Understanding the Recognition Mechanism of Repair Protein Fpg: Development and Application of Ensemble Approaches to Long Timescale Dynamics

A Dissertation Presented
by

Christina Bergonzo

to
The Graduate School
in Partial Fulfillment of the
Requirements
for the Degree of

Doctor of Philosophy
in
Chemistry

Stony Brook University

August 2012

Stony Brook University
The Graduate School

Christina Bergonzo

We, the dissertation committee for the above candidate for the
Doctor of Philosophy degree, hereby recommend
acceptance of this dissertation.

Carlos Simmerling – Dissertation Advisor
Professor, Department of Chemistry

Orlando Schärer - Chairperson of Defense
Professor, Departments of Pharmacological Sciences and Chemistry

Jin Wang
Associate Professor, Departments of Chemistry and Physics

Adrian Roitberg
Professor, Department of Chemistry
University of Florida

This dissertation is accepted by the Graduate School

Charles Taber

Interim Dean of the Graduate School

Abstract of the Dissertation

Understanding the Recognition Mechanism of Repair Protein Fpg:
Development and Application of Ensemble Approaches to Long Timescale Dynamics

by

Christina Bergonzo
Doctor of Philosophy
in
Chemistry
Stony Brook University
2012

Every day the genome sustains endogenous and exogenous damage, which is reversed by various repair pathways. One of the most prevalent types of damage, 8-oxoguanine (8OG), arises when a normal guanine base is oxidized, resulting in only a two atom difference from guanine in normal DNA. This type of damage is insidious, however; 8OG mispairs with adenine upon replication, giving rise to a transversion mutation from which the original DNA sequence cannot be recovered. To protect the genome from high mutation rates, the bacterial protein Formamidopyrimidine DNA glycosylase (Fpg) is responsible for searching DNA, recognizing the 8OG lesion when it is paired in a canonical Watson-Crick hydrogen bonding pattern opposite cytosine, promoting eversion of the 8OG out of the helix and into its active site where excision occurs. The dynamic aspects of recognition and base eversion have proved difficult to characterize using experiments. In this work, all-atom molecular dynamics simulations are used to model the recognition and repair pathway, and connect structure, dynamics, and energy. Due to the long timescale of these events, a time-independent method called partial nudged elastic band has been developed and used to link two crystal structures and predict the conformational rearrangements along a minimum energy pathway. Umbrella sampling can then be used to calculate the free energy surface along this pathway for transitions of interest. This method is used to probe the groove preference of the everting 8OG base. The simulations indicate 8OG is bound by Fpg in a conformation preferred by the damage, and is displaced towards the major groove. Eversion occurs through the major groove where there is a 3-4 kcal/mol free energy barrier, presenting a lower barrier to eversion than the minor groove pathway, which has a calculated 7-8 kcal/mol barrier.

Dedication Page

This dissertation is dedicated to my husband, Daniel Roe, who encourages me daily to be the best version of myself, and stands by me even when I am not.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Full Form of Abbreviation
8OG	8-oxoguanine
Å	Angstrom
B. st.	*Bacillus stearothermophilus*
BER	Base excision repair
bp	Base pair
COM	Center of mass
DNA	Deoxyribonucleic acid
E. coli	*Escherichia coli*
Fpg	Formamidopyrimidine DNA glycosylase
hOGG1	Human 8-oxoguanine glycosylase
L.l.	*Lactococcus Lactis*
LJ	Lennard-Jones
MEP	Minimum energy path
MMR	Mismatch Repair
MD	Molecular Dynamics
MM	Molecular Mechanics
MMGBSA	Molecular mechanics – Generalized Born Surface Area
NER	Nucleotide Excision Repair
PDB	Protein Data Bank
PHI	Eversion dihedral
PME	Particle Mesh Ewald
PMF	Potential of Mean Force
RMSD	Root mean square deviation
TIP3P	Transferable Intermolecular Potential 3 Points
TMD	Targeted Molecular Dynamics
UDG	Uracil DNA Glycosylase
VDW	van der Waals
WC	Watson-Crick
WHAM	Weighted Histogram Analysis Method

Acknowledgments

This dissertation would not have been possible without the help of so many people in so many different ways. I wish to thank, first and foremost, my advisor, Professor Carlos Simmerling. Carlos has been a thoughtful and patient mentor throughout our many years together, and I am extraordinarily grateful that he accepted me into his lab. He is an engaging teacher, helpful critic, and all around wonderful person to work for. He has provided me an excellent role model as a scientist and successful P.I., and has instilled in me a love for research.

