been determined for all of the glycosylases. There are some that recognize the damaged base in the helix, and others that recognize the base in the extrahelical state. In the former, the glycosylase is required to not only accelerate the hydrolysis of the N-glycosidic bond, but also the conformational change of the damaged base flipping out of the base stack. The recognition mechanism for uracil DNA glycosylases (UDG) has been investigated. It was demonstrated that family 1 of the UDG superfamily increased the lifetime of the uracil flipped-out state.[29] It is also known that UDG enzymes are active on single stranded DNA, which is unique among glycosylase enzymes.[2, 30, 31]

The last two chapters of this dissertation are dedicated to the recognition mechanism of monofunctional glycosylase enzymes. Chapter three covers the influence of electrostatics on base flipping conformational changes.[32] Free energy differences of base flipping were generated with umbrella sampling using implicit and explicit solvent models. Base flipping PMF profiles were constructed for the damaged bases uracil and xanthine. The electrostatic interactions of the flipping base are displayed, and show the significance of correctly representing solute-solvent interactions in the extrahelical state. We show that an implicit solvent model is not sufficient for modeling of base flipping conformational changes due to the lack of solute-solvent interactions in current models. After the DNA conformation studies, chapter four provides an examination of the binding equilibrium for *E.coli* mismatch-specific uracil DNA glycosylase (MUG) through active site interactions and the effects of several point mutations.[2] Mutations were studied to determine the roles that specific residues had in the recognition of base substrates within the MUG active site. Electrostatic interaction energies and structural analyses showed specific interactions that were critical for recognition in MUG. Overall, the recognition of the MUG enzyme has been thoroughly investigated by separating the recognition mechanism into the DNA conformational equilibrium, and the MUG binding equilibrium.

CHAPTER TWO

ROBUSTNESS OF THE MM/PBSA FREE ENERGY CALCULATION FOR DNA TRANSITIONS

This work has been published as:

Brice, A.R. and Dominy, B.N., *Analyzing the Robustness of the MM/PBSA Free Energy Calculation Method: Application to DNA Conformational Transitions*, Journal of Computational Chemistry, 32 (2011) 1431-1440.

Introduction

DNA adopts different conformations, which are necessary to sustain many different biological functions including transcription, translation, and replication.[10] The free energy difference between the A and B conformations of DNA is a crucial element in understanding many biological functions, including the binding equilibrium between DNA and other biomolecules.[33] Due to the importance of this subject, there have been many studies, experimental and theoretical, characterizing the conformational transitions of DNA.[11-13]

Two of the most prevalent and physiologically relevant conformational states in DNA are termed the A-form and B-form conformations. The B-form helical conformation of DNA is more elongated while the A-form structure is more compact with a larger helical diameter. The compact A-form conformation has a much deeper major groove and a shallower minor groove[10] than the more elongated B-form conformation. A number of geometric parameters distinguishes these two conformations

from one another. Pseudorotation of the deoxyribose is one such parameter, where the C3'-endo (A-form) and C2'-endo (B-form) conformations are distinct.[34] There are also base pair properties such as x-displacement, which is a measurement of the base pair deviation from the zy-plane, along the x-axis. The z-axis is defined as the helical axis, the y-axis is defined by the C1' – C1' virtual bond, and the x-axis is perpendicular to the zy-plane.[34] Structural variations between the A-form and B-form are primarily due to electrostatic effects resulting from repulsive interactions within the negatively charged phosphate backbone of DNA.[35] Electrostatic forces from surrounding environments are a significant factor in determining the conformational stability of NA systems.[14]

Conformational equilibria in DNA can be perturbed experimentally through manipulating the ionic strength of the solution, introducing different solvents or even introducing small molecules that preferentially bind to a conformational state of DNA. When DNA is in a low salt concentration aqueous solution, it will nearly always be observed in the B-form conformation. Depending on the sequence of the DNA, in a high salt environment it may display A-like properties or adopt the A-form conformation.[15, 36] Negatively charged phosphate groups on the backbone are more highly screened with added salt and reduce the significant electrostatic repulsion associated with the more compacted A-form.[37] A lower concentration of salt leads to more repulsion felt from the negatively charged phosphate groups, and a more elongated or B-like structure. In addition, specific enzymes, such as polymerases and endonucleases, are capable of binding to DNA, and shifting the equilibrium from B to A form conformations.[38] The

small molecules, neomycin, spermine, and hexaaminecobalt(III), have also been proven capable of converting B-form DNA into A-form DNA.[39]

In addition to influences arising from the phosphate backbone, DNA conformation is also sequence dependent. It has been found that A-T rich sequences are more likely to adopt the B-form conformation, and G-C rich sequences are more likely to adopt the A-form conformation.[15, 16] These sequence-dependent conformational preferences primarily arise from steric interactions between base steps.[40] As an example, cytosine bases are the least bulky, and tend to have minimal steric interactions with the deoxyribose ring, which have been found to allow transitions between the C3'-endo (A-form) and C2'-endo (B-form) conformations in the ribose sugar.[41]

Characterization of the conformational equilibria of DNA, as well as the effect of external perturbations, can be accomplished through computational free energy methods. Constructing a potential of mean force (PMF) with respect to some progress variable through the use of umbrella sampling[42] is a common approach for generating a free energy profile. Umbrella sampling is a method which employs a bias potential for the purpose of pulling the system from one thermodynamic state to another.[43, 44] The bias potential is based on an order parameter that defines a pathway between the two thermodynamic states of interest. As the order parameter varies, simulations are performed at each window or value of the order parameter resulting in a biased sampling of states within each window. The influence of the bias potential on the resulting sampling can be eliminated through the weighted histogram analysis method (WHAM)[45], leaving an unbiased distribution over the two endpoints from which the free

energy difference may be directly determined. This is a physically rigorous method, which is described within the framework of statistical mechanics.[46] The PMF approach can be very accurate when calculating free energy differences between two thermodynamic states.[47] However, in using umbrella sampling, the simulations corresponding to each window must be equilibrated. As a result, the total time of the free energy calculation is strongly dependent on the number of windows and the time required for each window to equilibrate.[44] For this reason, the approach can be computationally expensive.

In contrast to pathway methods, endpoint methods such as MM/PBSA[48] analysis estimate the free energy of the individual end-points and take the difference. Most commonly this approach is used to measure relative binding free energy[6, 7, 49, 50] and the two states are the bound and unbound states of a ligand-substrate complex. MM/PBSA estimates the enthalpic and entropic components of free energy through a post-analysis of conformational ensembles generated through molecular dynamics (MD) or Monte Carlo (MC) sampling.[51] Typically, an MD simulation is performed containing solute and solvent molecules, and free energies of the solute system are determined through post-analysis of solute trajectory snapshots using the same molecular mechanics force field combined with an implicit solvation model (Eq. 2.1).

$$G = \langle E_{MM} \rangle + \langle G_{PB} \rangle + \langle G_{SA} \rangle - T\langle S \rangle \qquad \text{Eq 2.1}$$

Free energy estimates from MM/PBSA are typically represented as a summation (Eq. 2.1) of the average gas-phase molecular mechanics energy (E_{MM}), the Poisson-Boltzmann polar solvation energy (G_{PB}), a surface-area dependent non-polar solvation energy (G_{SA}), and the solute entropy. E_{MM} (Eq. 2.2) is composed of intramolecular energies (E_{intra}) and intermolecular energies (E_{inter}). The intramolecular energy in a typical molecular mechanics potential (E_{MM}) accounts for bond stretching, bond angle bending, and dihedral angle rotations in the molecule of interest. Intermolecular energies are usually a summation of the nonbonded terms, which include the Coulomb and van der Waals energies. Both the Poisson-Boltzmann and GB implicit solvent models have been used in the past to generate the polar solvation free energy component, while the hydrophobic component is determined using an empirical surface area proportionate model.[48] Using normal mode analysis or quasi-harmonic analysis, the vibrational solute entropy (<S>) can be estimated.[6]

$$E_{MM} = E_{intra} + E_{inter} \qquad \text{Eq 2.2}$$

Endpoint methods are associated with both advantages and disadvantages. The primary advantages of the MMPB/SA endpoint method in determining free energies include both the modest computational complexity relative to pathway methods as well as the ability to trivially decompose the resultant free energy into various molecular mechanics and solvation terms. The advantages of calculation speed are obvious while the ability to decompose the free energy into various components can yield some

qualitative if not quantitative insight into the detailed nature of the thermodynamic process being studied. The disadvantages of the MMPB/SA method relate to its inaccuracy relative to pathway approaches.[52] Based on a thorough and rigorous description of the MMPB/SA method, McCammon and co-workers were able to demonstrate that while MMPB/SA was able to demonstrate reasonable accuracy in determining the binding free energy of 4-hydroxy-2butanone to FKBP12, the approach is particularly subject to challenges in determining changes in solute entropy.[6]

In order to further probe the utility of MMPB/SA methods in describing free energy trends in real systems, we have focused on conformational transitions in DNA. While the physiological importance of DNA conformational transitions is well-established, the literature suggests that MMPB/SA may be particularly well-suited to exploring these transitions. This suggestion is based on calculations that predict a relatively small conformational entropy change between different helical conformations of DNA.[48, 53] Using normal mode calculations within the AMBER force field, this was demonstrated in the case of A and B form DNA by Case and co-workers.[48] In a small dodecamer sequence of DNA, the conformational entropy change was found to be approximately -0.0017 kcal/mol K. This resulted in a contribution to the free energy difference ($-T\Delta S$ at 300K) of 0.5 kcal/mol, which favored the A-form conformation. Jayaram *et al.*[53], and showed that the entropy change between A and B-form DNA was a minor contribution. Quasiharmonic analysis was employed along with the AMBER force field in order to estimate the entropy difference between A-form and B-form DNA. In these circumstances where entropy contributions appear relatively insignificant in the

free energy change calculation (including relative binding free energy calculations between chemically similar systems), MM/PBSA can be a particularly appealing method.[54]

In addition to evaluating the validity of the final MM/PBSA calculated free energies, this work also explores the robustness of the calculated free energies to changes in the force field and the solvent model used within the conformational sampling engine. The influence of the force field on the free energy calculations results both from differences in the internal energy calculation performed on the sampled conformations, as well as differences in the conformational sampling itself. Studies performed recently on protein systems indicate that empirical force fields, demonstrate similar conformational sampling within the native basin.[55-57] Studies on smaller peptides and studies using enhanced sampling techniques indicate significant differences in conformational equilibria when comparing multiple empirical force fields.[58, 59] The impact of these effects on free energy calculations of DNA conformational equilibria are explored in the current study.

The conformational equilibria of biomolecules and nucleic acids (NA) in particular are also strongly influenced by the solvent environment, thus highlighting the importance of accurately modeling the solvent during free energy calculations.[14] Because systems constructed with explicit (atomistic) solvent models can easily increase the system size by a factor of 10-20, it is computationally very strenuous to simulate a large biomolecule over long timescales with an explicit solvent system.[60] Reducing the number of atoms and therefore degrees of freedom in the system greatly improves the

speed of these calculations, making accessible scientific questions involving larger solutes and longer timescales. Implicit solvent models offer that alternative. Implicit solvent models do not represent the solvent atomistically as in an explicit solvent model; rather the solvent is represented as a dielectric continuum.[61] Using the Poisson equation (Poisson-Boltzmann when including the influence of ionic strength) is a common way to estimate the polar solvation free energy of a system, however using traditional numerical solutions to the Poisson equation during dynamics generates a computational strain similar to that of an explicit solvent.[60] The Poisson equation can be solved analytically, however those solutions are typically restricted to simple geometric shapes.[21] There are several analytic generalized Born(GB) models that are optimized to reproduce Poisson solvation energies using rapidly solved parameterized equations [62-64]. 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. A TIP3P explicit solvent model and alternatively a generalized Born implicit solvent model will be used in generating conformational ensembles of the endpoint states to determine the influence of the solvent model on NA conformational free energies calculated using the MM/PBSA method.

In the present study, the robustness of the endpoint method MM/PBSA for calculating free energy changes associated with NA conformational transitions is evaluated. The reliability of the MM/PBSA method is evaluated by comparing DNA conformational free energy differences calculated using MM/PBSA to previously published calculations performed using pathway umbrella sampling.[65] The robustness

of the MM/PBSA method to variations in the force field is evaluated by comparing previously published calculations performed using the AMBER[66] force field with calculations performed using the CHARMM27[67] force field.[48] In addition, the robustness of MM/PBSA calculations are further examined by investigating the impact of conformational sampling performed upon explicitly and implicitly solvated solutes. These studies are performed on two previously examined DNA sequences: a hexamer $d(CTCGAG)_2$, and the Dickerson dodecamer $d(CGCGAATTCGCG)_2$.

Methods

In all simulations, the CHARMM c32b1 molecular mechanics package[68] and the CHARMM27[67] all-atom nucleic acid force field were used. Starting coordinates for the A-form and B-form conformations of the sequences d(CTCGAG)$_2$ and d(CGCGAATTCGCG)$_2$ were modeled with the program 3DNA.[34]

d(CTCGAG)$_2$

The initial A and B form hexamer structures generated in 3DNA[34] were minimized with the steepest descent algorithm in CHARMM for 500 steps. Simulation procedures from Roux *et al*.[65] were closely followed for the hexamer system in order to facilitate comparisons with earlier published work. Using the TIP3P water model, an explicit solvent system was constructed for this sequence within a rectilinear periodic box. Roux *et al*. used a periodic box with dimensions 43x34x38 Å^3, and the cylindrical restraint MMFP to reduce the translation and rotation of the DNA. The periodic box used in this study had dimensions of 55x55x55 Å^3, which required more simulation time, but allowed the DNA to freely diffuse eliminating the need for the MMFP restraints. The solvent system contained 5662 water molecules and 17374 total atoms. Long distance electrostatic interactions were accounted for using particle-mesh Ewald summation[69]. In order to achieve a neutral system necessary for the efficient calculation of long-range electrostatics using Ewald summation, 10 sodium ions were added. Since the sodium ions are expected to localize around the DNA molecule following equilibration, the 10 sodium ions were initially placed in close proximity to the DNA structure. A-form DNA

structures were restrained in the C1'-C2'-C3'-C4' torsion angle (υ_2) as done in previous studies [48, 70].

The Andersen thermostat[71] was used to heat the systems from 100 K to 300 K in increments of 10 K every 10 ps for a total of 200 ps. After the heating phase, the Andersen thermostat was used to equilibrate the system as a canonical ensemble (NVT) at 300K. The equilibration phase was carried out for 4 ns, and the production phase (used for further analysis) was run for an additional 5ns. A non-bonded cutoff starting at 9 Å with a shifting function was used to accelerate conformational sampling. The SHAKE constraint[72] was used on the hydrogen covalent bonds, which allowed for a 2 fs timestep during the heating, equilibration, and production phases.

The hexamer A form and B form structures were also simulated using a GB implicit solvent represented by the GBSW[73] algorithm in CHARMM. A salt concentration of 0 M ($\kappa = 0$), and smoothing length (sw) of 0.3 Å (default) were used. Non-bonded cutoffs of 30 Å with a switching function were implemented for the implicit solvent system. It was shown by Feig *et al.*[74] that MD simulations using GB with timesteps less than 2.0 fs yielded minimal energy drift during MD simulations, and for this reason a timestep of 1 fs was used for the GB simulations. Simulation times were the same as those described above for the explicitly solvated systems.

d(CGCGAATTCGCG)$_2$

Procedures from Srinivasan *et al.*[48] and Feig *et al.*[75] were closely followed for this system in order to facilitate comparisons to these previous works. The B and A form of

the Dickerson dodecamer[76] were simulated in a GB implicit solvent represented by the GBSW algorithm in CHARMM, and also in a TIP3P explicit solvent. A switching function was employed for the non-bonded cutoffs. All A-form DNA structures were sampled by restraining the C1'-C2'-C3'-C4' torsion angle as described in reference 21. Srinivasan *et al*.[48] used a periodic water box of dimensions 70x45x45 $Å^3$, and containing ~4000 water molecules. In the current study, a periodic water box of initially the same dimensions, containing 4100 water molecules was employed for both the B-form and A-form dodecamer. Long distance electrostatic interactions were accounted for using a particle-mesh Ewald summation[69]. A neutral system was achieved by adding 22 sodium ions to the system. Along with the restraining of all covalent hydrogen bonds by SHAKE, a 2 fs timestep was used in the TIP3P simulations. The non-bonded cutoffs were set to 12 Å in the TIP3P simulations. Starting structures of both solvent systems were minimized with the steepest descent algorithm for 100 steps. The system was heated from 200 K to 300 K in increments of 10 K every 10 ps for a total of 100 ps, and using the Langevin barostat[77] an isothermal-isobaric ensemble (NPT) was constructed for the equilibration. The system was then equilibrated for 4 ns as an NVT ensemble, followed by a production phase of 5 ns.

With regard to the simulations in implicit solvent, the GBSW algorithm was used with a smoothing length (sw) of 0.6 Å along with 12 Å non-bonded cutoff for the GBSW systems. The GB simulations were heated from 200K to 300K in increments of 10 K every 10 ps for a total of 100 ps with the Andersen thermostat. Simulation times for the

heating, equilibration, and production phases were the same as those described above for the explicitly solvated systems.

MM/PBSA

Conformational free energy differences were calculated for both the hexamer and dodecamer DNA sequences using the MM/PBSA method. Snapshots for MM/PBSA analysis were extracted from the relevant molecular dynamics trajectories every 5 ps for 500 ps. The averages were not carried out over a larger portion of the 5 ns production phase so that annealed B-form trajectories were being compared to annealed A-form trajectories when the free energy difference was calculated. During the endpoint simulations that were carried out in this study, the ends of the DNA naturally fray and in some cases re-anneal. Since the DNA did not fray during the umbrella sampling study performed by Roux and coworkers[65], it was important to analyze portions of the trajectory representing annealed DNA structures in order to compare corresponding states. While short, the length of the trajectory analyzed in this study (500 ps) is the same as that analyzed in an earlier MMPB/SA study of the Dickerson dodecamer sequence[76]. The average potential energy calculated using the CHARMM27 vacuum potential was determined over the 100 snapshots extracted from the 500 ps trajectory. Poisson-Boltzmann polar solvation free energies and surface areas were also determined from this conformational ensemble. The linear form of the Poisson-Boltzmann equation was solved within CHARMM using the PBEQ module to estimate the polar solvation energies. Srinivasan *et al*. demonstrated that a change in grid width from 0.5 Å to 0.2 Å

changed the solvation energy (G_{PB}) by 10 kcal/mol (or approximately 0.2%).[48] Further, the change in solvation energy was approximately the same in both the A and B form, resulting in a cancellation of error in determining the free energy of the conformational transition. The conformational free energy difference that was desired in the current study was therefore not greatly dependent on the grid width, and when solving for the Poisson-Boltzmann solvation energy, a grid cell width of 0.5 Å was used. The external dielectric constant was set to 80, and the internal dielectric constant was set to 1. A molecular surface (solvent accessible) was constructed for the solvent-solute dielectric boundary with a 1.4 Å radius probe. The sum of these averaged values is the free energy calculated by MM/PBSA ($G = \langle E_{MM} \rangle + \langle G_{PB} \rangle + \langle G_{SA} \rangle$). Vibrational entropic contributions were calculated through normal mode analyses of the same 100 snapshot conformational ensemble. Snapshots were minimized using the adopted Newton-Raphson algorithm under a distance dependent dielectric solvent environment with an ε coefficient of 4.0 under successively reduced harmonic restraints starting with a force constant of 5 kcal/mol Å and reducing in steps of 1 kcal/mol Å to a final force constant of 0 kcal/mol Å.