I would like to thank Professor Orlando Schärer, my committee chair, for his guidance and support. Especially, I'd like to thank him for the opportunity to take part in his lab's journal club meetings, which were incredibly informative for me. I would also like to thank Professor Jin Wang for his input during committee meetings throughout my graduate career. I'd like to thank Professor Adrian Roitberg for his good conversation and good advice throughout my graduate career, and for doing me the favor of agreeing to be my outside committee member.

I would like to thank the DNA repair project experimental collaborators Professor Arthur Grollman, Professor Carlos de los Santos and Professor Dimitry Zharkov, who provided exceptional reviews and continually challenged me to look at questions from an alternate point of view. I would like to acknowledge the 3MT-IGERT program for providing funding for most of my graduate career, as well as the NSF, NIH, XSEDE and NY Blue for funding this research. I would like to especially thank Katherine Hughes, Bev Campbell, Eileen Dowd, and Pam Harris for their administrative expertise throughout the years.

The graduate work outlined in this dissertation would never have become a reality if I didn't collaborate with some of the best coworkers anyone could have. I'd especially like to thank Kun Song, who showed incredible patience with me when I first joined the lab and asked him questions ten times a day. AJ Campbell became a good friend during the many years we worked on the DNA repair project together, and I thank him as well. I have been lucky to work with exceptional people, and would like to thank Asim Okur, Lauren Wickstrom, and all of the other past and present members of the Simmerling Lab.

I'd like to thank my friends for being an important part of my life, and especially for their support. Additionally I would like to thank the friends I have made here at Stony Brook, especially Yukiji Fujimoto, Edison Zuniga, and the Rizzo and Green lab members.

My family has rooted for me throughout my graduate career, and knowing I have their love and support has made all the difference in the world. I'd especially like to thank my Mom and Dad, and my brother and sister for their support. Last, I'd like to thank my husband Dan for his understanding and love. Luckily, he has been through this as well and I have been able to learn so much from him.

1 INTRODUCTION

1.1 Biology of DNA Damage and Repair

1.1.1 DNA structure

Deoxyribonucleic acid (DNA) is the biopolymer containing the genetic instructions for all living organisms. These instructions are arranged in genes, with other sequences providing structural or regulatory purposes. Along with RNA (ribonucleic acid) and proteins, DNA is part of the central dogma of biology. Transcription of DNA leads to mRNA processing, and to translation of mRNA by the ribosome to produce an amino acid chain that folds into an active protein. These three macromolecules are essential to all known forms of life.

DNA plays a complex role in life, but its chemical structure is simple and flexible. Each strand of DNA is made up of nucleotides, containing nucleobases attached to a sugar-phosphate backbone, as shown in Figure 1-1a. The sugar ring is a pentose 2-deoxyribose, and each phosphate group links the 3rd and 5th atom of the sugar rings. The nucleobases are heterocyclic aromatic molecules which are either purines or pyrimidines. Pyrimidines contain 4 carbon atoms and two nitrogen atoms at positions 1 and 3. Purines are a pyrimidine base joined with an imidazole ring. DNA consists of four nucleobases: Adenine (A), Thymine (T), Guanine (G), and Cytosine (C) (Figure 1-1b).

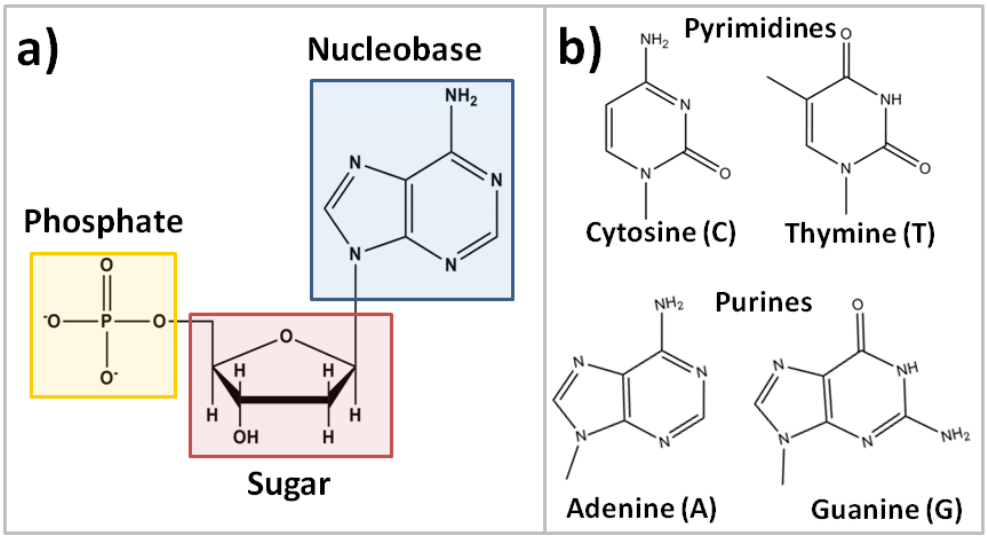

Figure 1-1 a) Nucleotide with base shown in blue, sugar in red and phosphate group in yellow. This is an Adenine nucleotide. b) The four bases which make up DNA: thymine and cytosine, which are pyrimidines, and adenine and guanine, which are purines.

Each DNA strand associates with a complementary one via hydrogen bonds between the nucleobases, running antiparallel to each other and forming a double stranded helix with a right-handed twist (Figure 1-2). Nucleobases pair in a 1:1 ratio combining purines and pyrimidines; A pairs with T and G pairs with C.[1] An A:T base pair forms two hydrogen bonds between complementary strands, and a G:C base pair forms three (Figure 1-2). The aromatic nature of the nucleobases results in favorable pi stacking interactions which further stabilize the helix. The structure of the DNA double helix, reported by Watson and Crick in 1953, is regarded as one of the most important breakthroughs in science.[2]

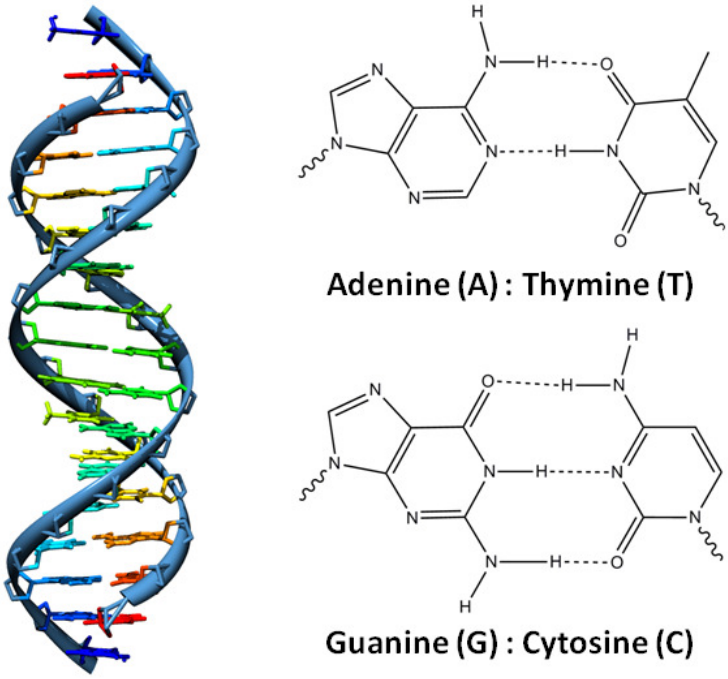

Figure 1-2 Left: DNA in a B-form double helix. Right: Base pair conformations. Adenine and Thymine form two hydrogen bonds, shown in dashed lines, and Guanine and Cytosine form three.