Results and Discussion

The following results are separated into three distinct sections for clarity. First, the results from the endpoint MM/PBSA analysis of the hexamer A form to B form conformational transition are described and compared to the corresponding results obtained using an umbrella sampling PMF approach. Next, the results of the MM/PBSA analysis of the Dickerson dodecamer A to B form conformational transition are discussed and compared to corresponding results obtained using an MM/PBSA analysis within the AMBER PARM94 force field. Finally, results obtained using ensembles derived from implicit solvent MD simulations are described and compared to MM/PBSA calculations performed on trajectories derived from explicitly solvated systems. Through these comparisons, questions involving the accuracy and robustness of the MM/PBSA approach for studying nucleic acid conformational equilibria are addressed.

MM/PBSA to PMF Comparison

	Hexamer MM/PBSA ΔΔGA→B (kcal/mol)					
	ΔG_B	Σ	ΔG_A	σ	$\Delta\Delta G_{A\rightarrow B}$	σ
<E_{coul}>	-488.7	2.2	-489.4	2.1	0.7	3.0
<E_{vdw}>	21.5	0.9	17.5	1.0	4.0	1.3
<E_{intern}>	820.6	1.5	851.9	1.8	-31.3	2.3
<E_{MM}>	353.4	3.0	380.0	3.0	-26.6	4.2
<E_{sa}>	15.9	0.0	15.1	0.0	0.8	0.0
<E_{pb}>	-1866.5	2.0	-1882.4	1.9	15.9	2.8
<E_{tot}>	-1497.2	1.8	-1487.3	2.0	-9.9	2.7
<E_{tot}> (0.1M)	-1508.4	1.9	-1498.5	2.0	-9.9	2.8
<E_{tot}> (0.3M)	-1511.6	1.9	-1501.8	2.0	-9.8	2.8
-T<S>	-309.5	0.4	-308.6	0.1	-0.9	0.4
<ΔG>	-1806.7	1.9	-1795.9	2.0	-10.8	2.8
<ΔG> (0.1M)	-1817.9	1.9	-1807.1	2.0	-10.8	2.8
<ΔG> (0.3M)	-1821.1	1.9	-1810.4	2.0	-10.7	2,8

Table 2.1 – Columns labeled as σ represent the standard error of the calculated free energies provided in the preceding column. Standard errors in the means are calculated as s/sqrt(N), where s is the standard deviation. Summation of individual energy components (<E_{tot}> - T<S>) yields the free energies, represented as <G>.

The free energy difference of the conformational transition, determined using the MM/PBSA method applied to TIP3P solvated simulations of the hexamer DNA, (-10.8 kcal/mol, table 2.1) is very comparable to the free energy difference found by Roux *et al*. (-13.5 kcal/mol) under corresponding simulation conditions.[65] The MM/PBSA analysis also quickly estimates the impact of ionic strength on the conformational equilibria of DNA. Consistent with experiment and as expected, the hexamer demonstrates a

preference for B-form over A-form in a low salt concentration.[78] When the ionic strength is increased, the free energy differences indicate a slight shift towards the A-form.

A similar assessment of the accuracy of the MM/PBSA approach in comparison to a pathway sampled approach has been recently described for applications to protein / small molecule binding free energies. Lee and Olson *et al*.[52] have compared MM/PBSA and PMF methods for calculating the absolute binding free energy of protein-ligand complexes. Using CHARMM, they examined binding free energies of the FKBP protein bound to 4-hydroxy-2-butanone (BUQ) and FK506. It was shown that MM/PBSA calculations did not accurately reproduce free energy differences obtained using umbrella sampling with either a GBSA solvent model or a hybrid solvent model. Using the GBSA[79] implicit solvent model and the multiple simulation method on the BUQ ligand, MM/PBSA results deviated from the umbrella sampled results by approximately 23% (or 0.9 kcal/mol), and approximately 29% (or 3.3 kcal/mol) with respect to the FK506 ligand. Similarly, in the current study, MM/PBSA calculations of the DNA conformational equilibrium for the hexamer deviated from the PMF results (Roux *et al*.[65]) by approximately 20%. When comparing to experimental values, Lee and Olson's absolute binding energies for protein-ligand complexes yielded more uncertainty in the MM/PBSA calculations than those binding energies determined with the umbrella sampling approach.

To further explain the small dissimilarities in free energy differences calculated using MM/PBSA and umbrella sampling, the vibrational entropy contributions are

examined (Table 2.1). When harmonic estimates of the vibrational entropy are neglected, the predicted conformational transition free energies from MM/PBSA suggest a relatively less stable B-form conformation, while still qualitatively reproducing the umbrella sampling result (Table 2.1). While other analyses of A and B form DNA conformational equilibria have found it appropriate to neglect conformational entropy changes,[80] within the context of the CHARMM force field, the entropic contribution does not appear to be insignificant. The flexibility of the B-form conformation is considerably greater than the A-form conformation, which is the reason that the changes in entropy favor the B-form in this study. It has been shown by Zakrzewska and coworker[81] that the B-form conformation of DNA populates more modes of low frequencies than either the A-form or Z-form conformations of DNA. The low frequency modes are more populated for the B-form conformation than for the A-form conformation, and therefore significant changes in entropy should not be unexpected. The MM/PBSA method calculates the enthalpy and entropy contributions separately for the purpose of generating a free energy, however calculating accurate vibrational entropy can be difficult.[6] This may be the cause of discrepancies between the MM/PBSA and umbrella sampling methods.

In addition to the difficulty in calculating entropy changes, another factor that should be considered is the order parameter used to characterize the A and B form DNA conformations. The order parameter defines the end states from which the calculated free energy difference is based. Roux *et al*. demonstrated that using different order parameters produced differences in the calculated free energy difference.[65] Based on the demonstrated sensitivity of the calculated conformational free energy differences toward

the order parameter, it is suggested that relatively small differences in the A or B ensemble generated during the umbrella sampling or the MMPB/SA analysis could account for a substantial portion of the observed free energy differences calculated using these two methods. Similar challenges in defining the end states during umbrella sampling have been noted in determining absolute binding free energies.[52] In order to address the possibility that the choice of order parameter may affect the comparison between MM/PBSA and the umbrella sampling methods, a structural analysis was performed on the A and B form ensembles.

In an effort to analyze the origin of the small difference between the free energies calculated using the MM/PBSA endpoint method and the umbrella sampling pathway approach, conformational analysis of the hexamer trajectories were compared with results from Roux *et al*.[65] The DNA structural parameter z_p was used by Roux *et al*. to identify the two end states, A-form and B-form DNA, of the PMF profile and determine a free energy difference between those two states. A z_p greater than 1.5 Å was described as an A-form structure, while any z_p lower than 1.5 Å was described as a B-form structure. This classification of the A-form and B-form conformations was found to be consistent with the z_p values for the hexamer structures generated in the current study. The z_p of the A-form hexamer structures averages 2.0 Å (σ=0.23 Å) for the trajectory, and the z_p of the B-form hexamer structures averages 0 Å (σ=0.32 Å) for the trajectory. This implies that for the comparison to Roux *et al.*[65], similar structures are being analyzed as the end-points of the B to A transition in the current study. However, structural differences that

are not captured by this crude analysis of the z_p parameter may help explain the small differences in the free energies calculated using the two methods.

Force Field Comparison

	Dodecamer MM/PBSA ΔΔGA→B (kcal/mol)					
	ΔG_B	σ	ΔG_A	σ	$\Delta\Delta G_{A\rightarrow B}$	σ
$\langle E_{coul} \rangle$	1305.6	4.3	1393.1	5.6	-87.5	7.1
$\langle E_{vdw} \rangle$	-4.9	1.3	0.4	1.2	-5.3	1.8
$\langle E_{intern} \rangle$	1587.7	2.2	1621.0	2.0	-33.3	3.0
$\langle E_{MM} \rangle$	2888.5	4.6	3014.5	5.8	-126.0	7.4
$\langle E_{sa} \rangle$	27.0	<0.05	26.2	<0.05	0.8	0.0
$\langle E_{pb} \rangle$	-6098.1	3.8	-6216.1	5.3	118.0	6.5
$\langle E_{tot} \rangle$	-3182.6	1.9	-3175.4	2.0	-7.2	2.8
$\langle E_{tot} \rangle$ (0.1M)	-3228.9	1.9	-3224.3	1.9	-4.6	2.7
-T<S>	-608.5	0.7	-601.3	0.2	-6.9	0.7
<ΔG>	-3790.8	1.9	-3776.7	2.0	-14.1	2.8
<ΔG> (0.1M)	-3837.1	1.9	-3825.6	1.9	-11.5	2.7

Table 2.2 – Standard errors (σ) in the means are calculated as s/sqrt(N), where s is the standard deviation. Summation of individual energy components ($\langle E_{tot} \rangle$ - T<ΔS>) yields the free energies, represented as <ΔG>.

In order to further investigate the robustness of the MMPB/SA free energy calculation method, comparisons are made between DNA conformational free energy differences calculated within the CHARMM force field and those calculated within the AMBER force field.[48] Free energies calculated using the MM/PBSA method within the CHARMM27 force field for the Dickerson dodecamer DNA sequence are found in table

2.2. With 0 M salt, the A→B free energy difference calculated within the AMBER force field was determined to be -13.0 kcal/mol[48], and within the CHARMM27 force field it was determined to be -14.1 kcal/mol. The conformational free energy differences between the two force fields are qualitatively and even quantitatively similar, however there is an obvious distinction between the relative contributions of the MM/PBSA and vibrational entropy components of the force fields. MM/PBSA calculations in the CHARMM force field favor the B-form by 7.2 kcal/mol, while in the AMBER force field, the B-form is favored by 13.0 kcal/mol. On the other hand, while vibrational entropy contributions to the conformational equilibria are negligible in AMBER, they favor the B-form conformation by 6.9 kcal/mol in CHARMM.

This discrepancy displayed in the enthalpic and entropic contributions to the free energies of the two force fields is examined further. The AMBER results demonstrate that most of the difference in the conformational free energy change results from the molecular mechanics and solvation energy terms, while very little contribution arises due to the vibrational entropy. Alternatively, the CHARMM results indicate that the configurational preference for B form arises roughly equally from the MM/PBSA term and the vibrational entropy component. One explanation for this difference may be linked to a form of enthalpy-entropy compensation. An example of this was shown in a recent study that calculated free energies of solvation for methane in explicit water-tert-butanol solvent systems.[82] Results were collected over several different mole fractions of the solvent mixtures, and the entropy and enthalpy demonstrate compensating effects over these different solvent systems.[82]

Small perturbations on the Hamiltonian may lead to similar free energy estimates with distinct and compensating contributions from the entropy and enthalpy terms.[83] The CHARMM and AMBER force fields were parameterized differently, and these differences are particularly reflected in the partial charge parameters[66, 68]. In this case, the perturbation on the system is the variation of force field partial charges, which creates dissimilar molecular mechanics energies. Coulombic interactions are more favorable with the AMBER partial charges (-220.5 kcal/mol)[48] than with CHARMM partial charges (-87.5 kcal/mol). This allows greater flexibility in the CHARMM trajectories, and yields a larger vibrational entropy change.

In order to validate the premise of entropy-enthalpy compensation arising from differences in force field parameterizations, a model calculation was performed using the B-form structure of the Dickerson dodecamer. In this model, partial charges associated with the base pair hydrogen bonds were randomly perturbed by 1% on average. Potential energies and harmonic vibrational entropies were calculated for the B-form structure of the dodecamer. Changes in both the potential energy and the vibrational entropy were determined relative to results obtained using the original CHARMM 27 charge parameters. Changes in the potential energy and the vibrational entropy contribution to free energy ($-T\Delta S$) are found to be strongly negatively correlated (figure 2.1). Results from this model calculation suggest that enthalpy-entropy compensation can arise between different force fields, potentially resulting in greater agreement in calculated free energies but weaker agreement in calculated potential energies or entropies.

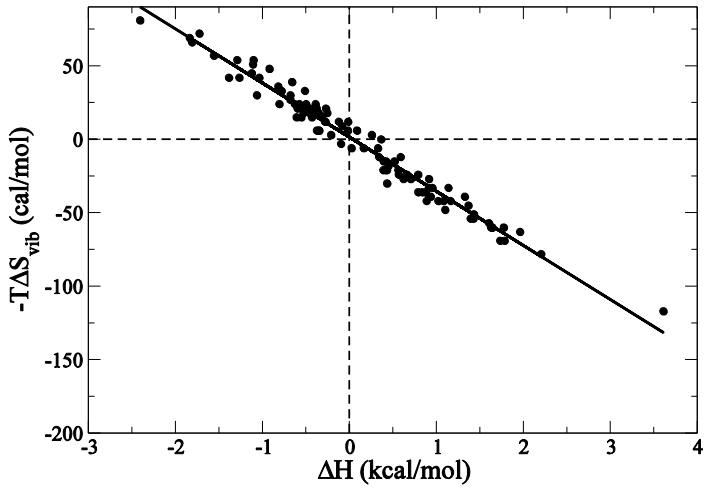


figure 2.1 – Entropy-enthalpy compensation is demonstrated with the CHARMM27 force field by randomly varying H-bond donor and acceptor partial charges on the dodecamer. The calculated molecular mechanics energies (ΔH) and vibrational entropy contributions (-TΔS) from normal mode analysis are plotted to show the entropy-enthalpy relationship.

Contributions of different energy terms

In order to further understand the differences between the CHARMM and AMBER results, the differences in the molecular mechanics energies and solvation energies are examined in more detail. The Coulomb energy differences contributing to the A form / B form equilibria vary by 133.0 kcal/mol between the two force fields, while the van der Waals, hydrophobic solvation and bonded terms differ by 7.8, 0.4 and 12.9 kcal/mol respectively. Considering the anti-correlated Coulomb and Poisson-Boltzmann solvation terms together as the electrostatic contribution to the A/B equilibria, the differences between CHARMM and AMBER amount to 26.1 kcal/mol. As expected, this suggests that electrostatics play a significant (if not dominant) role in determining the

nucleic acid conformational equilibria[37] in each of these force fields. The large positive Coulomb energies observed in both the A and B form conformations of DNA in both the CHARMM and AMBER analyses suggest that the electrostatic influences on conformational preference arise substantially from repulsion in the phosphodiester backbone of DNA.

Due to the significance of the DNA backbone on the conformational equilibrium,[37] focus is placed on the partial charge parameters surrounding the phosphates within the two force fields. In the AMBER PARM94 force field, P = 1.1659, O3' = -0.5232, OP (2) = -0.7761, and O5' = -0.4954 .[66] Whereas in the CHARMM27 force field, P = 1.50, O3' = -0.57, OP (2) = -0.78, O5' = -0.57, and the CH_2 groups 3' and 5' to the phosphate group each have a partial charge of 0.1.[67] While these partial charges (which include the phosphate group and bonded, neighboring atoms) in both force fields sum to a net charge of -1.0, differences in how the partial charges are distributed within the phosphate and neighboring atoms can result in significant changes to the electrostatic energy dominated by short range interactions.

Since electrostatic interactions have a large role in the conformational changes of DNA,[37] it is not surprising that these same conformational changes are significantly influenced by the environmental salt concentration. The increased ionic strength enhances the screening of repulsive phosphate-phosphate interactions, facilitating conformational transitions that reduce the inter-phosphate distance (e.g. the B-form to A-form conformational transition). As a result of the role played by ionic strength in the conformational equilibria of nucleic acids, the effect of ionic strength on the A-form to

B-form free energy change is assessed and compared to results obtained with the AMBER force field.

When the ionic strength is increased during the MM/PBSA calculation, the resulting conformational free energy differences also show similar trends for both the CHARMM and AMBER force fields. With a 0.1M salt concentration, the A→B free energy difference calculated within the AMBER force field is determined to be -10.9 kcal/mol,[48] and within the CHARMM force field is determined to be -11.5 kcal/mol (table 2.2). By raising the salt concentration from 0M to 0.1M, the AMBER free energy increases or stabilizes the A-form by 2.1 kcal/mol. The same addition of salt in CHARMM, stabilizes the dodecamer A-form by 2.6 kcal/mol. While the electrostatic solvation free energy terms (PB) calculated for the A and B forms of DNA differ in the two force fields by ~39% and the Coulomb energies differ by over 100%, the change in PB energies due to the addition of salt (or the salt effects) differ by only 20%. When analyzing the impact of the energy components on the conformational equilibria ($\Delta E_{A->B}$), the Coulomb effects differ by ~60% between CHARMM and AMBER and the PB energy contributions differ by 48% while the influence of ionic strength differs by only ~24%. This suggests that the salt effects are less sensitive to changes in the force field, or specifically the partial charge model, than the Coulomb or PB solvation energy terms.

Poisson-Boltzmann ΔG_{solv} for GC sequence			
Salt Conc (M)	**6bp (kcal/mol)**	**12bp (kcal/mol)**	**18bp (kcal/mol)**
0	-1927.05	-6325.6	-11969.3
0.05	-1936.33	-6365.3	-12053.4
0.1	-1938.43	-6372.8	-12067.1
Range	-11.38	-47.2	-97.8

Table 2.3 – Polar solvation energies of a GC DNA sequence calculated with the Poisson-Boltzmann equation at different salt concentrations.

Increasing the salt concentration from 0M to 0.1M had different impacts on the hexamer sequence and the dodecamer conformational equilibria, but similar effects were observed between the dodecamer in CHARMM and the dodecamer in AMBER. This indicated that the effect of ionic strength on conformational equilibria was either sequence or length dependent. When the salt was added to a GC hexamer sequence, the free energies of the individual conformations decreased by approximately 11-12 kcal/mol (table 2.3), which is similar to the hexamer result in table 2.1. After the salt addition to a GC dodecamer sequence, the free energies decreased by approximately 40-50 kcal/mol, which resembles the dodecamer result in table 2.2. This demonstrates that the impact from the addition of salt is not primarily dependent on the sequence, but rather on the length of the DNA strand. In comparing the Dickerson dodecamer sequence to the alternating GC dodecamer sequence, the number of hydrogen bonds is greater for the latter. The salt effect also displays little dependence on the number of hydrogen bonds. All of this indicates that, within the context of this model, the impact of ionic strength is

based on the screening of phosphate charges. Furthermore, the robustness of the calculated impact of ionic strength on conformational equilibria across the CHARMM and AMBER force fields is a result of the preferential screening of long-range interactions between the phosphate groups.[84]

Due to the differences in partial charges, the Coulombic energy contributions to free energy are inconsistent between the two force fields, however the force fields do yield similar sensitivities to ionic strength. It can be concluded from the similar trend in the CHARMM and AMBER salt effect that the salt screening is dependent on the long-range phosphate interactions of the DNA backbone. The partial charges of the atoms within the phosphate group are different between the two force fields, but the overall net charge (-1.0) on the phosphate groups are equivalent. Diverging results observed in the Coulombic energies are due to the short-range interactions between the varying atomic partial charges. Salt concentration and ionic strength are represented in the Poisson-Boltzmann equation with the inverse Debye length (κ), which increases as the interaction distance between the charges increases.[85] Therefore, as the chain length, or number of base pairs (bp) in a DNA sequence increases, the salt screening of the negative phosphate backbone will also increase.

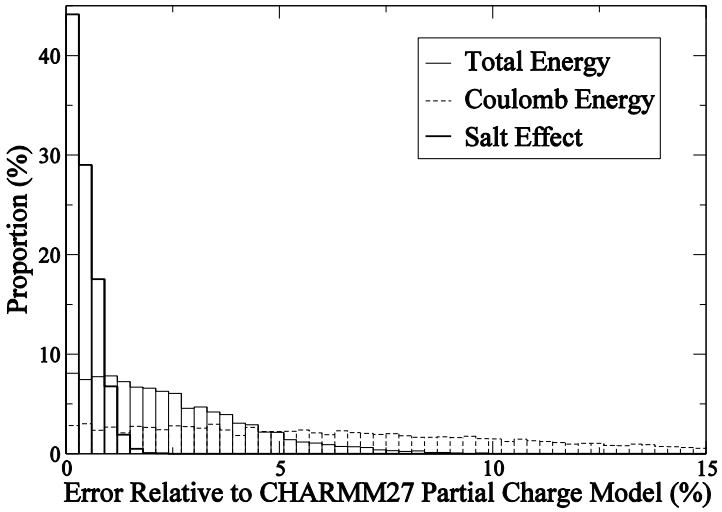


Figure 2.2 – Perturbing the partial charges of the ataoms within the phosphate groups, and keeping an overall net charge of -1.0, the small change in salt effect is demonstrated. As the CHARMM27 partial charges are altered, the variation from the CHARMM potential energy, Coulomb energy and salt effect are shown with the relative error.