1.1.2 DNA Damage and Repair

DNA is damaged by endogenous and exogenous sources. Endogenous damage arises from replication and recombination errors, metabolic by-products, as well as errors in repair. Exogenous damage is caused by environmental factors, such as UV radiation, free radicals or transition metals.[3] DNA damage causes problems in maintaining the integrity of the genetic code by blocking replication, transcription and causing mutations.[4] Normal cells maintain a balance between damaging agents and repair rates. Abnormal cells cannot maintain repair rates and accumulate DNA damage, leading to mutagenesis, carcinogenesis and age-related diseases.[5]

Repair pathways are in place which combat DNA damage. In general, there are several ways to repair damage which depend on the type of damage sustained by the DNA. Non bulky lesions or damage which affects a single base are primarily repaired via a base excision repair (BER) pathway. In this pathway, a DNA glycosylase recognizes the damage and removes the base, creating an apurinic (AP) site. An AP endonuclease cuts the phosphodiester bond, leaving a one nucleotide gap in the DNA, which is filled by a polymerase and sealed with a DNA ligase. Bulky, helix-distorting lesions are repaired by nucleotide excision repair (NER). This pathway involves many cofactors which play a role in lesion recognition, act as scaffolding proteins, distort the helix, cut the DNA backbone up-and-downstream from the damage, and recruit polymerases to fill the gaps. The third major repair pathway deals with errors made during replication and recombination, and is called mismatch repair (MMR). This locates undamaged but mispaired nucleotides which have been incorrectly incorporated into the DNA.[6; 7]

1.1.3 8-oxoguanine and the GO repair system

The most prevalent result of oxidative stress to DNA is 8-oxoguanine (8OG) damage, occurring on the order of about 1,000 bases per cell per day in eukaryotes. This lesion is of particular interest since it is only slightly different from undamaged G, but promotes transversion mutations.[8] 8OG differs from a canonical G by two atoms, as illustrated in Figure 1-3a: the N7 is protonated and oxygen takes the place of the hydrogen atom at the C8 position. 8OG has the ability to form a Hoogsteen base pair with A, promoting a G to T transversion mutation from which the original sequence is lost.

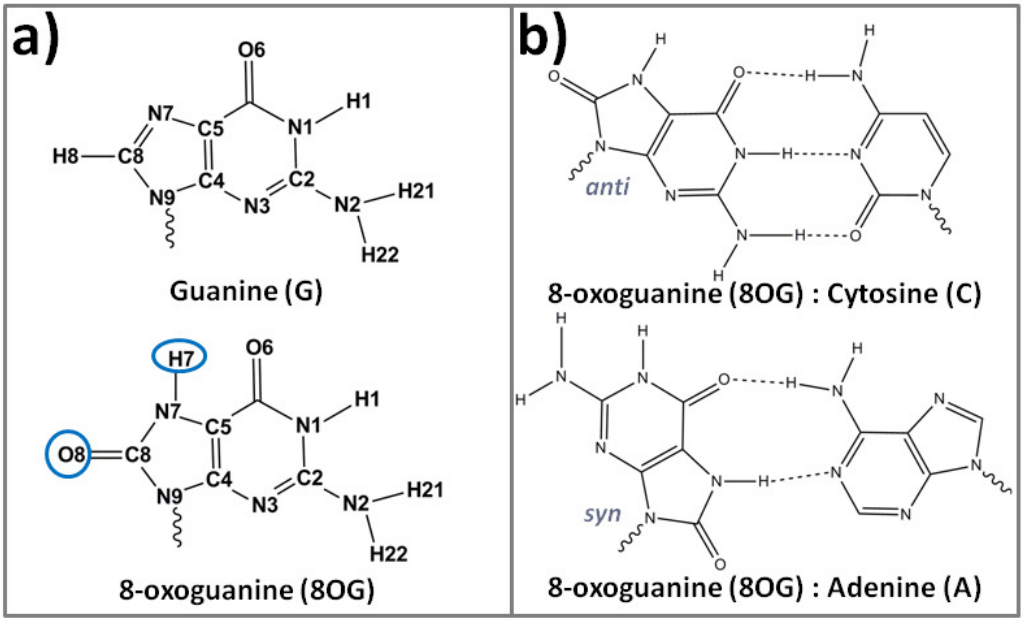

Figure 1-3 a) G (top) is oxidized to 8OG (bottom). There is a two atom difference (highlighted by blue circles) between the undamaged base and the lesion. b) 8OG can form a canonical Watson-Crick base pair with G (top) when it is in the *anti* orientation, and can also form a Hoogsteen base pair with A (bottom) when it is in the *syn* conformation.

DNA repair pathways exist in all forms of life to cope with the highly mutagenic 8OG. This is called the GO repair pathway, and is a form of short patch base excision repair. In bacteria, MutT, MutY and MutM (also called Fpg – Formamidopyrimidine glycosylase) constitute the GO pathway. The human analogs are MTH1, MUTYH, and OGG1. MutT (MTH1) hydrolyses deoxyribose 8-oxoguanine tri-phosphate (d8OGTP) to deoxyribose 8-oxoguanine mono-phosphate (d8OGMP) so it cannot be incorporated into the DNA during replication. MutY (MUTYH) is responsible for excising the A which mispairs opposite the 8OG lesion in a Hoogsteen orientation after replication. The undamaged base is excised in order to maintain the blueprint of the original DNA sequence. The 8OG damage is removed by the glycosylase

Fpg/MutM (OGG1) via cleavage of the glycosyl bond which connects the nucleobase to the sugar ring. Further processing leads to a one nucleotide gap in the DNA. A summary of the GO repair pathway is shown below in Figure 1-4.[9; 10]