A proof of principle calculation is provided to demonstrate that while differences in the partial charge model can result in relatively large differences in the corresponding potential energy, the impact on the calculated salt effect is relatively small. In this model calculation, the CHARMM27 partial charges for atom types associated with the phosphate groups in the Dickerson dodecamer are randomly perturbed according to a Gaussian distribution with a mean corresponding to 0% change and a standard deviation corresponding to a 10% change. Each randomly perturbed partial charge model is generated to preserve the net charge of -1.0 for the phosphate group in order to realistically represent differences in charge models seen in different available force fields. The results demonstrate that the fractional change in the salt effect is typically much

smaller than the change observed in either the total potential energy or Coulomb energy for a given partial charge model (figure 2.2). This suggests that the reason for the close agreement in the salt effects between the CHARMM and AMBER force fields is a result of the physical nature of the salt effect that efficiently screens long range electrostatic interactions.

Solvent Model Comparison

An evaluation of the solvent model implemented during conformational sampling is performed for the purpose of further assessing the reliability of the MM/PBSA free energy calculation. The MM/PBSA free energy results (table 2.4) of ensembles generated with a generalized born implicit solvent model (GBSW) are compared with those generated in an explicit solvent model (TIP3P).

Hexamer GBSW $\Delta G_{A\rightarrow B}$ (kcal/mol)						
	ΔG_B	σ	ΔG_A	σ	$\Delta\Delta G_{A\rightarrow B}$	σ
$\langle E_{coul}\rangle$	-510.8	1.9	-497.7	1.9	-13.1	2.7
$\langle E_{vdw}\rangle$	11.5	0.9	12.5	0.9	-1.0	1.3
$\langle E_{intern}\rangle$	810.4	1.9	833.1	1.7	-22.7	2.5
$\langle E_{MM}\rangle$	311.1	2.3	347.9	2.3	-36.8	3.3
$\langle E_{sa}\rangle$	14.9	<0.05	14.6	<0.05	0.3	0.0
$\langle E_{pb}\rangle$	-1844.0	1.4	-1874.1	1.6	30.1	2.1
$\langle E_{tot}\rangle$	-1517.9	1.8	-1511.6	1.7	-6.3	2.5
0.1M	-1529.2	1.8	-1522.8	1.7	-6.4	2.5
$-T\langle S\rangle$	-311.1	0.4	-308.7	0.2	-2.4	0.4
$\langle\Delta G\rangle$	-1829.0	1.8	-1820.3	1.7	-8.7	2.5
$\langle\Delta G\rangle$ (0.1M)	-1840.3	1.8	-1831.5	1.7	-8.8	2.5
Dodecamer GBSW $\Delta G_{A\rightarrow B}$ (kcal/mol)						
	ΔG_B	σ	ΔG_A	σ	$\Delta\Delta G_{A\rightarrow B}$	σ
$\langle E_{coul}\rangle$	1369.1	3.0	1495.8	3.6	-126.7	4.7
$\langle E_{vdw}\rangle$	-12.7	1.4	-12.4	1.3	0.3	1.9
$\langle E_{intern}\rangle$	1635.1	2.5	1683.0	2.6	-47.9	3.6
$\langle E_{MM}\rangle$	2991.5	3.6	3166.4	4.4	-174.9	5.7
$\langle E_{sa}\rangle$	26.3	<0.05	25.6	<0.05	0.7	0.0
$\langle E_{pb}\rangle$	-6156.1	2.6	-6322.4	3.1	166.3	4.0
$\langle E_{tot}\rangle$	-3138.3	2.4	-3130.4	2.4	-7.9	3.4
0.1M	-3185.3	2.4	-3180.0	2.4	-5.3	3.4
$-T\langle S\rangle$	-604.1	0.3	-599.8	0.7	-4.3	0.8
$\langle\Delta G\rangle$	-3742.4	2.4	-3730.2	2.5	-12.2	3.5
$\langle\Delta G\rangle$ (0.1M)	-3789.4	2.4	-3779.8	2.5	-9.6	3.5

Table 2.4 – Free energy differences of the hexamer and dodecamer calculated with the MM/PBSA method within the CHARMM27 force field. MD trajectories of the two sequences sampled with GBSW solvent model. Summation of individual energy components ($\langle E_{tot}\rangle - T\langle\Delta S\rangle$) yields the free energies, represented as $\langle\Delta G\rangle$.

MM/PBSA analyses of simulations performed on the DNA hexamer sequence within a TIP3P explicit solvent model and a GB implicit solvent model demonstrate some differences. The A to B free energy difference of the hexamer is -8.7 kcal/mol (table 2.4) when simulated in implicit solvent, which is qualitatively similar to the free energy

difference of -10.8 kcal/mol (table 2.1) when the hexamer is simulated in explicit solvent. Looking closer at the individual components of the conformational free energy difference in both solvent models, the components largest in magnitude are the Coulomb, Poisson-Boltzmann, and bonded energy terms. These terms also demonstrate the largest differences when comparing the analysis of the implicit and explicit solvent ensembles. As expected, the differences observed in the Coulomb energy are largely canceled by differences seen in the Poisson-Boltzmann energy term. This is due to the anti-correlated nature of these two energy terms as a result of the cross-polarization screening term in the polar solvation energy. This cancellation results in overall better agreement in the conformational free energy differences based on ensembles derived from implicitly solvated and explicitly solvated simulations. The remaining difference of ~2 kcal/mol is largely attributed to the bonded energy terms.

In analyzing the energy terms derived for the dodecamer simulated in implicit and explicit solvent, some differences are also apparent. The A to B free energy difference of the dodecamer is -12.2 kcal/mol (table 2.4) when simulated in the implicit solvent and -14.1 kcal/mol (table 2.2) in the explicit solvent, demonstrating a stronger preference for the B form DNA ensemble generated in the TIP3P explicit solvent. This agrees with the trend observed for the hexamer where the implicit solvent ensembles demonstrated a weaker preference for the B form conformation. The solute entropy component contributes the most in this comparison between free energy differences of the implicitly and explicitly solvated dodecamer. Pseudorotation analysis (figure 2.3) of the DNA backbone sugar shows that dodecamer trajectories generated in the TIP3P explicit solvent

exhibited a greater degree of conformational fluctuation than the trajectories generated in the GB implicit solvent. This difference in conformational fluctuation for the dodecamer trajectories is consistent with the 2.6 kcal/mol variation between the vibrational entropy components of the implicit and explicit solvent models.

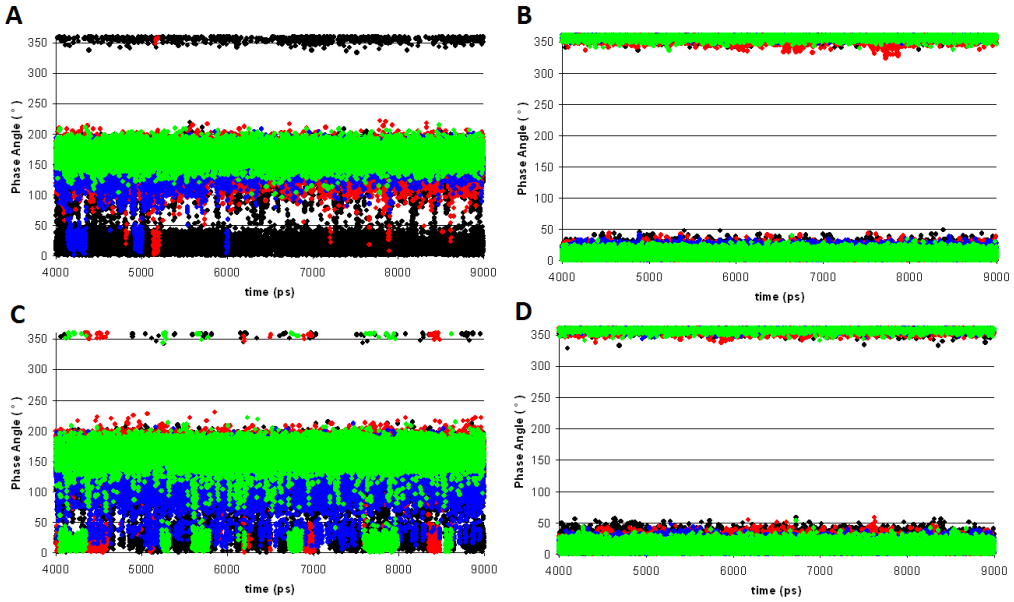


Figure 2.3 – Pseudorotation of the DNA deoxyribose, for 5000 ps (4 ns – 9ns) of dodecamer MD trajectories. The black points are cytosine nucleotides, green points are adenine nucleotides, red points are guanine nucleotides, and blue points are thymine nucleotides. A) unrestrained B-form dodecamer in GB implicit solvent B) restrained A-form dodecamer in GB implicit solvent C) unrestrained B-form dodecamer in TIP3P explicit solvent D) restrained A-form dodecamer in TIP3P explicit solvent.

In this example of the Dickerson dodecamer, the choice of solvent model appears to have a greater impact on the calculated conformational free energy differences than the choice of force field.

There is very close agreement for both the hexamer and dodecamer between the solvent models with respect to the salt effect analysis. The sensitivity of the DNA dodecamer conformational equilibria to ionic strength is in qualitative agreement (19%) between the CHARMM and AMBER results. This sensitivity to ionic strength is equivalent for the ensembles generated with explicit and implicit solvent models. Therefore, the influence of ionic strength on the A to B conformational transition free energy is more strongly perturbed by changes in the force field, rather than changes in the solvent model.

With significant differences in the free energies of the hexamer and the dodecamer suggesting differences in the conformational ensembles generated in implicit and explicit solvent, a detailed conformational analysis was indicated. Time resolved all-atom rmsd measurements, and pseudorotation measurements indicated structures consistent with A-form and B-form structures (figure 2.3). In the pseudorotation analysis, the unrestrained B-form simulations show fluctuations between the C2'-form (phase angle 160°) and C3'-endo (phase angle 0°-30°) conformations of the ribose sugar.[34] These fluctuations are greater in ensembles generated with the explicit solvent than with the implicit solvent.

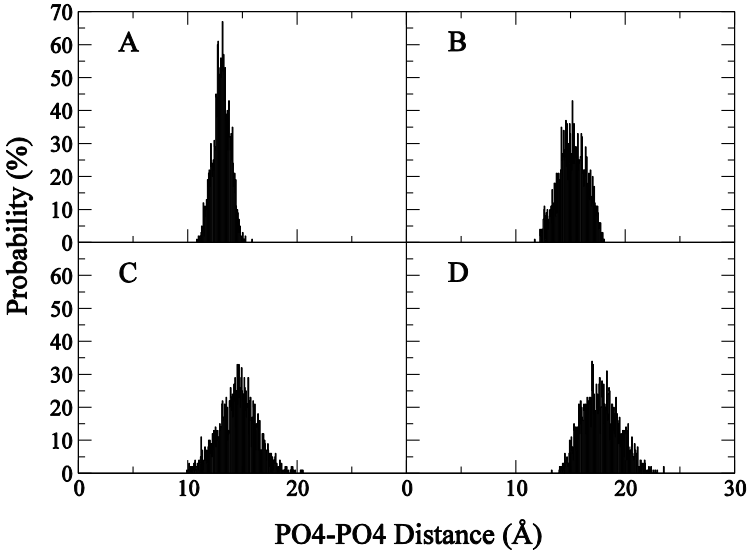


Figure 2.4 - Phosphate to phosphate distances across dodecamer major grooves. Distances between major groove phosphate groups are calculated and averaged over A.) A-form DNA structures sampled with the GB model, B.) B-form DNA structures sampled with the GB model, C.) A-form DNA structures sampled with the TIP3P model, and D.) B-form DNA structures sampled with the TIP3P model.

In order to probe directly the structural features likely responsible for the discrepancies between the solvent models, phosphate-phosphate distances were measured for the GB and TIP3P dodecamer trajectories. In comparing the phosphate-phosphate distances in the A-form ensembles generated in implicit and explicit solvent, significant differences were apparent. The distance between phosphate groups across the major groove (figure 2.4) were found to be narrowly distributed around 13.0 ± 0.8Å in the implicit solvent ensemble while the distances were considerably larger, 14.6 ± 1.7Å, in the case of the explicit solvent ensemble. The B-form structures show the same trend for

the major groove phosphate-phosphate distances. The explicit solvent model yields larger distances (17.8 ± 1.1Å) than the implicit solvent model (15.2 ± 1.2Å). Major groove phosphate-phosphate distances are divergent between the two solvent models by 15% in the B-form and by 11% in the A-form. Therefore, the inconsistency of between the free energy differences of the two solvent models is an effect of the dissimilarity in phosphate-phosphate distances.

Conclusions

The practical utility of end-point free energy methods lies in their ability to quickly evaluate free energy differences between two well-defined states and in their ability to decompose the estimated free energy into meaningful components that may permit qualitative assessment of the key physical features contributing to the equilibria.[52] While these features are valuable, their value is based on the contention that the estimated free energies resulting from endpoint methods are reliable to an acceptable level of error. In order to address this contention, the reliability and robustness of a popular endpoint method, termed MM/PBSA, was evaluated in this study. Specifically, the reliability of this technique for evaluating conformational equilibria of nucleic acids was investigated by comparing MMPB/SA free energies evaluated in the CHARMM force field to free energies determined using: 1) an umbrella sampling pathway approach utilizing the same CHARMM force field, 2) the same MM/PBSA approach using a different force field and 3) the MM/PBSA approach using a different solvent model to mediate conformational sampling.

In order to investigate the robustness of the MM/PBSA free energy calculation, we evaluated the influence of solvent model, force field, and free energy algorithm. The conformational free energy differences for the hexamer, using MM/PBSA, showed agreement with the results based on the PMF free energy profiles of Roux *et al*.[65] When the ensembles were generated in explicit solvent, the A→B form DNA conformational free energy difference for the hexamer was similar to the PMF result.[65] The MM/PBSA calculations of the ensembles generated in the implicit solvent produced free energy

differences that are qualitatively representative of the correct conformational trend (favoring B-form) in an aqueous solution. However, the explicitly solvated hexamer resulted in MM/PBSA free energies that quantitatively resembled the PMF result[65] more than the implicitly solvated hexamer. The use of implicit solvent models during conformational sampling resulted in ensembles exhibiting phosphate-phosphate distances distinct from those observed in ensembles generated in an explicit solvent. This resulted in quantitatively divergent conformational free energy differences, while not significantly impacting the salt effect or the free energy change upon increasing the ionic strength. The free energies of ensembles generated in both the explicit and implicit solvent model showed similar sensitivities to ionic strength. A comparable salt effect was also exhibited in the comparison of the MM/PBSA free energies[48] using the CHARMM[67] and AMBER[66] force fields. It was demonstrated that the two force fields produced MM/PBSA conformational free energies that were similar, but the enthalpic and entropic contributions to free energy varied between the force fields. These differences can be explained by an entropy-enthalpy compensation effect.

In summary, we found the MM/PBSA method to be a reliable approach for estimating the conformational free energy difference of oligomer DNA sequences. The MM/PBSA method performed well in reproducing the results of the more rigorous umbrella sampling method. When the explicit solvent model was used during the conformational sampling, regardless of which force field was chosen, the MM/PBSA method was qualitatively robust in calculating the free energy difference of the A to B form equilibrium.

CHAPTER THREE

ELECTROSTATIC INFLUENCES ON BASE FLIPPING

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.[86, 87] However, base flipping has been shown to occur spontaneously[88], and in some cases enzymes utilize base flipping for catalysis.[28] For example, uracil DNA glycosylase enzymes, target the exposed base, and stabilize the flipped-out state for the purpose of base excision repair. [2, 29]

Several studies have investigated the effects of the base flipping conformational transition on enzyme function.[29, 88, 89] 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.[29] These experiments yield base opening rates as well as the equilibrium ($K_{flip} = k_{op}/k_{clsd}$) between open and closed state. Umbrella sampling[90, 91] 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.[90, 91] An umbrella bias potential is applied to sample across the chosen reaction coordinate, from one end-point to the other. Since the path between the flipped-out and flipped-in states is known, reaction coordinates are easily constructed for base flipping.[92-94]

When molecular dynamics is used to describe conformational changes of proteins or nucleic acids, a suitable force field is critical.[95, 96] Priyakumar *et al*.[97] tested the performance of three force fields (CHARMM27[98], AMBER4.1[99], and BMS[100]) 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[92] 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[101] 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.[101] Equilibrium constants for base flipping measured with NMR proton exchange[102] were compared with the free energy difference results from the force fields. In comparison with experimental values, free energies generated with CHARMM and AMBER were more similar to experimental values than those generated with BMS.[97]

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[96, 103], 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 in 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.[104, 105]

Implicit solvent models offer an alternative, representing the solvent as a function of the solute configuration. Many implicit solvent models have been developed.[106-108] Typically, the free energy of solvation ($\Delta G_{solvation}$) is broken down into the polar and nonpolar contributions (equation 3.1).

$$\Delta G_{Solvation} = \Delta G_{pol} + \Delta G_{nonpol} \quad \text{Eq. 3.1}$$

$$\Delta G_{nonpol} = \sum_i \gamma_i A_i \quad \text{Eq. 3.2}$$

The nonpolar contribution (equation 3.2) is the cost of creating a cavity within the solvent, which is proportional to the surface area (A_i) of the solute.[106, 109] When studying DNA, the polar solvation term is the dominant contribution to solvation due to the highly negative DNA backbone.[110] 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}). The Poisson equation (equation 3.3) [111] is commonly solved numerically by a finite difference method[112, 113], which can be computationally expensive when implemented in molecular dynamics or monte carlo simulations. When the influence of ionic strength is factored in, equation 3.3 becomes the Poisson-Boltzmann (PB) equation.[111]

$$\nabla[\varepsilon(r)\nabla\varphi(r)] = -4\pi\rho(r) \qquad \text{Eq. 3.3}$$

The PB equation yields an electrostatic potential φ(r), where ε(r) is the distance dependent dielectric, and ρ(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.[111, 114]

$$\propto_i = -\frac{1}{2}\left(\frac{1}{\varepsilon_p} - \frac{1}{\varepsilon_w}\right)\frac{1}{G^i_{pol}} \qquad \text{Eq. 3.4}$$

The Born equation (equation 3.4)[115] 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 Born radius (α) is the distance between atom i and the solvent boundary. An extension of the Born equation is the generalized Born (GB) equation (equation 3.5), where the

empirical factor F may range from 2 to 10, while 4 is the most common value.[116] The Debye length (κ), which is proportional to the square root of the electrolyte ionic strength ($(I)^{1/2}$), is applied to represent salt effects.[117] There are several analytic generalized Born (GB) models that are optimized to reproduce Poisson solvation energies using rapidly solved parameterized equations.[106-108, 118, 119] 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.

$$\Delta G_{pol} = -\frac{1}{2}\left(\frac{1}{\varepsilon_p} - \frac{-e^{\kappa f_{GB}}}{\varepsilon_w}\right)\sum_{i,j}\frac{q_i q_j}{f_{GB}} \qquad \text{Eq. 3.5}$$

$$f_{GB} = \left[r_{i,j}^2 + \alpha_i \alpha_j exp\left(-r_{i,j}^2/F\alpha_i\alpha_j\right)\right]^{1/2} \qquad \text{Eq. 3.6}$$

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[8], two sets of atomic radii were compared, the atomic van der Waals radii and the atomic radii developed by Banavali and Roux (BR).[120] 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.[80, 121, 122] 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 mechanism of undamaged and damaged DNA bases. For the purpose of examining the performance of the GB model on the base flipping mechanism, the GB and TIP3P solvent models were used during umbrella sampling 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*.[8] The natural base pair GC was analyzed along with the damaged DNA bases uracil and xanthine.

Methods

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

The CHARMM c32b1 molecular mechanics package and the CHARMM27[98, 123] all-atom nucleic acid force field were used in all molecular dynamics simulations. The starting coordinates of the dodecamer sequence d(GTCAGXGCATGG)$_2$ were generated within CHARMM. The base to be flipped out of the helix is X. This sequence was chosen because it has been used in many base flipping studies previously[94, 97] and provides an easy comparison. Using the program 3DNA[124], the canonical B-form DNA structure of the sequence d(GTCAGCGCATGG)$_2$ was constructed. The base complementary to X was systematically modeled as guanine, adenine, cytosine and thymine. The flipping base (X) in each of these DNA models was systematically modeled as cytosine, uracil or xanthine.

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*.[97] Umbrella sampling was performed to calculate the PMF associated with base flipping using an explicit and implicit solvent. A pseudo-dihedral angle defined through the centers of mass (COM)[92] 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 was used as the reaction coordinate. This is the same pseudo-dihedral angle used as a flipping reaction coordinate in previous work.[97] The corresponding molecular dynamics simulations were run in a generalized Born solvent

and also TIP3P solvent. The PMF profiles were created by incrementing the pseudo-dihedral angle 5° in each simulation 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 base-opened or flipped out state. Starting structures for these simulations were created by minimizing 100 steps with the adopted basis Newton-Rapheson, and using the miscellaneous mean field potential (MMFP) module in the charmm package to increment the pseudo-dihedral angle with a force constant of 10,000 kcal/mol/rad^2. Starting structures were varied ±5° from the final structures of the previous minimization. The starting structures were then used in simulations with explicit solvent and implicit solvent. A harmonic umbrella potential $w_i(x) = k_i (x - x_i)^2$ 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[119] was used as the implicit solvent system since it was determined by Feig *et al*. [125] to closely reproduce PB solvation energies. For GBMV, we used a β value of -20, and a water probe radius of 1.4Å. The inverse Debye length (κ) [117]was set to 0.129Å^{-1}, which corresponds to the physiological salt concentration (0.15M). Nonpolar contributions (equation 3.2) to the solvation free energy are accounted for here as the product of the solvent accessible surface area (A) and the surface tension (γ).[106, 109] The surface tension was set to 0.03kcal/mol/Å^2 since it was used in previous studies to calculate the nonpolar solvation energy.[118] 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[118] 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.[126] 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.[127] The interaction energies between explicit solvent and flipping base (X) 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. Possible base pair interactions that were within 3.5Å of the flipping base and linear were designated as hydrogen bonds. Then a hydrogen bond percentage over the trajectory was generated. These calculations were performed over the base paired umbrella sampling windows to determine the hydrogen bond fractions of the base paired state.

Free Energy of Base Flipping

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.

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*.[97] The results from base flipping, where cytosine was the flipping base, are displayed in figure 3.1. 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 (Figure 3.1A black), and TIP3P solvent model (Figure 3.1A 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 base paired

state at ~10°, which is consistent with results from Priyakumar *et al*.[97] 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 equation 3.5) are typically chosen in order to account for electronic polarizability of the solute molecule.[111] 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. [128, 129] Figure 3.1B 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 (Figure 3.1A red). Electrostatic interaction energies between the flipping base 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 destabilized to -11.1 ± 1.08 when the higher dielectric of 2.0 was used. The destabilization of the base pair interactions likely caused the decrease in base flipping free energy after the interior dielectric adjustment.

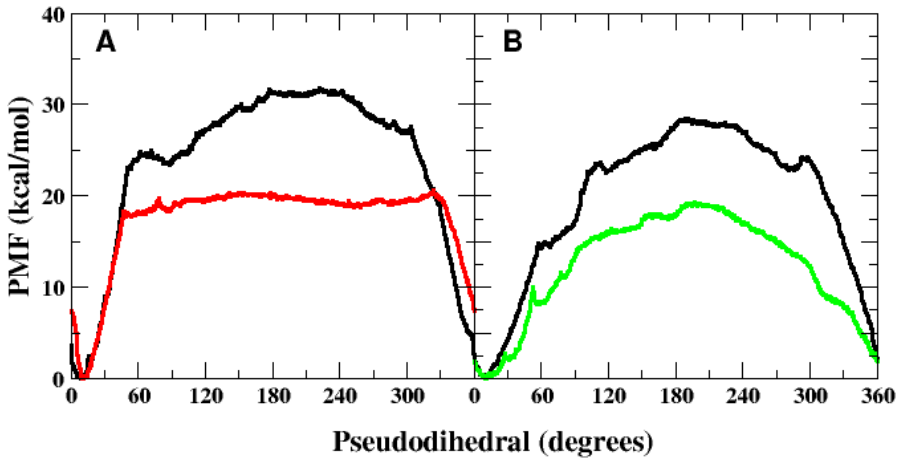


Figure 3.1 - Potentials of mean force (PMF) of base flipping for GC base pairs along the pseudodihedral angle coordinate. Watson-Crick base pairing is approximately 10°-30° pseudodihedral angle and the flipped out state is approximately 190° (line). **A.** Umbrella sampling performed with GBMV implicit solvent (black) 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).

Damaged DNA Base Flipping

In order to examine the base flipping equilibria of damaged bases, the free energy difference between the base opened and base closed states were represented with PMF

profiles. In figure 3.2, 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 (Figure 3.2A and 3.2D). This implies the flipping base favors the extrahelical state in explicit solvent over implicit solvent. Since the implicit solvent model does not include solute-solvent interactions, the hydrogen bonds between the extrahelical base and explicit waters are most likely responsible for the discrepancy.

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 (Figure 3.2A) are greater in energy than those generated using TIP3P (Figure 3.2B), 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 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 (Figure 3.2D) yields greater base flipping free energy differences than the TIP3P solvent model (Figure 3.2E). 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 base paired state. The number of hydrogen bonds in the base paired 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 extrahelical base, but also the interactions between the WC base pairs.

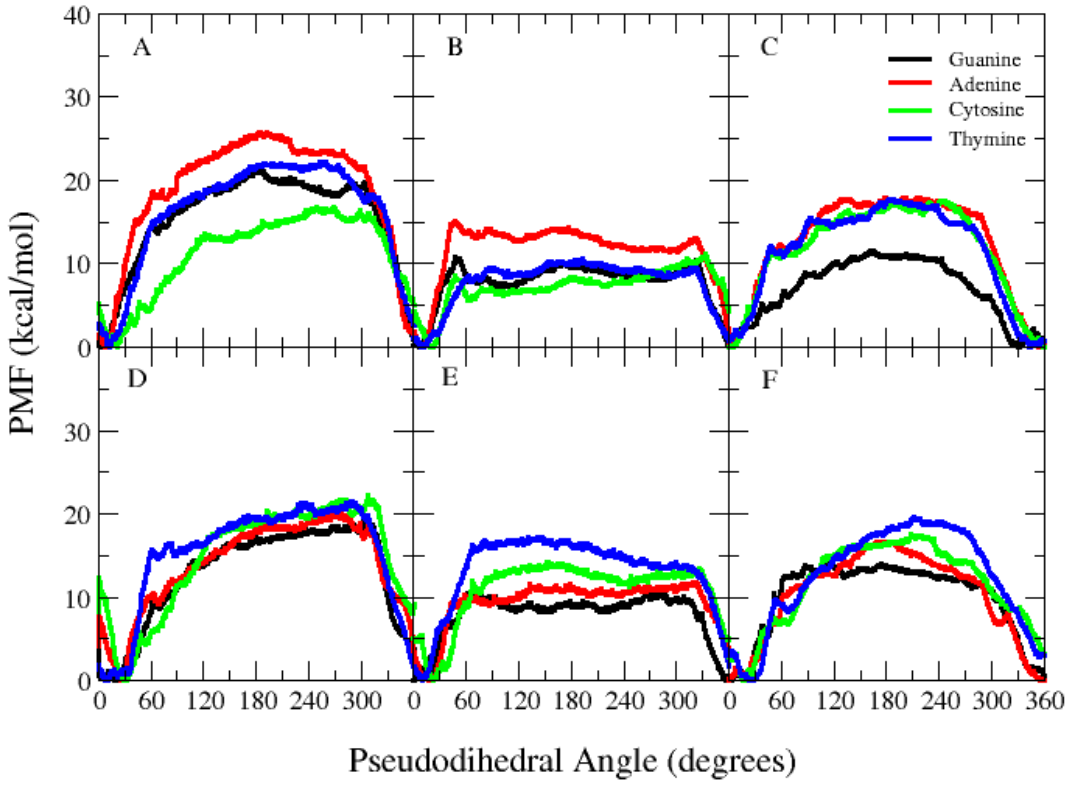


Figure 3.2 – Potentials of mean force (PMF) of uracil- containing (A,B,C) and xanthine-containing (D,E,F) base pairs along the pseudodihedral angle coordinate. Watson-Crick base pairing is approximately 10°-30° pseudodihedral angle and the flipped out state is approximately 190° (line). **A.** Base flipping PMF profiles generated with GB solvent model **B.** Base flipping PMF profiles generated with TIP3P solvent model **C.** Base flipping PMF profiles generated with GB solvent model and ε_p=2.0 **D.** Base flipping PMF profiles generated with GB solvent model **E.** Base flipping PMF profiles generated with TIP3P solvent model **F.** Base flipping PMF profiles generated with GB solvent model and ε_p=2.0

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 (Figure 3.1B), 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 Figure 3.2C and 3.2F. 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 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.

Solute-Solvent Interactions

Interacting with the Extrahelical 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*. (~20kcal/mol) for the CHARMM27 force field, and with the experimental GC interaction energies used in the parameterization of the CHARMM27 force field.[97, 98] Therefore the discrepancies observed in the base flipping PMF profiles can be attributed to solvent-solute interactions in the extrahelical 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 flipping base were included within a 5Å cutoff. It can be seen in figure 3.3 that interactions between the flipping base and water molecules are the most favorable in the flipped out state (~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.[97] Interaction energies show approximately a 30 kcal/mol difference between the flipped in and flipped out state. This interaction energy in the extrahelical state significantly stabilizes the extrahelical state of the explicit solvent simulations. Since implicit solvent models lack these interactions, they most likely contribute to the difference observed between the GB and TIP3P PMF profiles.

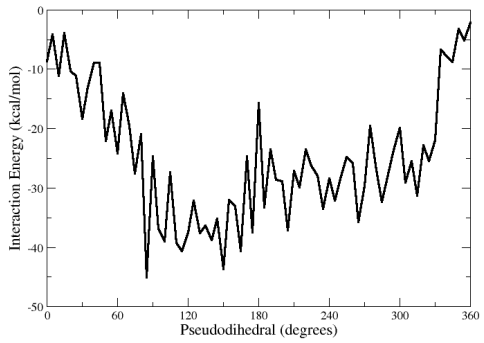


Figure 3.3 – Interactions between water molecules and extrahelical base. A.) Interaction energies between cytosine flipping base and TIP3P explicit water molecules (within 5Å of base). Energies averaged over 500ns of production trajectory for each pseudodihedral simulation window.

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 3.1. 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.

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%

Table 3.1 - Uracil and Xanthine H-bond fractions

Conclusions

Umbrella sampling was performed over the base flipping pathways of natural and damaged DNA base pairs. 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*.[97] However, the PMF profiles of the GC and uracil base pairs showed the WC base paired state was overstabilized in the implicit solvent model. It was hypothesized that differences in solute-solvent interactions in the extrahelical 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 extrahelical state. These interactions are not represented in the GB solvent model, and therefore the extrahelical 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 base paired state. Hydrogen bond fractions of the WC base paired 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 base

paired 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[8, 121, 122]. 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 arrangement of the uracil profiles did not agree with previous GB and TIP3P solvent results. Although GB models used here did not exactly reproduce free energy differences from the explicit solvent model, the adjustment of the interior dielectric constant may allow for the efficient and accurate use of GB solvent models during base flipping calculations.

CHAPTER FOUR

MUG ACTIVE SITE INTERACTIONS WITH DEAMINATED BASES

Sections of this work have been published as:

Lee, H.W., Brice, A.R., Wright, C.B., Dominy, B.N., Cao, W., *Identification of Escherichia coli Mismatch-specific Uracil DNA Glycosylase as a Robust Xanthine DNA Glycosylase*, Journal of Biological Chemistry, 285 (2010) 41483-41490.

Introduction

During initiation of the base excision repair (BER) process, uracil DNA repair enzymes recognize and excise deaminated cytosine bases in order to prevent the irreversible modification of the genome.[28, 130-132] Glycosylase enzymes require damaged DNA bases to be in a flipped-out state in order for catalysis of the cleavage to occur.[25, 26] For glycosylase enzymes, cleavage takes place at the N-C1' bond of the nucleotide.

Chapter one briefly discussed the uracil DNA glycosylase (UDG) superfamily, and the following provides more detail on the first two families of the superfamily. All of the enzymes within the UDG superfamily are capable of cleaving uracil from the DNA helix. The active site among all the families is structurally homologous, although there are several differences in the active site residues between families 1 and 2 specifically (Table 4.1).[28, 133] Family 1 is called UDG, and enzymes have been taken from both human (hUDG) and *E.coli* (eUDG) organisms. [26, 134] Mismatch-specific uracil DNA

glycosylase (MUG) is a family 2 enzyme that has been characterized from *E.coli*.[135] The asparagine at position 18 (N18) in *E.coli* MUG is conserved within the MUG/TDG family, which is almost certainly due to its significant role as the catalytic residue.[3] In all UDG enzymes, the role of the catalytic residue is to initiate a nucleophilic attack at the C1' position on the flipped-out nucleotide. In family 1, this is carried by an aspartate (ASP) residue, which acts as a nucleophile in the catalysis of the N-glycosidic bond cleavage. However, in family 2 the N18 orients a water molecule within proximity (1.7Å) of the C1' of the deoxyribose sugar in the DNA backbone.[3, 135] Then, the water molecule acts as a nucleophile that initiates the hydrolysis and cleavage of the N-C1' bond. Despite structural active site conservation, it is interesting that the most critical residue in regard to the glycolysis reaction is unconserved between the individual UDG families. For some UDG enzymes, the catalytic residue is known to also stabilize the resulting transition state.[136] Table 4.1 shows the catalytic residues that have been reported for family 1 and family 2 enzymes.[28] The transition state for all UDG enzymes is an oxocarbenium ion, which was established through kinetic isotope effect experiments.[137] In family 1, a histidine residue (residue 187 or 268) plays a role in stabilization of the cleaved base transition state.[136, 138, 139] However, the corresponding N140 in MUG and S271 in thymine DNA glycosylase (TDG) are only capable of nonspecific binding to the DNA backbone.[2] Further analysis of the function of individual amino acids revealed details of the MUG and TDG recognition mechanisms.

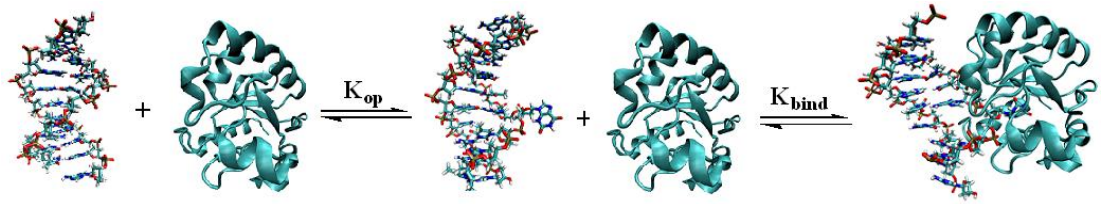
Significant UDG Superfamily Residues				
hUDG	**eUDG**	**eMUG**	**hTDG**	**Description of Residue**
D145	D64	N18	N140	Catalytic Residue/Transition State stabilization
H268	H187	N140	S271	Non-specific interaction with DNA phosphate/Specific interaction with uracil
L272	L191	L144	R275	wedge interacts with complementary strand
		G143/R146	A274	wedge interacts with damaged strand
F158	F77	F30	Y152	π-stacking with substrate
Q145	Q63	I17	I139	Specific H-bond with Substrate

Table 4.1 – Conserved residues and their reported role in UDG function.[28]

In the recognition of damaged bases, it is unclear whether all glycosylase enzymes actively increase the rate of base flipping, or there are some that play a more passive role by thermodynamically stabilizing the flipped out state.[27, 140] The conserved wedge region of UDG enzymes intercalates into the abasic site of the DNA to stabilize the flipped out state, or possibly to induce the flipping process.[3, 138, 141] Residue L144 of the wedge may have the ability to push the base out of the stack. The wedge residues R146 and G143 have been shown to form interactions within the abasic site of the DNA, and stabilize the flipped-out state.[3]

In chapter two, studies were described that have examined the base flipping process through the use of NMR experiments.[29, 88] In Stivers *et al.* [29], NMR spectroscopy techniques were used to monitor exchange rates of imino protons with solvent. With this method, the flipped-out and flipped-in (Watson-Crick base pair) conformations were characterized with and without the enzyme. Family 1 UDG enzyme binding was studied, which is specific for only the uracil base. It was determined that the

rate of the base flipping was unaffected by the presence of UDG enzyme (family 1). However, the lifetime of the flipped out state was increased 100-fold when the UDG enzyme was added. This implies the UDG enzyme was not actively flipping out the deaminated base, but recognizing, and binding to the flipped out base. Since the UDG enzyme did not increase the rate of flipping for the deaminated base, the overall binding equilibrium can be separated into two equilibria. Figure 4.1 shows the equilibria, which include the base opening (K_{op}), and the enzyme binding to the flipped-out base (K_{Bind}). The next step in the mechanism is the cleavage of the damaged base at the N-glycosidic bond catalyzed by the enzyme (k_{cat}).



dsDNA(closed) + Enzyme $\overset{K_{op}}{\rightleftharpoons}$ **dsDNA(open) + Enzyme** $\overset{K_{bind}}{\rightleftharpoons}$ **dsDNA(open)-Enzyme**

Figure 4.1 – The binding equilibrium for MUG separated into two equilibria: the base flipping mechanism (K_{op}) and the binding of the enzyme to the opened base (K_{bind}).