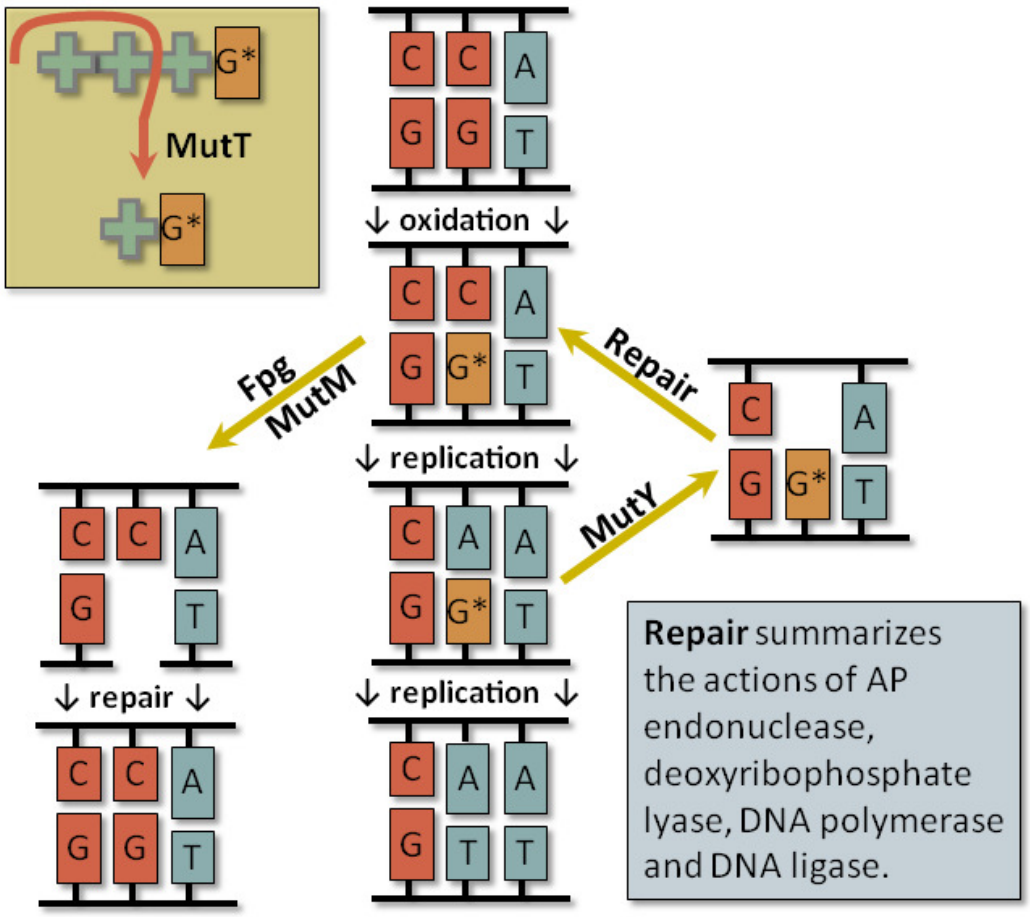

Figure 1-4 Workflow diagram of the GO cycle in prokaryotes. Following the middle column, the G in the central G:C base pair is oxidized to 8OG:C, and after two rounds of replication the sequence is mutated to T:A. To combat 8OG damage before the first round of replication, Fpg/MutM excises the 8OG, and repair of the one nucleotide gap maintains the original sequence (left branch). After the first round of replication, MutY excises the A opposite 8OG (right branch). MutT (shown in the yellow box) acts on d8OGTP, reducing it to the d8OGMP form which cannot be incorporated into DNA by the replicative polymerase.

1.1.4 8OG damage recognition by Fpg

The process by which Fpg recognizes and excises 8OG has been an active area of research for a long time. In 1979 the glycosylase was discovered by Chetsanga and Lindhal to excise alyklated lesions and opened imidazole ring structures.[11] In 1988, the *mutM* mutator locus in *E. coli* was found to promote G:C to A:T transversions.[12] Tchou and Grollman discovered that oxidized substrates were repaired in addition to the previously known opened ring structures, and determined that 8OG was the primary substrate for Fpg/MutM.[13]

Fpg is a bifunctional enzyme, meaning it possesses both glycosylase and AP lyase activity. The function of Fpg is to excise 8OG damage when paired opposite C. The catalytic mechanism for removal of lesions has multiple steps. First, catalysis is initiated by a nucleophilic attack on the C1' of the 8OG deoxyribose ring by the N-terminal proline, as shown in Figure 1-5, Step 1.[14; 15] This forms a covalently bonded Schiff base intermediate (Figure 1-5, Step 2).[16] This ring-opened structure undergoes β and δ elimination reactions which cleave the 3' and 5' phosphate groups, leaving a one nucleotide gap in the DNA, denoted as end products in Figure 1-5, Step 3.[17] The Schiff base intermediate can be trapped using $NaBH_4$ (Step 4) and has been used to generate crystal structures of Fpg bound to DNA.[18]