Activity cleavage assays of several MUG mutants were performed by Lee *et al.*[2] Enzyme activity and binding assays (figure 4.2) were conducted for the MUG enzyme with the damaged bases (figure 4.2), and all of the natural DNA complementary bases. Since MUG is part of the UDG superfamily, it is known to have activity towards uracil.[142]

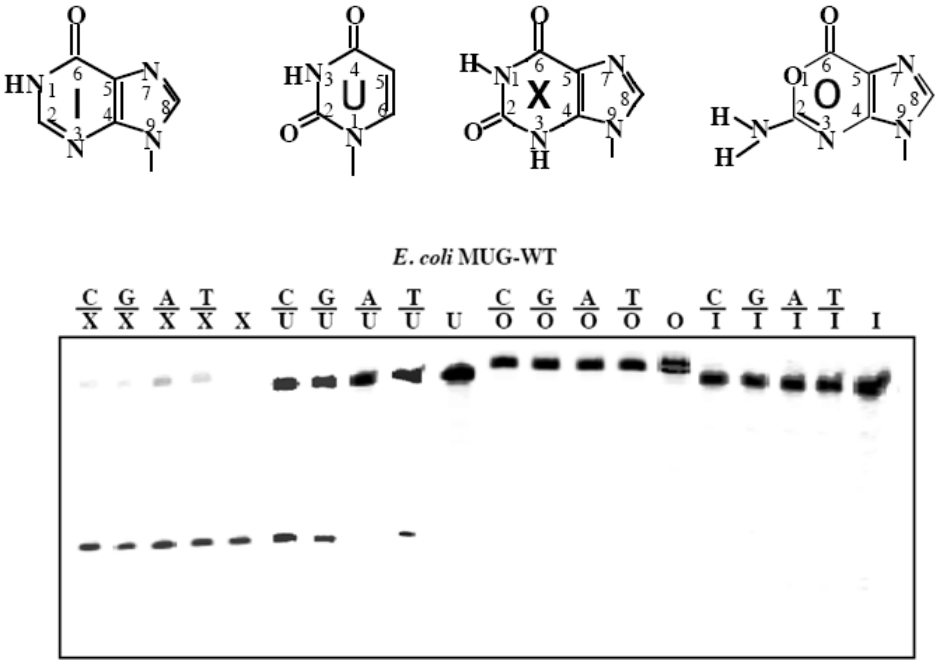


Figure 4.2 – Activity assay for *E.coli* MUG. Chemical structures of deaminated bases, Inosine (I), Uracil, (U), Xanthine (X), and Oxanine (O)[2]

Lee *et al.* [2] discovered that the wild type of MUG has strong activity on both uracil and xanthine (figure 4.2). The enzyme shows greater specificity for all of the xanthine base pair substrates than for the uracil base pair substrates. Since MUG is active for more than one substrate, the Michaelis-Menten constant (K_M or $1/K_{bind}$) will not be negligible, and the rate of the enzyme-catalyzed reaction will be given by equation 1.4. Then, equation 1.4 is manipulated to show the contribution of the base flipping equilibrium on the overall rate of catalysis:

$$\frac{d[P]}{dt} = \frac{k_{cat}}{K_M} K_{op}[E]_0[dsDNA]_0 \qquad \text{Eq. 4.1}$$

Where the K_{op} is the equilibrium constant for the base flipping process. The K_{bind} accounts for the total free energy of binding when the enzyme binds to the flipped-out base. The influence of K_{op} and K_{bind} on the catalytic rate were observed in the enzyme activity results for MUG (figure 4.2).[2]

The base flipping equilibria of the damaged bases have been examined previously, where the free energy difference between the flipped-out and flipped-in states were represented with PMF profiles.[2, 32] We know from Stivers *et al.* the base flipping mechanism occurs independently of the UDG enzyme binding.[29, 86] Human UDG and *E.coli* MUG are enzymes of the same UDG superfamily, thus it is feasible for these two enzymes to undergo similar mechanisms of binding. In order to determine the relationship between base pair stability and enzyme activity, free energy differences of base flipping were generated for each complementary base (figure 3.2). There have not been many studies on the stability of xanthine-containing base pairs.[2, 143] However, it is known that the AU base pair is the most stable uracil base pair, which is expected since AU forms a natural Watson-Crick base pair.[144] Activity assays from Lee *et al.* show that the AU base pair is the least active of the four uracil base pairs, and the free energy differences from the uracil base flipping profiles exhibit the same trend.(figure 3.2) These results indicate the uracil flipping mechanism (K_{op}) strongly influences enzyme

recognition. To understand the MUG recognition completely, the conformational transitions of the enzyme should be considered.

In a recent study, the effect of mutations on a family 3 UDG enzyme was analyzed using molecular dynamics and flexibility calculations.[31] Similar to MUG, it was determined that SMUG1 from the UDG family 3, has xanthine and uracil activity. Electrostatic interaction energies were calculated for minimized structures between the enzyme and substrates xanthine and uracil. Interaction energies indicated that active site interactions were more favorable for SMUG1 with xanthine, than with uracil. A more detailed analysis of the active site interactions revealed a favorable interaction with the N7 of xanthine. Site-directed mutagenesis was performed to show the influence of specific residues on the enzyme activity. The SMUG1 mutants G63P and M57L both removed or reduced xanthine activity. In order to show the flexibility of the enzyme, root mean squared fluctuations (RMSF) were calculated per residue over molecular dynamics trajectories. It was shown that mutants G63P and M57L both increased the flexibility of the wedge region, and concluded that the higher flexibility may cause lower specificity for xanthine.

In order to fully understand the influence of active site interactions (K_{bind}) on the recognition of *E.coli* MUG, several point mutations of MUG were analyzed. Site-directed mutagenesis experiments were performed by Cao and coworkers, and the corresponding mutations were then modeled for comparison.[2] The following studies used molecular modeling, and interaction energy calculations to illustrate the effects of the mutations on active site interactions. Several of the mutations were in close proximity to the active site,

which made interaction energies sufficient for displaying the difference between wild type and mutation. However, the mutations that were not directly contacting the substrate or active site, correlated motions were determined in order to show possible pathways for the long-range mutation effects.

Methods

Molecular models of the unbound and bound conformations of wild type *E.coli* MUG were used as initial structures. The crystal structure of *E.coli* MUG (pdb accession code 1mug) was used as a model for the unbound MUG enzyme.[135] The molecular model of the wild-type *E.coli* MUG complexed with a DNA decamer sequence (AAAGATGACA) containing uracil was constructed based on the crystal structure of UDG bound to a DNA dodecamer (pdb accession code 1emh).[145] Using the Swiss-Pdb Viewer (SPDBV) program[146], the model of *E.coli* MUG bound to the decamer was generated by performing a structural alignment between the crystal structure of MUG and the crystal structure of the UDG/decamer complex. The UDG structure was then removed, leaving a structural model of MUG bound to DNA.

Molecular dynamics simulations were performed on the bound MUG structures using the CHARMM 32b1 molecular mechanics software package[147], and the CHARMM27 force field.[98, 101] The solvent molecules were represented with the explicit TIP3P water model. A solvent box was constructed that resulted in a minimum water layer of 10Å between the solute and the boundary of the box, which yielded ~17700 water molecules. 14 sodium atoms were added for electrical neutrality. Periodic boundaries and Ewald summation[126] were used to account for long-distance electrostatics. The starting structures were gently minimized with the adopted basis Newton-Raphson (ABNR) module in CHARMM, for 100 steps to remove any unfavorable clashes. The system was heated for 200ps, from 200K to 300K in increments of 1K every 2ps. Using an integration timestep of 2fs, a canonical ensemble was

generated for 2ns of production. Interaction energies consisting of Coulomb and van der Waals energies were calculated over the trajectory between the active site residues and the substrates. Active site residues were defined as any atom of the enzyme within 8Å of the substrate.

Mean square fluctuations (MSF) were calculated using the CHARMM 32b1 molecular mechanics software package.[147] The RMSF (coor dyna) for each amino acid was calculated for the unbound enzymes, over the production portion of molecular dynamics trajectories. These calculations were performed for both *E.coli* MUG and all of the S22 mutations.

Normal mode trajectories were generated for the calculation of a covariance matrix. The free protein crystal structure of MUG (1mug.pdb) was minimized using ABNR with a harmonic restraint on each heavy atom. A loop was used to decrease the restraint from 10 kcal/mol*Å^2 in decrements of 1 kcal/mol*Å^2 until the restraint was zero. The VIBRAN module in CHARMM was used generate the normal mode trajectory at a temperature of 300K. Correlated motion calculations entailed determining the covariance of atomic displacements over the normal mode trajectories. The COOR COVA module in CHARMM was used to construct the covariance matrices, which converged after the first 200 normal modes were superimposed. Below, S_{ij} is the covariance (Eq 4.2) of the displacement of the protein backbone atoms.

$$S_{ij} = < x_{(i)} \bullet x_{(j)} > / \sqrt{< x_{(i)}^2 >< x_{(j)}^2 >} \qquad \text{Eq 4.2}$$

The displacements of the C_α atoms for residues i and j relative to the average coordinates are represented by x(i) and x(j). In the generated matrix, non-zero covariance values indicate residues are strongly correlated, while covariance values close to zero signify residues that are weakly correlated. A positive covariance implies the corresponding residues are moving similarly, while a negative covariance implies the residues are moving in opposite directions.

Mutation Effects on E. coli MUG

While MUG has been identified as active on uracil, its activity against xanthine was previously unknown.[2] Molecular models were constructed to characterize the possible interaction between MUG and xanthine. The similarity between the Watson-Crick faces of uracil and xanthine suggests that similar hydrogen bond donor/acceptor patterns may partly explain the ability of some uracil glycosylases to also interact with xanthine. The bound molecular models were generated on this premise (Fig. 4.3) and illustrated the potential for xanthine to form hydrogen bonds similarly to uracil within the MUG active site. In Figure 4.3, it was also demonstrated that the N7 of xanthine was capable of forming a hydrogen bond with the sidechain from S23, while uracil was not capable of this hydrogen bond. Energies were generated in order to quantify the difference in interactions.

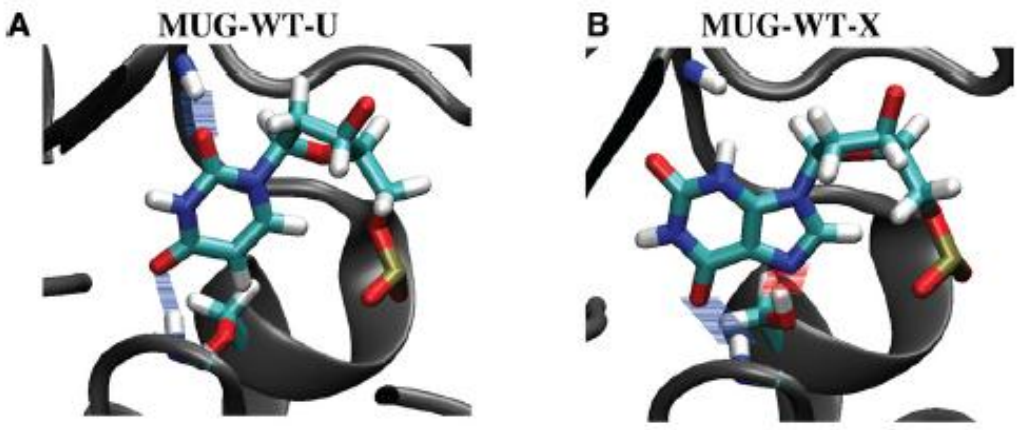


Figure 4.3 – Molecular modeling of *E. coli* Mug recognition **A.** Interactions between wt *E. coli* MUG and uracil. Mainchain hydrogen bonding between N18, F30 and uracil are shown in blue. **B.** Interactions between wt *E. coli* MUG and xanthine. Mainchain hydrogen bonding between F30 and uracil is shown in blue. Sidechain hydrogen bonding between S23 and N^7 of xanthine is shown in red.

A description of the active site interactions provided insight into the function and activity of the enzyme. Perturbations on these interactions from point mutations were easily observed in some of the bound models, while others required further conformational sampling to observe the effects. In order to further understand the specificity for xanthine, Coulombic and van der Waals interaction energies were calculated over molecular dynamics trajectories. Interaction energies were determined between the *E.coli* MUG active site, and the substrates xanthine and uracil. Even though this is not an accurate method for calculating the free energy of binding, the interaction energies did provide insight for the binding equilibrium (K_{bind}). These active site

interaction energies with the bound DNA substrate reveal a significant enthalpic component of binding free energy. More importantly, the comparison of these active site interactions for the different substrates allow for a qualitative method of determining the critical interactions necessary for specificity. These results (Figure 4.4) demonstrated more favorable interaction energies with xanthine than with uracil, which reinforced the model description (Figure 4.3). This analysis provided further support to the significance of the S23 residue. The hydrogen bond with S23 may clarify the enzyme activities of the wtMUG enzyme, where xanthine was more active than uracil.

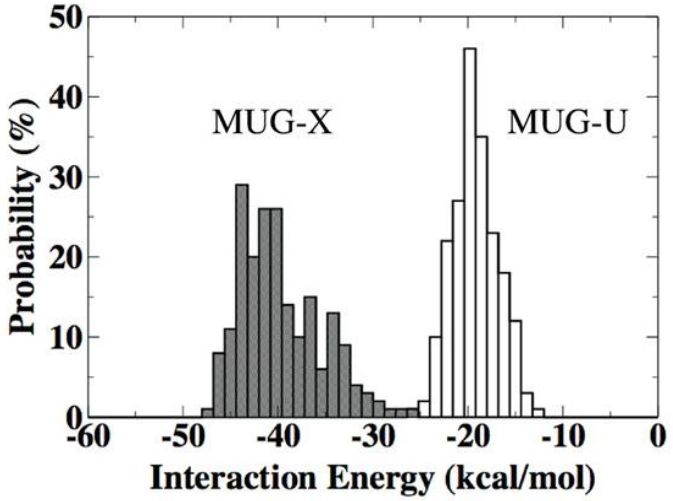


Figure 4.4 – Interaction Energies for *E. coli* MUG with xanthine and uracil. Energetics of wt *E. coli* MUG interactions with G/X (solid bars) and G/U base pairs (blank bars).

S23 Residue Provides Xanthine Specificity

In order to explain the difference in activities for the wtMUG and S23A mutant, differences in interaction energies were examined over MD trajectories. The results

(Figure 4.5) indicate that wtMUG has stronger electrostatic and van der Waals interactions with xanthine than with the S23A mutant. This difference in energy is due to the loss of a hydrogen bond between the sidechain hydroxyl of S23 and N7 in xanthine.

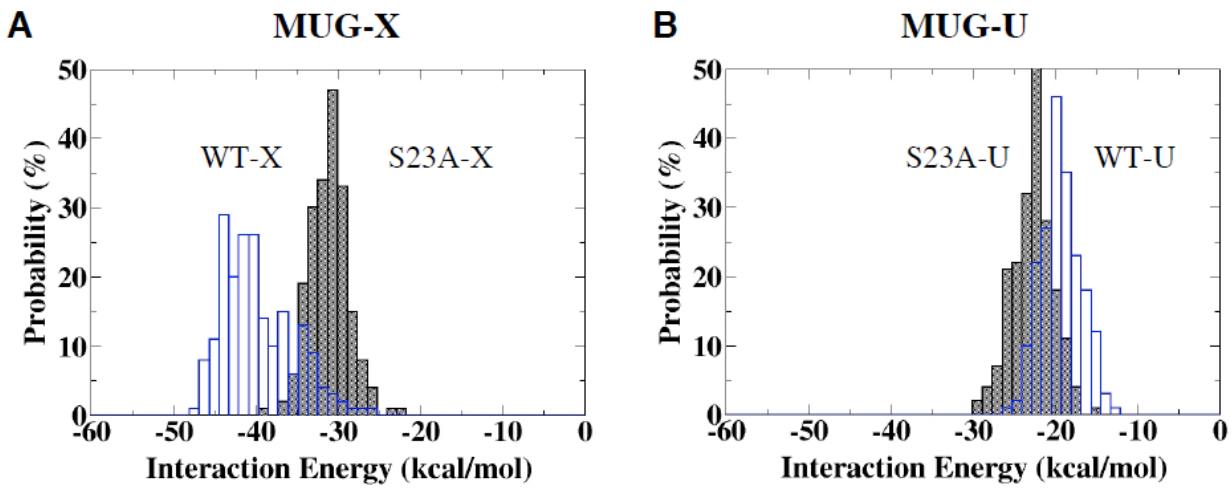


Figure 4.5 – Effect of the S23A mutant on active site interactions **A**.Energetics of *E. coli* MUG interactions with G/X. Blank bars, MUG-WT; solid bars, MUG-S23A. **B.**Energetics of *E. coli* MUG interactions with G/U. Blank bars, MUG-WT; solid bars,MUG-S23A.

Unlike xanthine, uracil is more active with S23A than with wtMUG. The possible hydrogen bond donors and acceptors within the MUG active site are reduced in the S23A mutation, however since S23 does not interact with uracil the loss of this hydrogen bond donor is insignificant to UDG activity. Uracil still forms stronger interactions with the active site of S23A, which are mostly attributed to a stronger hydrogen bond with the F30 and N18 mainchain. The MSF in the region of the short α-helix bordering the active site

is greater for S23A, implying that S23A is more flexible than MUG in that region (figure 4.6). The greater flexibility of the short α-helix allows the uracil substrate to adopt a more favorable configuration, and form the stronger interactions within the active site.

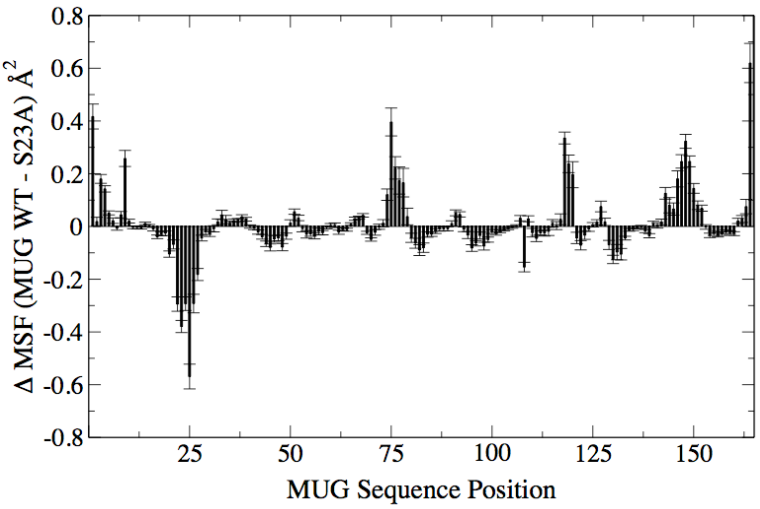


Figure 4.6 – Difference in isotropic mean squared fluctuations between the MUG-WT and MUG-S23A. MSF values were calculated within CHARMM using the "coor dyna" command, and the error bars correspond to the standard error of the ΔMSF values over the molecular dynamics trajectory. Positive ΔMSF indicates that C-α's in the S23A mutant are more rigid.

DNA Backbone Interactions with N140

Although most of the mutants studied maintain activity on xanthine, two of the N140 mutants constructed showed no detectable XDG activity (Figure 4.7 and 4.8). N140M results in a complete loss of xanthine and uracil activity, while N140H loses xanthine activity and reduces the uracil activity. Given that the wtMUG is much more

robust on xanthine than uracil, the complete loss of xanthine activity while still maintaining some uracil activity is dramatic. These results may indicate the significance of N140 in xanthine activity. Molecular models of the N140 mutants bound to uracil and xanthine were constructed in order to investigate the interactions at position 140. In the modeled MUG-uracil complex structure, N140 in MUG interacts with the phosphate backbone through hydrogen bonding (Figure 4.7A), which may contribute to the stabilization of uracil base pair DNA. Although N140 in MUG is sequentially aligned with M269 in hTDG, the structural alignment of these enzymes, performed with SPDBV [146], superimposes N140 of MUG with S271 of hTDG. Likewise, S271 of hTDG forms

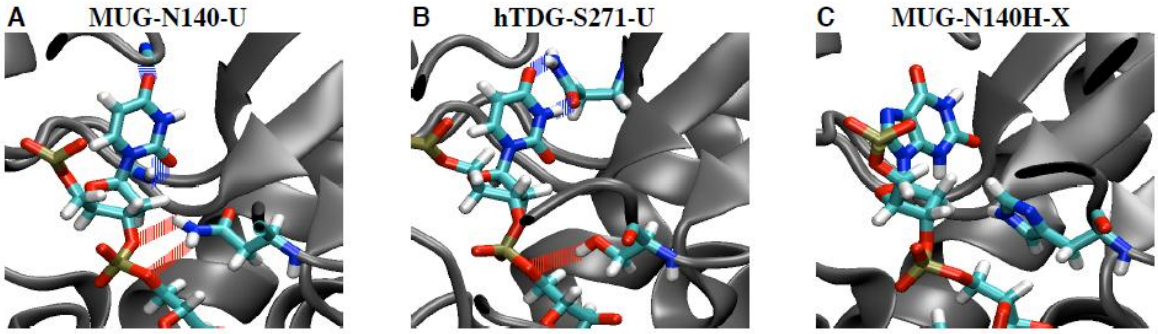


Figure 4.7 - Modeled structures of *E. coli* MUG and human TDG. **A.** Interactions of the sidechain of N140 with 3'-phosphate in the DNA backbone in *E. coli* MUG. DNA and N140 are shown in color. **B.** Interactions of the sidechain of S271 with 3'-phosphate in the DNA backbone in human TDG. DNA and S271 are shown in color. **C.** Lack of interactions between *E. coli* MUG-N140H and xanthine. DNA and N140H are shown in color.

equivalent hydrogen bonding with the phosphate backbone (Figure 4.7B). In the modeled N140H-uracil structure, N140H potentially can form a hydrogen bond with C^2-keto of uracil and form a weak hydrogen bond with the 3'-phosphate (Figure 4.8A). The presence of these favorable interactions may underscore the weak UDG activity of the N140H mutant. However, these potential interactions are lost when the uracil is substituted by xanthine (Fig. 4.7C), which may explain the loss of XDG activity. The loss of both XDG and UDG activity in N140M can be viewed as due to the loss of DNA backbone interactions as seen in N140 of MUG and S271 of hTDG or loss of direct hydrogen bonding to uracil as seen in N140H. The lack of favorable interactions with the backbone or the base may lead to the complete loss of both XDG and UDG activity (Figure 4.8B). These analyses are consistent with the previous study that identifies the role of +/- 1 phosphate interactions in transition state stabilization in the family 1 UNG enzymes [148].

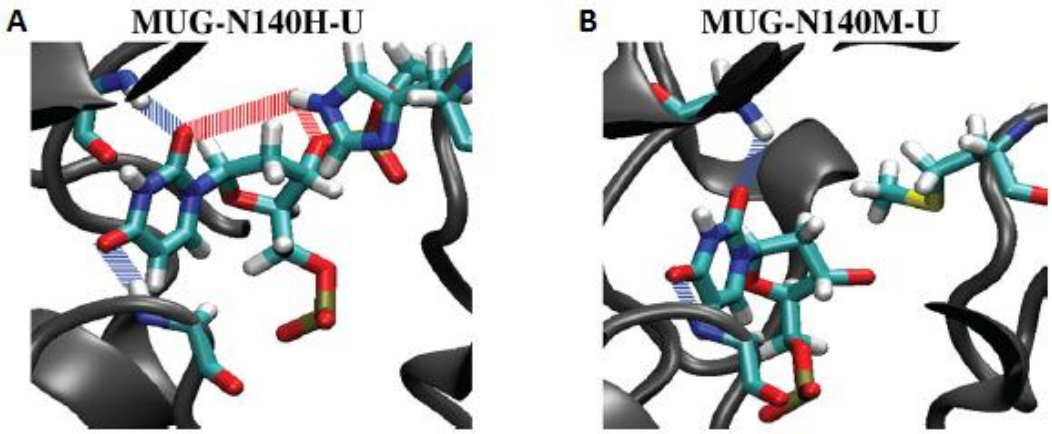


Figure 4.8 – Molecular modeling of N140 mutants **A.** Interactions between *E. coli* MUG-N140H and uracil. Hydrogen bonding between the sidechain of N140H and the uracil and that of the 3'-phosphate are shown in red. Mainchain hydrogen bonding between N18, F30 and uracil are shown in blue. **B.** Interactions between *E. coli* MUG-N140M and uracil. Mainchain hydrogen bonding between N18, F30 and uracil are shown in blue.

Comparing Active Sites of MUG and TDG

A distinct difference between *E. coli* MUG and hTDG is that while the former demonstrates a highly robust xanthine DNA glycosylase activity, the latter is void of the same activity completely. To understand the structural differences that may underlie the functional distinction, we created bound models of hTDG to compare the differences between how hTDG interacts with uracil and xanthine. Figure 4.9 shows uracil was stabilized by sidechain interactions provided by asparagine 191 (N191) , xanthine appeared to have fewer favorable interactions in the active site. The favorable sidechain hydrogen bonding involving S23 of MUG is not available because the position is

occupied by an ALA residue in hTDG. The reduction of XDG activity observed in MUG-S23A mutant illustrates the role of this interaction in xanthine recognition.

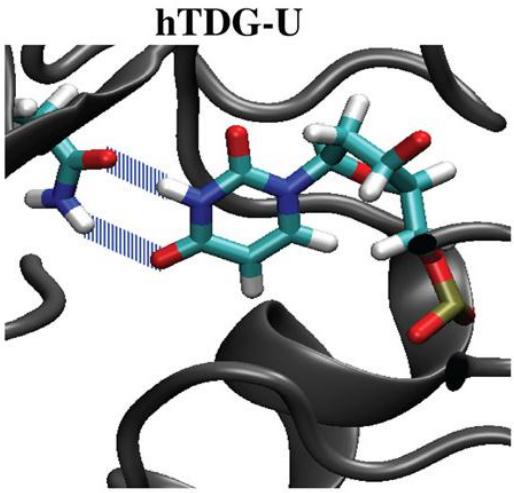


Figure 4.9 – Interactions between human TDG and uracil are shown. Side-chain hydrogen bonding between Asn-191 and uracil are shown in *blue*.

Increased UDG Activity in K68N

Specific increases in active site interactions were discovered for the K68N mutation, which was created to better understand the recognition mechanism of the MUG/TDG family. Wild type MUG showed activity for GU, CU, and TU, but not the AU pair. In the corresponding K68 position of MUG, both human TDG and *S.Pombe* TDG have an asparagine at this position within the active site. The mimic of TDG, the K68N mutation, was demonstrated as a significant mutant since it increased the activities for uracil base pairs, and as a result yielded activity for the AU pair. The K68 residue is located in the MUG active site, and replacing it with an asparagine produces a direct effect on the substrate-active site interactions. The molecular model of K68N displays

two hydrogen bonds that are formed with uracil in the mutant, but not the wild type. Increased uracil activities are attributed to the increase in hydrogen bonds in the K68N-uracil complex.

Effects of Mutating S22

Even though the S22 residue is not in close proximity to the MUG active site, mutations of the S22 position greatly affect the MUG activity. As stated above, *E. coli* MUG was determined to be active for xanthine and uracil. Mutagenesis experiments at the S22 position resulted in interesting enzyme activities (Table 4.2), specifically for the inosine and oxanine substrates. In hTDG, the corresponding S22 residue is a methionine, and in *S. Pombe* TDG the corresponding residue is a threonine. Therefore, the S22M and S22T mutations were created to mimic the hTDG and *S. Pombe* TDG enzymes. These mutations did not affect the activities of xanthine or uracil base pairs, however there were slight increases in activity for the GI and GO base pairs. It was also determined that S22M and S22T strengthened the binding of the enzyme for the GI and GO substrates. According to Eq. 4.1, increasing the binding affinity (K_{bind}) would enhance the enzyme activity. Assuming the binding affinity increases independent of the complementary base, the base pair with the lowest base flipping free energy difference (highest K_{op}) will have the highest activity increase. Least stable base pairs have a higher concentration of the flipped out base than the more stable base pairs. Figure 4.10 displays the base flipping PMF profiles (refer to procedures in chapter 3) for inosine and oxanine with the four complementary bases (G,A,C,T). The profiles demonstrate that GI and GO base pairs have the lowest base flipping free energy differences for inosine and oxanine. Therefore, the increase in GI and GO activity observed in the S22 mutants can be attributed to the base flipping free energy difference.

	wtMUG	hTDG	SpoTDG	S22M	S22T	S22L	S22V	S22E	S22F	S22Y	S22I
Xanthine	A		A	A	A	A	A	A	A	A	A
Uracil	A	A	A	A	A	A	A	A			
Inosine		A	A	SA	SA	A	A	A			
Oxanine			SA	SA	SA	A	A	A			

Table 4.2 – Activities for S22 mutations. The activity is indicated by A = high activity , SA = slight activity, and a blank = no activity.

Several other mutation experiments were made at the S22 position, and resulted in extreme changes in MUG activity. The S22L, S22V, and S22E increased activity on most, if not all of the inosine and oxanine base pairs. There were three S22 mutations (S22F, S22Y, S22I) that did not increase the inosine or oxanine activity, but did completely remove uracil activity.

Changes in Protein Dynamics

The S22 residue is not located near the active site of MUG, and for that reason must affect the overall activity by altering the equilibrium conformation. After generating an ensemble for the unbound wild type, the mean square fluctuations (MSF) were calculated to determine the flexibility of the protein. Flexibilities per residue for *E. coli* MUG are shown in Figure 4.11A. The regions that have the greatest MSF values are the most flexible, and correspond to the loop regions in MUG.

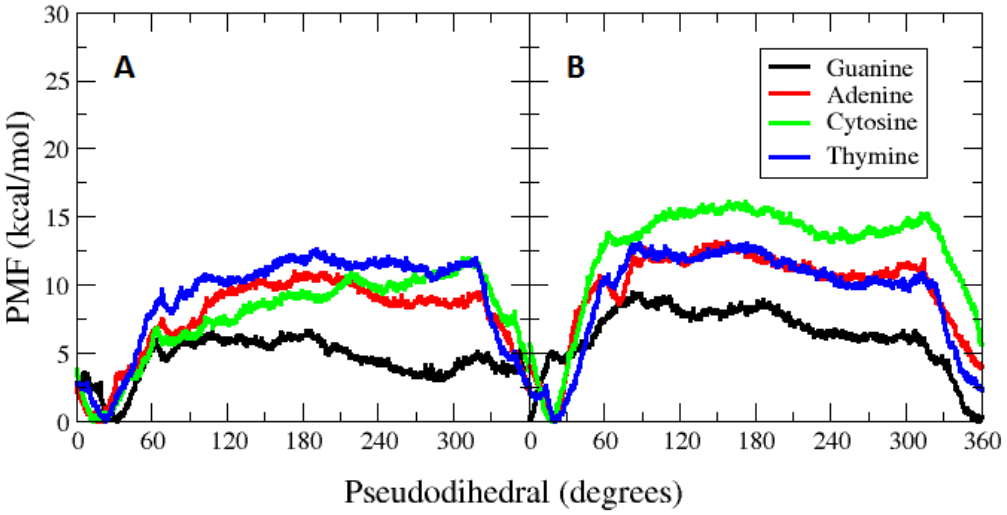


Figure 4.10 – Potentials of mean force (PMF) of oxanine- containing **A**. and inosine-containing **B**. base pairs along the pseudodihedral angle coordinate. TIP3P explicit solvent used during the umbrella sampling simulations. Watson-Crick base pairing is approximately 10°-30° pseudodihedral angle and the flipped out state is approximately 190° (line).

Similar to allosteric effects, perturbing the S22 position indirectly influences the enzyme-substrate interactions. Differences in mean square fluctuations (ΔMSF) were calculated (Figure 4.11B) over unbound protein MD trajectories, in order to determine the effect of mutation on the local flexibility. MSF was calculated per residue, and the difference between the wild type and the mutation was then determined. The flexibility of the active site residues (16-30, 68, 140) were only slightly affected by these mutations. We found that the primary effect of the S22 mutations was on the flexibility of three

regions (residues: 73-79, 109, 120) that interact with the DNA non-specifically. In figure 4.11, it is shown that most of the S22 mutations increase the flexibility of these regions. These regions are not in direct contact with residue S22, therefore the correlation between S22 and the flexibilities of these distal residues was hypothesized to be similar to an allosteric relationship. As demonstrated by Mukherjee *et al.*[149], the correlation between large protein motions could indicate a means for distal residues to have an influence on the enzyme function. Therefore, in order to reveal which regions of MUG have potential for influencing the flexibilities of distal residues correlated motion of the lowest frequency modes was determined.

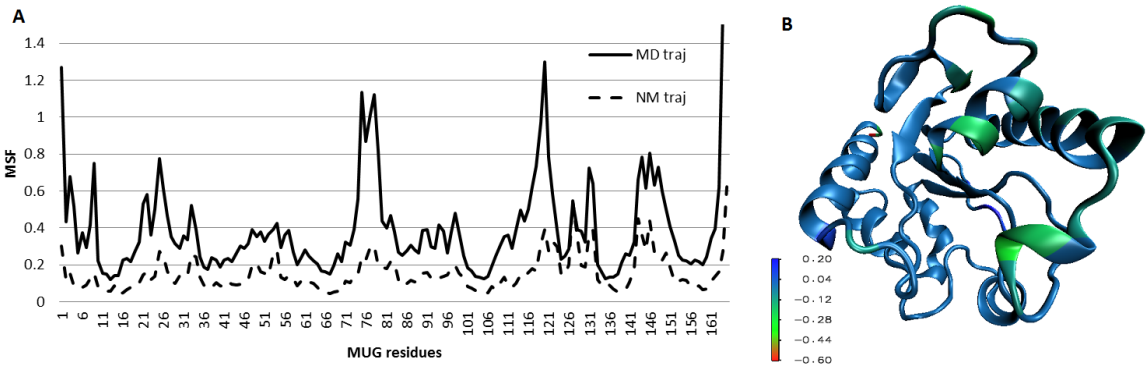


Figure 4.11 – Flexibility analysis of wtMUG and the S22 mutations. **A.**Per residue MSF analysis of wtMUG over free protein MD and NM trajectories. Greater MSF indicates a region of flexibility in the protein. Solid line = MD trajectory Broken line = NM trajectory **B.**Surface map of average flexibility changes post-mutation. The MSF over free protein MD trajectories. Differences between the wild type and S22 mutants were taken per residue, and then averaged over all of the mutants.

Previous studies have shown that enzyme dynamics are central in the mechanisms of recognition and function.[150, 151] Large collective motions in proteins have been studied as dynamics, using normal mode analysis (NMA).[152-154] At low computational cost this method generates low frequency and high frequency normal modes of biomolecules with a harmonic approximation.[152] The low frequency modes are usually the most relevant, since they include large-scale conformational changes of biomolecules.[155] Others have shown that correlated motion can be utilized for connecting the dynamics and activity of an enzyme. Brooks et al. demonstrated correlated motions in the Micaelis-Menton complex of the dihydrofolate reductase enzyme correlated well with the allosteric effects of the protein as well as the changes in activity after point mutations.[156, 157] Correlated motion has not only been used to study allosteric effects of proteins, but also the functional effects of distal mutations on enyzmes.[157] Changes in the negatively correlated motion may be a significant relationship between collective regions within the enzyme. Positively correlated motion is less clear, since those correlations may also be caused by neighboring secondary structure within the same collective region of the protein.[156] A point mutation in one of these correlated regions can alter the recognition or catalysis of the enzyme.

In order to identify coupled regions of the protein, correlated motions were determined for wild type MUG. The changes in flexibility over the S22 mutations revealed specific regions of the protein (residues: 73-79, 109, 120) that were affected most by the mutations. Previous studies have examined coupled motions of proteins over long-time scales calculating the correlated motions over the lowest frequency normal

modes.[149, 158] Here, covariance matrices were constructed for the the lowest frequency normal modes (not including the first six) of wtMUG. It has been established that converged correlated motions are necessary when using NM trajectories.[158] In the current study, covariance matrices converged after the first 200 modes. Positive correlation is represented by the yellow to red colors of the spectrum, and negative correlation is represented by the dark blue regions. Positive correlation implies the two collective regions are moving together in the same direction. Negative correlation indicates the collective regions of interest are moving in opposite directions. Figure 4.12A shows that motion of residue S22 is positively correlated ($S_{ij} \neq 0$) with the 73-77 residue region. This correlation is also one of the few significant (<0.05 P-value) correlated motions of MUG, as can be seen from the P-values (Figure 4.12B). Residue 76 is not conserved across the sequences of the MUG/TDG family. However, this region has conserved hydrogen bond donor side chains, which can interact with the DNA phosphate backbone.

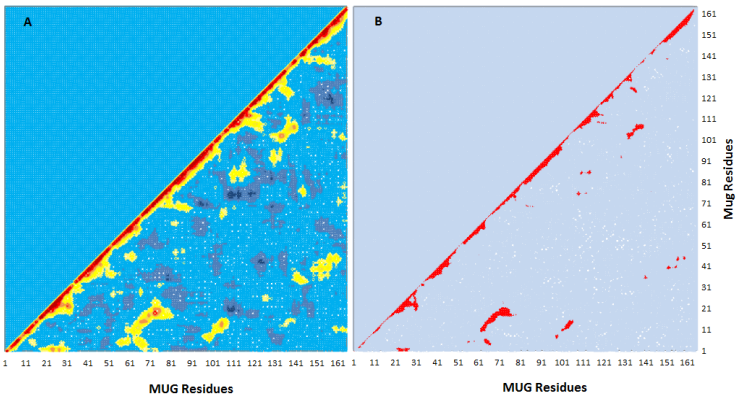


Figure 4.12 – Correlated motion of *E. coli* MUG. **A**. Covariance matrix taken over normal mode trajectory of MUG. yellow to red = positive correlation, light blue to dark blue = anti-correlation. **B**. Two tailed P-values for correlated motion. Red indicates P-values < 0.05.

Covariance matrices were also constructed for select S22 mutations in order to demonstrate the influence of mutations on coupled motion of the enzyme. Firstly, mutations that gained activity on inosine and oxanine (S22V,S22E,S22M) were examined. The correlated motions of the S22 mutants over the first 200 normal modes are displayed in figure 4.13. The significant correlated motions of S22V, S22E, and S22M are very similar to the wild type enzyme (figure 4.12B). The three mutants retain xanthine and uracil activity similar to the wild type. Therefore, it may be necessary for the enzyme to have coupled motions similar to the wild type in order to maintain the xanthine and uracil function. Although the three mutants increase the activity on inosine and oxanine, they do not alter the coupled motion. Increases in inosine and oxanine activity are most likely a result of changes to the protein motion or average structure upon binding to DNA.

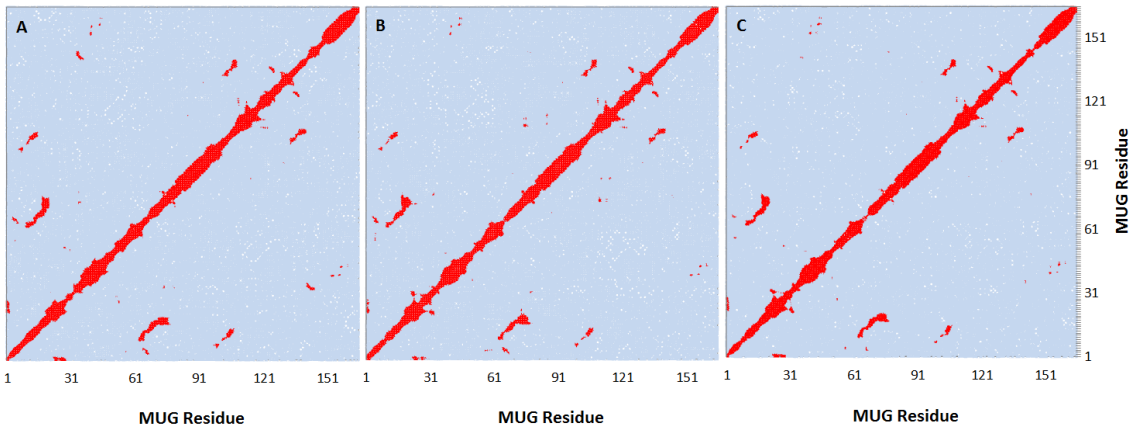


Figure 4.13 – Correlated motion of S22 mutants active for inosine and oxanine. Covariance matrix was taken over normal mode trajectory. Two tailed P-values for correlated motion of **A.** S22M **B.** S22E **C.** S22V were calculated. Where red indicates P-values < 0.05.

Several of the S22 mutants displayed a loss in uracil activity, and the correlated motion was analyzed to connect the protein dynamics to the activity. While the wild type of MUG is active for xanthine and uracil, the mutants S22F, S22Y, and S22I lost their uracil activity. The correlated motions of these three mutants over the first 200 normal modes are displayed in figure 4.14. It can be seen that significant negatively correlated motions (negative correlation not shown) are greater in the mutants than in the wild type enzyme (figure 4.12B). It has already been established that negatively correlated motions have an impact on the activity.[156] Therefore, the increases in the negatively coupled motion could be linked to the loss of uracil function. The S22F and S22Y produce the greatest loss of uracil activity, and also the greatest increases in correlations. Increases in

negatively correlated motion from the wild type ordered from greatest to least are: S22F > S22Y > S22I. These increases are highlighted in figure 4.14. Root mean square deviations (RMSD) from the wild type covariance values follow the same trend. The RMSD values are able to distinguish the S22 mutations that lost uracil activity from those that gained activity on inosine and oxanine. Since there was an increase in correlated motion with the mutants that lost uracil activity, the RMSD values (S22F=0.1521, S22Y=0.1458, S22I=0.1430) are greater than the RMSD values for the mutants that gained activity on inosine and oxanine (S22V=0.1410, S22M=0.1418, S22E=0.1399) This trend correlates well with the loss of uracil activity.

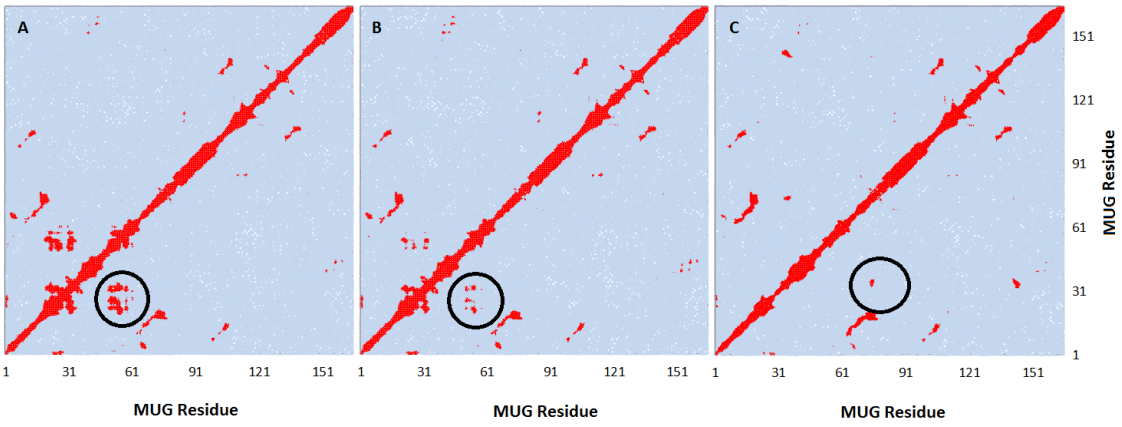


Figure 4.14 – Correlated motion of S22 mutants inactive for uracil. Covariance matrix was taken over normal mode trajectory. Differences from wild type correlated motion circled. Two tailed P-values for correlated motion of **A.** S22F **B.** S22Y **C.** S22I were calculated. Where red indicates P-values < 0.05.

Mba Catalyic Residue

After MUG enzymes are bound to the target base, the N-C1' bond is hydrolyzed, and the damaged base is cleaved out of the DNA helix. From crystal structures[3], we know that each uracil glycosylase has a catalytic residue within the active site, which positions a water molecule into the active site for catalysis with the substrate. A water bridge is formed between the ribose sugar of the damaged base, and the catalytic residue. In the MUG/TDG family, the catalytic residue is an asparagine.[3] Through the water bridge, the aparagine attacks the C1' of the substrate base, and catalyzes the removal of the base (Figure 4.15).

An investigation was performed on an archeal MUG enzyme (*M. barkeri* Mba), and discovered that the catalytic mechanism was distinct from other MUG/TDG enzymes. The corresponding N18 position in Mba is a LEU, which cannot form the necessary water bridge with C1' of the ribose. Mutations of Mba were created through site-directed mutagenesis on all of the possible catalytic residues, in effect any asparagine that was a water bridge distance from the substrate. The Mba was inactive on all substrates when residue N39 (N35 in MUG) was mutated, implying that the N39 residue is required for the cleavage of the substrate base. To understand the details of the interactions between the substrate and residue N39, homology models were created for Mba. Homology modeling is used to create atomic coordinates for a target protein sequence with unknown structure. A sequence alignment with the target sequence is carried out, and sequences similar to the target with a known structure can then be used as the template for the modeling. With the target sequence aligned to the template

sequence, the backbone atoms of the target sequence are generated according to the sequence alignment. Mba was modeled by using a multi-structure template, where all of the similar sequences with known structures were used in the template. The structures included in the template were either TDG or MUG structures (2rba.pdb, 2c2p.pdb, 1wyw.pdb, 1mug.pdb). In comparing the *E. coli* MUG and *M. barkeri* Mba models (Figure 4.15), the position of the catalytic asparagine was unique for each enzyme. The Mba model demonstrated that N39 was oriented in a position that could form a water bridge with the C1' of the ribose. These results supported the the enzyme activity results, which showed the N39 residue was the catalytic residue for Mba.

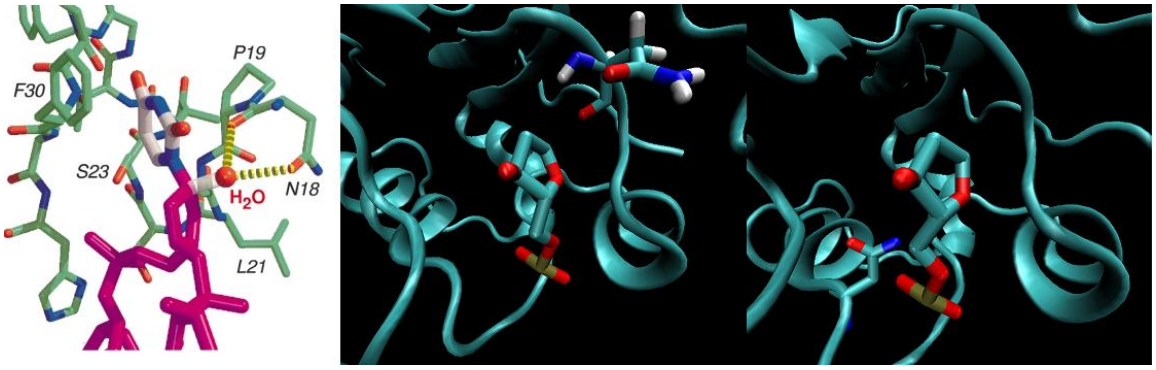


Figure 4.15 – Catalytic residues for MUG and MBA. **A**. Crystal structure of MUG. The water bridge between N18 and C1' of the ribose sugar is displayed.[3] **B**. Minimized structure of MUG bound to uracil, highlighting the N18 catalytic residue and the ribose sugar. **C**. Homology model of MBA enzyme bound to uracil, highlighting N39 catalytic residue and the ribose sugar.

Conclusions

Active site interactions of *E. coli* MUG have been analyzed at a molecular level. While wtMUG is active for both xanthine and uracil, it shows specificity for the xanthine substrate. The molecular models and interaction energies of wtMUG indicate that the MUG active site interactions favored xanthine over uracil, which agrees with the experimental activity assays.[2] More specifically, the S23 sidechain in MUG forms a hydrogen bond with N7 of xanthine and provides the specificity for xanthine. In order to confirm its role in MUG recognition, site-directed mutagenesis was performed on the S23 residue. Interaction energies between the S23A mutant and xanthine were less favorable than energies between the wtMUG and xanthine. This confirmed that mutation S23A lacked the hydrogen bond between the S23 sidechain and N7 of xanthine.

Site-directed mutagenesis was performed on the N140 of MUG in order to mimic human TDG and *E.coli* UDG enzymes. It was shown through molecular models that N140 forms hydrogen bonds with the DNA phosphate backbone. Point mutations were created to reproduce the mutagenesis experiments, which included N140M and N140H. N140M removed uracil and xanthine activity, which is most likely due to the loss of hydrogen bond with the DNA backbone. The H187 of the *E.coli* UDG enzyme is critical for both stabilizing the transition state, and forming hydrogen bonds with the DNA phosphate backbone.[136] The N140H mutant lost xanthine activity, but not uracil activity. More details were provided in the molecular model, and displayed a hydrogen bond forming between the sidechain of H140 and the DNA backbone. The specificity with uracil was generated from a hydrogen bond between the histidine and uracil.[2]

The K68N mutation resulted in an increase in uracil activity. This was discovered to be due to the gain of two hydrogen bonds with uracil when lysine is substituted with asparagine. Interaction energies showed that K68N was more favorable than the wild type, and the two hydrogen bonds were clearly observed in the bound uracil-K68N model.

Mutagenesis at the S22 position displayed increases in inosine and oxanine activities. The wtMUG is not active against these substrates, so this was an especially interesting result. It was shown that GI and GO were the most active of the inosine and oxanine base pairs. Base flipping PMF profiles of inosine and oxanine demonstrated that GI and GO have the lowest barrier for base flipping. In the MUG enzyme, the base flipping equilibrium plays a significant role in the recognition of damaged bases. S22 is not close enough to the active site, where it could affect the active site interactions directly. However, the correlated motion analysis shows that collective regions of the protein are affected by the mutants that remove UDG activity. The changes in correlated motion after mutation, follow the same trend as the changes in uracil activity for these mutants. Disruptions of the large scale motion may be responsible for differences in the uracil activity.

APPENDICES

APPENDIX A

CMG2 DOCKING STUDY

The protective antigen (PA) domain of the anthrax protein binds to the cell surface at the cell surface receptor in order to induce toxicity. PA either binds to tumor endothelial cell marker 8 (TEM8) or capillary morphogenesis protein 2 (CMG2). Previous work has shown that the CMG2 receptor is present during angiogenesis, and because of this, ligands with high binding affinity towards CMG2 (i.e. inhibitor capability) are of great interest. Our collaborator Michael Rogers and Ken Christensen determined 46 ligands that were good binders to CMG2 and 23 ligands that were poor binders to CMG2. With the ability to efficiently predict which ligands are strong binders to the CMG2 receptor, the process of finding an inhibitor for slowing or reducing angiogenesis would be accelerated.

In this study we optimized docking protocols to predict the strong binding CMG2 ligands from the weak binding CMG2 ligands. The docking program AUTODOCK4 was applied to perform docking simulations on all of the provided ligands. Informational entropy is displayed for the docking simulations, and distinguishes the strong binging ligands from the weak binding ligands. Furthermore, analyses of the structural qualities are provided.

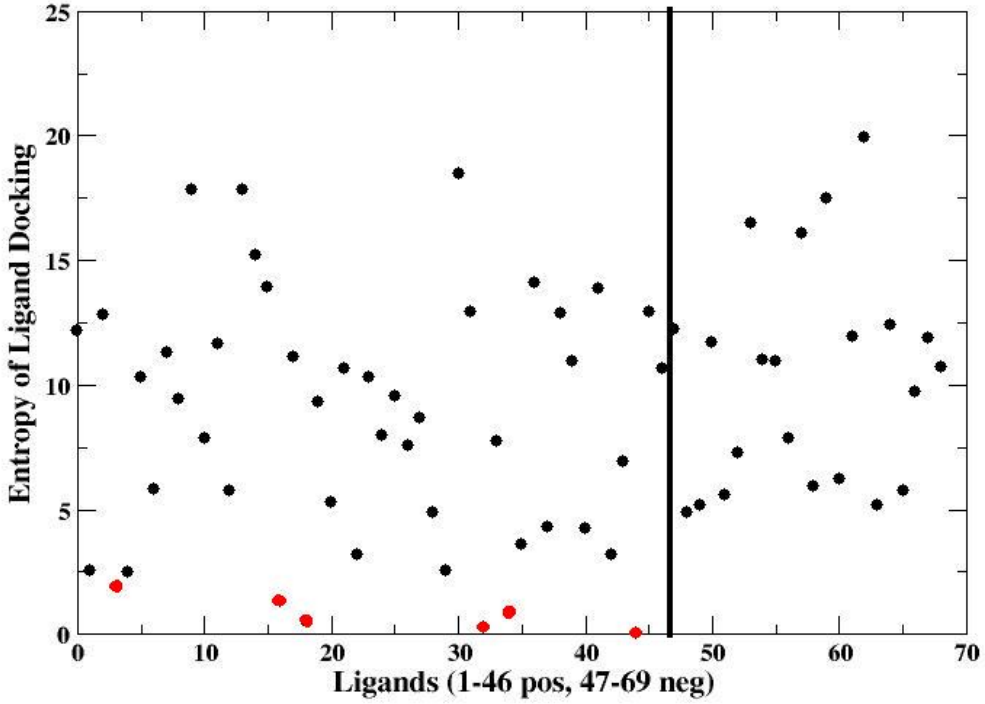
The crystal structure of CMG2 (1SHU.pdb) from the RCSB protein database was used as the receptor during the docking simulations. With the database of ligands our

collaborator provided (69 total) we attempted to distinguish positive hits from negative hits. A fixed protein structure was applied for the docking simulations, while the ligand was flexible. AUTOGRID4 included within the AUTODOCK4 package was used with 80X80X110 grid points, and grid spacing of 0.375Å. Docking runs were performed using the Lamarckian genetic algorithm (LGA), 100 docking runs, and 10000000 maximum energy evaluations.

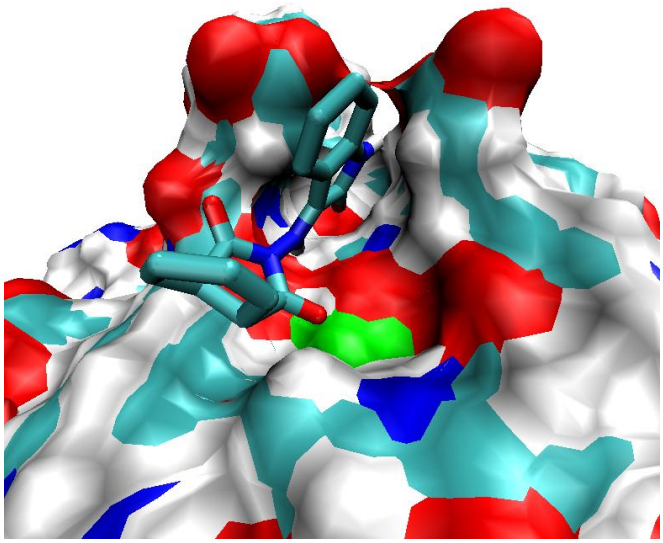
We decided the first step in the project was to validate our docking methods. Docking simulations with AUTODOCK4 had already been performed on this database of ligands by our collaborators. These results were used as a control set. Before optimizing the protocol, the control set was reproduced. A correlation coefficient of 0.95 was achieved with the control.

Using the AUTODOCK4 docking program, we attempted to optimize a protocol for predicting strong CMG2 binding small molecules. It was found that the AUTODOCK4 binding free energy was not capable of distinguishing the good binders from the poor binders. However, it was also observed the CMG2 dockings of the weak binding ligands were less localized on the protein surface than the strong binding ligands. The informational entropy (-Plog(P)) was calculated for both the poor and good binding ligands, and was able to distinguish them. Figure A1 shows the informational entropy for all of the CMG2 ligands, where the ligands with lowest entropy are highlighted in red. The low entropy of these ligands distinguishes them from the weak binding ligands. Details of the structural qualities for the binding interactions of these low entropy binders are displayed in Figure A2. In general, the ligands with low informational entropy had an

acidic group that favorably interacted with the Mg^{2+} in the CMG2. It can also be seen in Figure A2 the low entropy ligands mostly had an aromatic group that interacted with the nearest pocket on the protein surface. These favorable interactions resulted in the localized docking, and low informational entropy.



Appendix A1 - Informational entropy (-Plog(P)) of ligand dockings on surface of CMG2 receptor protein. Autodock used for docking of 69 total ligands, where 1-46 were known positive hit ligands, and 47-69 were known negative hit ligands. Highlighted in red are ligands with low entropy.



Appendix A2 - Docked structure of one of the low entropy ligands from figure A1 (ligand 1538E09). Common structural characteristics for the low entropy dockings were discovered. An acidic group interacts favorably with the Mg^{2+} (green residue), and there is also an aromatic group that interacts favorably inside nearest pocket.

REFERENCES

[1] H. Lee, Brice, A.R., Wright, C.B., Dominy, B.N. and Cao, W., Identification of Escherichia coli MUG as a Robust Xanthine DNA Glycosylase, Journal of Biological Chemistry, (*In Press*).
[2] 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.
[3] T.E. Barrett, O.D. Scharer, R. Savva, T. Brown, J. Jiricny, G.L. Verdine, L.H. Pearl, Crystal structure of a thwarted mismatch glycosylase DNA repair complex, EMBO J, 18 (1999) 6599-6609.
[4] L. Michaelis, M.L. Menten, Biochemische Zeitschrift, 49 (1913) 333-369.
[5] D. Thompson, P. Plateau, T. Simonson, Free-Energy Simulationsand Experiments Reveal Long-Range Electrostatic Interactions and Substrate-Assisted Specificity in an Aminoacyl-tRNA Synthetase, ChemBioChem, 7 (2006) 337-344.
[6] J.M.J. Swanson, R.H. Henchman, J.A. McCammon, Revisiting free energy calculations: A theoretical connection to MM/PBSA and direct calculation of the association free energy, Biophysical Journal, 86 (2004) 67-74.
[7] A. Weis, K. Katebzadeh, P. Soederhjelm, I. Nilsson, U. Ryde, Ligand Affinities Predicted with the MM/PBSA Method: Dependence on the Simulation Method and the Force Field, Journal of Medicinal Chemistry, 49 (2006) 6596-6606.
[8] 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.
[9] A.D. Mackerell Jr., L. Nilsson, Molecular Dynamics Simulations of Nucleic Acid-Protein Complexes, Current Opinion in Structural Biology, 18 (2008) 194-199.
[10] W. Saenger, Principles of Nucleic Acid Structure, 1984.
[11] J.M. Vargason, K. Henderson, P.S. Ho, A crystallographic map of the transition from B-DNA to A-DNA, Proceedings of the National Academy of Sciences of the United States of America, 98 (2001) 7265-7270.
[12] L. Yang, B.M. Pettitt, B to A transition of DNA on the nanosecond time scale, Journal of Physical Chemistry, 100 (1996) 2564-2566.
[13] D. Jose, D. Porschke, The Dynamics of the B-A Transition of Natural DNA Double Helices, Journal of the American Chemical Society, 127 (2005) 16120-16128.
[14] T.E. Cheatham, M.F. Crowley, T. Fox, 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.
[15] A.G.W. Leslie, S. Arnott, R. Chandrasekaran, R.L. Ratliff, Polymorphism of DNA double helices, Journal of Molecular Biology, 143 (1980) 49-72.
[16] A.K. Mazur, Electrostatic Polymer Condensation and the A/B Polymorphism in DNA: Sequence Effects, Journal of Chemical Theory and Computation, 1 (2005) 325-336.

[17] R. Rohs, X. Jin, S.M. West, R. Joshi, B. Honig, R.S. Mann, Origins of Specificity in Protein-DNA Recognition, Annual Review of Biochemistry, 79 (2010) 233-269.
[18] C.W. Garvie, C. Wolberger, Recognition of Specific DNA Sequences, Molecular Cell, 8 (2001) 937-946.
[19] N.C. Seeman, J.M. Rosenberg, A. Rich, Sequence-specific Recognition of Double Helical Nucleic Acids by Proteins, PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, 73 (1976) 804-808.
[20] R. Rohs, S.M. West, A. Sosinsky, P. Liu, R.S. Mann, B. Honig, The Role of DNA Shape in Protein-DNA Recognition, Nature, 461 (2009) 1248-1254.
[21] K.A. Sharp, B. Honig, Calculating total electrostatic energies with the nonlinear Poisson-Boltzmann equation, Journal of Physical Chemistry, 94 (1990) 7684-7692.
[22] J.I. Friedman, A. Majumdar, J.T. Stivers, Nontarget DNA binding shapes the dynamic landscape for enzymatic recognition of DNA damage, Nucleic Acids Research, 37 (2009) 3493-3500.
[23] R.B. Setlow, W. Carrier, PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, 51 (1964) 226.
[24] R. Boyce, P. Howard-Flanders, PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, 51 (1964) 293.
[25] S.S. David, S.D. Williams, Chemistry of Glycosylases and Endonucleases Involved in Base-Excision Repair, Chemical Reviews, 98 (1998) 1221-1261.
[26] J.T. Stivers, Y. Jiang, A Mechanistic Perspective on the Chemistry of DNA Repair Gycosylases, Chemical Reviews, 103 (2003) 2729-2759.
[27] C.Y. Cao, Y.L. Jiang, J.T. Stivers, F.H. Song, Dynamic opening of DNA during the enzymatic search for a damaged base, Nature Structural & Molecular Biology, 11 (2004) 1230-1236.
[28] 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.
[29] 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.
[30] L. Dong, R. MI, R.A. Glass, J.N. Barry, W. Cao, Repari of Deaminated Base Damage by *Schizosaccharomyces pombe* Thymine DNA Glycosylase, DNA Repair, 7 (2008) 1962-1972.
[31] R. Mi, L. Dong, T. Kaulgud, K.W. Hackett, B.N. Dominy, W. Cao, Insights from Xanthine and Uracil DNA Glycosylase Activities of Bacterial and Human SMUG1: Switching SMUG1 to UDG, Journal of Molecular Biology, 385 (2009) 761-778.
[32] A.R. Brice, B.N. Dominy, Examining Electrostatic Influences on Base-Flipping: a Comparison of TIP3P and GB Solvent Models, Communications in Computational Physics, (*In Preparation*).
[33] T.E. Cheatham, Simulation and modeling of nucleic acid structure, dynamics and interactions, Current Opinion in Structural Biology, 14 (2004) 360-367.

[34] 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.
[35] K.J. McConnell, D.L. Beveridge, DNA Structure: What's in Charge?, Journal of Molecular Biology, 304 (2000) 803-820.
[36] L.E. Minchenkova, A.K. Shchelkina, B.K. Chernov, V.I. Ivanov, CC/GG contacts facilitate the B to A transition of DNA in solution, Journal of Biomolecular Structure & Dynamics, 4 (1986) 463-476.
[37] W. Saenger, Hunter, William N., Kennard, Olga, DNA conformation is determined by economics in the hydration of phosphate groups, Nature, 324 (1986) 385-388.
[38] X.-J. Lu, Z. Shakked, W.K. Olson, A-form Conformational Motifs in Ligand-bound DNA Structures, Journal of Molecular Biology, 300 (2000) 819-840.
[39] D.P. Arya, Aminoglycoside-nucleic acid interactions: the case for neomycin, Topics in Current Chemistry, 253 (2005) 149-178.
[40] J. Mazur, A. Sarai, R.L. Jernigan, Sequence dependence of the B-A conformational transition of DNA, Biopolymers, 28 (1989) 1223-1233.
[41] M. Banyay, A. Graslund, Structural Effects of Cytosine Methylation on DNA Sugar Pucker Studied by FTIR, Journal of Molecular Biology, 324 (2002) 667-676.
[42] 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.
[43] A.R. Leach, Molecular Modelling: Principles and Applications, 2000.
[44] D. Frenkel, B. Smit, Editors, Understanding Molecular Simulation: From Algorithms to Applications, 1996.
[45] 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.
[46] M.K. Gilson, J.A. Given, B.L. Bush, J.A. McCammon, The statistical-thermodynamic basis for computation of binding affinities: a critical review, Biophysical Journal, 72 (1997) 1047-1069.
[47] P. Kollman, Free energy calculations: Applications to chemical and biochemical phenomena, Chemical Reviews (Washington, DC, United States), 93 (1993) 2395-2417.
[48] J. Srinivasan, T.E. Cheatham, III, P. Cieplak, P.A. Kollman, D.A. Case, Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate-DNA Helixes, Journal of the American Chemical Society, 120 (1998) 9401-9409.
[49] T. Lazaridis, A. Masunov, F. Gandolfo, Contributions to the binding free energy of ligands to avidin and streptavidin, Proteins: Structure, Function, and Genetics, 47 (2002) 194-208.
[50] H. Luo, K. Sharp, On the calculation of absolute macromolecular binding free energies, Proceedings of the National Academy of Sciences of the United States of America, 99 (2002) 10399-10404.
[51] P.A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D.A. Case, T.E. Cheatham, III, Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models, Accounts of Chemical Research, 33 (2000) 889-897.

[52] M.S. Lee, M.A. Olson, Calculation of absolute protein-ligand binding affinity using path and endpoint approaches, Biophysical Journal, 90 (2006) 864-877.
[53] B. Jayaram, M. Sprous, A. Young, D.L. Beveridge, Free Energy Analysis of the Conformational Preferences of A and B Forms of DNA in Solution, Journal of the American Chemical Society, 120 (1998) 10629-10633.
[54] A. Ferrari, Degliesposti, G, Sgobba, M, Rastelli, G, Validation of an automated procedure for the prediction of relative free energies of binding on a set of aldose reductase inhibitors, Bioorganic & Medicinal Chemistry, 15 (2007) 7865-7877.
[55] D.J.B. Price, C.L., III, Modern Protein Force Fields Behave Comparably in Molecular Dynamics Simulations, Journal of Computational Chemistry, 23 (2002) 1045-1057.
[56] A.F. Villa, H.; Wassenaar, T.; Mark, A.E., How Sensitive are Nanosecond Molecular Dynamics Simulations of Proteins to Changes in the Force Field?, Journal of Physical Chemistry B, 111 (2007) 6015-6025.
[57] T. Yoda, Sugita, Y, Okamoto, Y, Secondary-structure preferences of force fields for proteins evaluated by generalized-ensemble simulations, Chemical Physics, 307 (2004) 269-283.
[58] N. Todorova, Legge, FS, Treutlein, H, Yarovsky, I, Systematic comparison of emirical force fields for molecular dynamic simulation of insulin, Journal of Physical Chemistry B, 112 (2008) 11137-11146.
[59] Y. Mu, Kosov, DS, Stock, G, Conformational dynamics of trialanine in water. 2. comparison of AMBER, CHARMM, GROMOS, and OPLS force fields to NMR and infrared experiments, Journal of Physical Chemistry B, 107 (2003) 5064-5073.
[60] 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.
[61] C.J. Cramer, D.G. Truhlar, Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics, Chemical Reviews (Washington, D. C.), 99 (1999) 2161-2200.
[62] M.S. Lee, F.R. Salsbury, Jr., C.L. Brooks, III, Novel generalized Born methods, Journal of Chemical Physics, 116 (2002) 10606-10614.
[63] 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.
[64] B. Dominy, Brooks, CL, Development of a generalized Born model parameterization for proteins and nucleic acids, Journal of Physical Chemistry B, 103 (1999) 3765-3773.
[65] N.K. Banavali, B. Roux, Free Energy Landscape of A-DNA to B-DNA Conversion in Aqueous Solution, Journal of the American Chemical Society, 127 (2005) 6866-6876.
[66] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, Jr., D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, P.A. Kollman, A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules, Journal of the American Chemical Society, 117 (1995) 5179-5197.

[67] N. Foloppe, A.D. Mackerell, 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 (2000) 86-104.
[68] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus, CHARMM: a program for macromolecular energy, minimization, and dynamics calculations, Journal of Computational Chemistry, 4 (1983) 187-217.
[69] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: an N.log(N) method for Ewald sums in large systems, Journal of Chemical Physics, 98 (1993) 10089-10092.
[70] D. Cremer, J.A. Pople, General definition of ring puckering coordinates, Journal of the American Chemical Society, 97 (1975) 1354-1358.
[71] H.C. Andersen, Molecular dynamics simulations at constant pressure and/or temperature, Journal of Chemical Physics, 72 (1980) 2384-2393.
[72] J.P. Ryckaert, G. Ciccotti, H.J.C. Berendsen, Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, Journal of Computational Physics, 23 (1977) 327-341.
[73] W. Im, M.S. Lee, C.L. Brooks, III, Generalized Born model with a simple smoothing function, Journal of Computational Chemistry, 24 (2003) 1691-1702.
[74] 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 (2006) 719-729.
[75] 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.
[76] H.R. Drew, R.M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura, R.E. Dickerson, Structure of a B-DNA dodecamer. I. Conformation and dynamics, Proceedings of the National Academy of Sciences of the United States of America, 78 (1981) 2179-2183.
[77] S.A. Adelman, J.D. Doll, Generalized Langevin equation approach for atom/solid-surface scattering: General formulation for classical scattering off harmonic solids, Journal of Chemical Physics, 64 (1976) 2375-2388.
[78] M.C. Wahl, S.T. Rao, M. Sundaralingam, Crystal structure of the B-DNA hexamer d(CTCGAG): model for an A-to-B transition, Biophysical journal, 70 (1996) 2857-2866.
[79] 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.
[80] 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.
[81] T. Duong, Zakrzewska, K, Calculation of low frequency normal modes for DNA, Journal of Computational Chemistry, 18 (1997) 796-811.
[82] M. Lee, Van der Vegt, NFA, Molecular thermodynamics of methane solvation in tert-butanol-water mixtures, Journal of Chemical Theory and Computation, 3 (2007) 194-200.
[83] H. Qian, Hopfield, JJ, Entropy-enthalpy compensation: Perterbation and relaxation in thermodynamics systems, Journal of Chemical Physics, 105 (1996) 9292-9298.

[84] C. Bertonati, B. Honig, E. Alexov, Poisson-Boltzmann Calculations of Non-specific Salt Effects on Protein-Protein Binding Free Energies, Biophysical Journal, 92 (2007) 1891-1899.
[85] J. Srinivasan, M.W. Trevathan, P. Beroza, D.A. Case, Application of a pairwise generalized Born model to proteins and nucleic acids. Inclusion of salt effects, Theoretical Chemistry Accounts, 101 (1999) 426-434.
[86] J.T. Stivers, Extrahelical Damaged Base Recognition by DNA Glycosylase Enzymes, Chemistry A European Journal 14 (2008) 786-793.
[87] J.T. Stivers, Progress in Nucleic Acid Research and Molecular Biology, 44 (2004) 37-65.
[88] K. Snoussi, J.L. Leroy, Imino proton exchange and base-pair kinetics in RNA duplexes, Biochemistry, 40 (2001) 8898-8904.
[89] U.D. Priyakumar, A.D. MacKerell, Computational approaches for investigating base flipping in oligonucleotides, Chemical Reviews, 106 (2006) 489-505.
[90] 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.
[91] P. Kollman, Free-Energy Calculations - Applications to Chemical and Biochemical Phenomena Chemical Reviews, 93 (1993) 2395-2417.
[92] 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.
[93] 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.
[94] U.D. Priyakumar, A.D. Mackerell Jr., Computational Approaches for Investigating Base Flipping in Oligonucleotides, Chemical Reviews, 106 (2006) 489-505.
[95] T.E. Cheatham, P.A. Kollman, Molecular dynamics simulation of nucleic acids, Annual Review of Physical Chemistry 51 (2000) 435-471.
[96] J. Norberg, L. Nilsson, Molecular dynamics applied to nucleic acids, Accounts of Chemical Research, 35 (2002) 465-472.
[97] 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.
[98] 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.
[99] 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.
[100] 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.
[101] 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.
[102] U. Dornberger, M. Leijon, F. H., High base pair opening rates in tracts of GC base pairs, Journal of Biological Chemistry, 274 (1999) 6957-6962.
[103] 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.
[104] 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.
[105] 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.
[106] 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.
[107] 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.
[108] B. Roux, T. Simonson, Implicit Solvent Models, Biophysical Chemistry, 78 (1999) 1-20.
[109] 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.
[110] K.J. McConnell, D.L. Beveridge, DNA Structure:What's in charge?, Journal of molecular Biology, 304 (2000) 803-820.
[111] K.A. Sharp, B. Honig, Electrostatic Interactions in Macromolecules: Theory and Applications, Annual Review of Biophysics and Biophysical Chemistry, 19 (1990) 301-332.
[112] 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.
[113] 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.
[114] B.H. Zimm, M. Lebret, Counterion Condensation and System Dimensionality Journal of Biomolecular Structure and Dynamics, 1 (1983) 461-471.
[115] M. Born, Z. Phys., 1 (1920) 45-48.

[116] 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.
[117] 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.
[118] W. Im, M.S. Lee, C.L. Brooks, Generalized Born Model with a Simple Smoothing Function, Journal of Computational Chemistry, 24 (2003) 1691-1702.
[119] 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.
[120] N. Banavali, B. Roux, Atomic Radii for Continuum Electrostatic Calculations on Nucleic Acids, Journal of Physical Chemistry B, 106 (2002) 11026-11035.
[121] 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.
[122] 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.
[123] 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.
[124] 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.
[125] Chocholousova J., 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.
[126] T. Darden, D. York, L.J. Pedersen, Journal of Chemical Physics, 98 (1993) 10089.
[127] S. Kumar, D. Bouzida, R.H. Swendsen, P.A. Kollman, J.M. Rosenberg, The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules .1. The Method, Journal of Computational Chemistry, 13 (1992) 1011-1021.
[128] 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.
[129] T. Simonson, Macromolecular electrostatics: continuum models and their growing pains, Current Opinion in Structural Biology, 11 (2001) 243-252.
[130] T. Lindahl, R.D. Wood, Quality control by DNA repair, Science, 286 (1999) 1897-1905.

[131] B.K. Duncan, J.H. Miller, Mutagenic deamination of cytosine residues in DNA, Nature, 287 (1980) 560-561.
[132] T.M. Hitchcock, L. Dong, E.E. Connor, L.B. Meira, L.D. Samson, M.D. Wyatt, W. Cao, Oxanine DNA glycosylase activity from Mammalian alkyladenine glycosylase, J Biol Chem, 279 (2004) 38177-38183.
[133] L.H. Pearl, Structure and function in the uracil-DNA glycosylase superfamily, Mutation Research-DNA Repair, 460 (2000) 165-170.
[134] L.H. Pearl, Structure and function in the uracil-DNA glycosylase superfamily, Mutat Res, 460 (2000) 165-181.
[135] T.E. Barrett, R. Savva, G. Panayotou, T. Barlow, T. Brown, J. Jiricny, L.H. Pearl, Crystal structure of a G:T/U mismatch-specific DNA glycosylase: mismatch recognition by complementary-strand interactions, Cell, 92 (1998) 117-129.
[136] R. Savva, K. McAuley-Hecht, T. Brown, L.H. Pearl, The structural basis of specific base-excision repair by uracil-DNA glycosylase, Nature, 373 (1995) 487-493.
[137] R.M. Werner, J.T. Stivers, Kinetic Isotope Effect Studies of the Reaction Catalyzed by Uracil DNA Glycosylase: Evidence for an Oxocarbenium-ion anion intermediate Biochemistry, 39 (2000) 14054-14064.
[138] G. Slupphaug, C.D. Mol, B. Kavli, A.S. Arvai, H.E. Krokan, J.A. Tainer, A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA, Nature, 384 (1996) 87-91.
[139] S.S. Parikh, C.D. Mol, G. Slupphaug, S. Bharati, H.E. Krokan, J.A. Tainer, Base Excision Repair Initiation Revealed by Crystal Structures and Binding Kinetics of Human Uracil-DNA Glycosylase with DNA, EMBO, 17 (1998) 5214-5226.
[140] D.J. Krosky, F.P. Schwarz, J.T. Stivers, Linear free energy correlations for enzymatic base flipping: How do damaged base pairs facilitate specific recognition?, Biochemistry, 43 (2004) 4188-4195.
[141] A.Y. Lau, O.D. Scharer, L. Samson, G.L. Verdine, T. Ellenberger, Crystal structure of a human alkylbase-DNA repair enzyme complexed to DNA: mechanisms for nucleotide flipping and base excision, Cell, 95 (1998) 249-258.
[142] R. Shapiro, Damage to DNA caused by hydrolysis, in: E. Seeberg, K. Kleppe (Eds.) Chromosome Damage and Repair, Plenum Press, New York, 1981, pp. 3-18.
[143] T. Suzuki, Y. Matsumura, H. Ide, K. Kanaori, K. Tajima, K. Makino, Deglycosylation susceptibility and base-pairing stability of 2'-deoxyoxanosine in oligodeoxynucleotide, Biochemistry, 36 (1997) 8013-8019.
[144] P. Liu, J.A. Theruvathu, A. Darwanto, V.V. Lao, T. Pascal, W. Goddard, 3rd, L.C. Sowers, Mechanisms of base selection by the Escherichia coli mispaired uracil glycosylase, J Biol Chem, 283 (2008) 8829-8836.
[145] S.S. Parikh, C.D. Putnam, J.A. Tainer, Lessons learned from structural results on uracil-DNA glycosylase, Mutat Res, 460 (2000) 183-199.
[146] N. Guex, M.C. Peitsch, SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling, Electrophoresis, 18 (1997) 2714-2723.
[147] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus, Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations, Journal of Computational Chemistry, 4 (1983) 187-217.

[148] J.B. Parker, J.T. Stivers, Uracil DNA glycosylase: revisiting substrate-assisted catalysis by DNA phosphate anions, Biochemistry, 47 (2008) 8614-8622.
[149] S. Mukherjee, Law, S.M., Feig, M., Deciphering the Mismatch Recognition Cycle in MutS and MSH2-MSH6 Using Normal-Mode Analysis, Biophysical Journal, 96 (2009) 1707-1720.
[150] E.Z. Eisenmesser, e. al., Enzyme Dynamics During Catalysis Science, 295 (2002) 1520-1523.
[151] R.M. Daniel, e. al., The Role of Dynamics in Enzyme Activity Annu. Rev. Biophys. Biomol. Struct., 32 (2003) 69-92.
[152] B. Brooks, M. Karplus, Harmonic Dynamics of Proteins - Normal Modes and Fluctuations in Bovine Pancreatic Trypsin-Inhibitor PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, 80 (1983) 6571-6575.
[153] J.P. Ma, M. Karplus, The allosteric mechanism of the chaperonin GroEL: A dyanmic analysis, PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, 95 (1998) 8502-8507.
[154] A.W. Van Wynsberghe, G.H. Li, Q. Cui, Normal-mode analysis suggests protein flexibility modulation throughout RNA polymerase's functional cycle, Biochemistry, 43 (2004) 13083-13096.
[155] I. Bahar, A.J. Radar, Coarse-grained normal mode analysis in structural biology Current Opinion in Structural Biology, 15 (2005) 586-592.
[156] J.L. Radklewlcz, C.L. Brooks, Protein Dynamics in Enzymatic Catalysis: Exploration of Dihydrofolate Reductase Journal of the American Chemical Society, 122 (1999) 225-231.
[157] T.H. Rod, J.L. Radklewlcz, C.L. Brooks, Correlated Motion and the Effect of Distal Mutations in Dihydrofolate Reductase, PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, 100 (2003) 6980-6985.
[158] A.W. Van Wynsberghe, Q. Cui, Interpreting Correlated Motions Using Normal Mode Analysis, Structure, 14 (2006) 1647-1653.