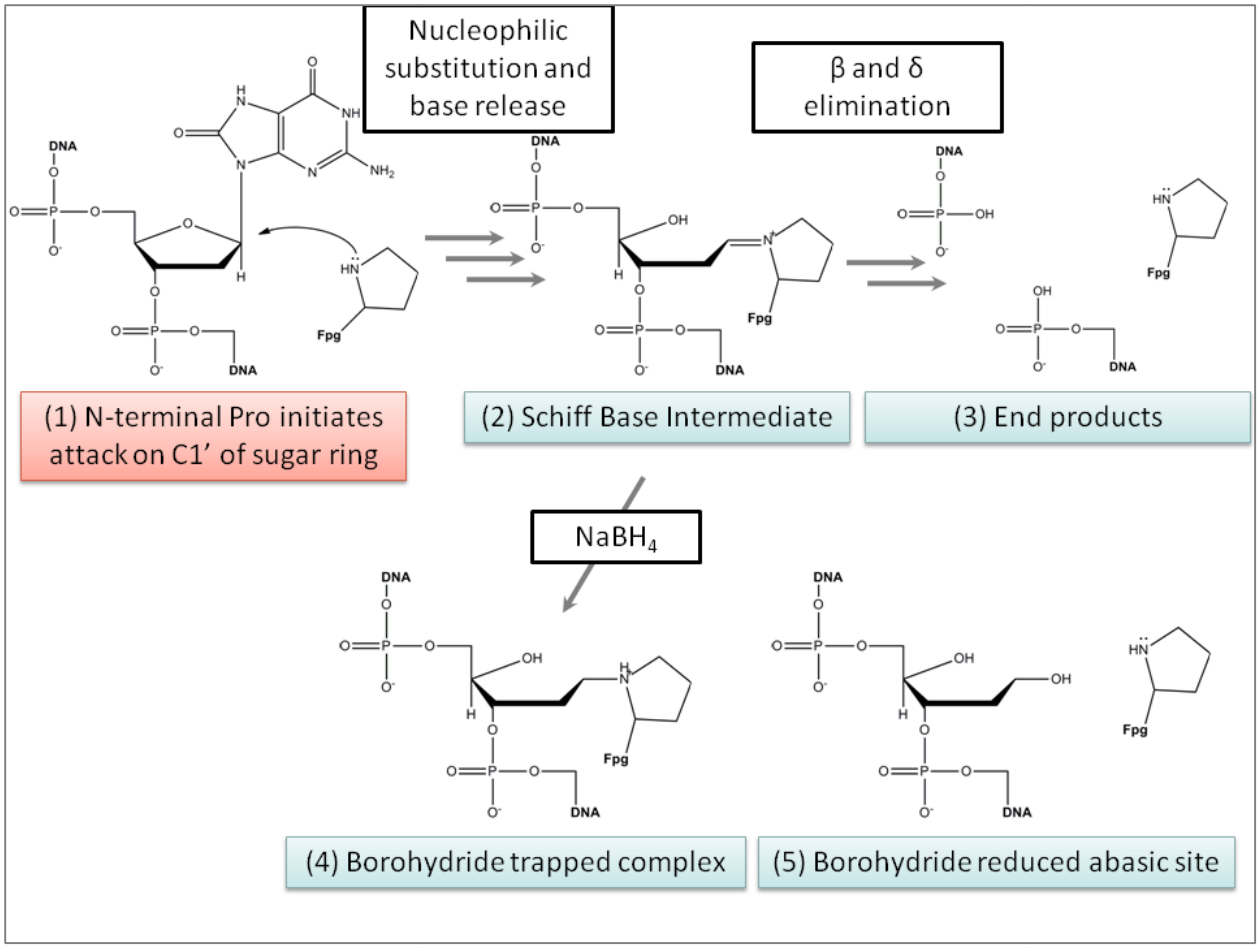

Figure 1-5 Reaction mechanism for Fpg's glycosylase/AP lyase activity. (1) Attack on the C1' of the deoxyribose ring of the 8OG base is initiated by the N-terminal proline. Glycosylation occurs via nucleophilic substitution, followed by base release. (2) A Schiff base intermediate is formed, covalently linking the protein and DNA prior to β and δ elimination reactions, resulting in the end products shown (3). By adding sodium borohydride, a covalent intermediate after base release can be trapped (4), when reduced it produces an abasic site (5).

The structure of apo Fpg and Fpg bound to DNA has been solved under diverse conditions. In general, the protein contains around 270 residues, depending on the species. Fpg is a two domain protein with a central hinge region where the N-terminal is located and where catalysis takes place (Figure1-6). DNA binds in this central cleft via zinc finger and helix two turn helix DNA binding motifs. Upon binding a 60^{o} bend in the DNA occurs. Here, the 8OG base is trapped via a catalytic mutant.[19] Hydrogen bonds formed between the O6 of the 8OG base and the H of the backbone amide groups of the local amino acid residues orients the base in the active site pocket beneath the βF-α10 loop. A summary of available crystal structures is given in Table 1-1.

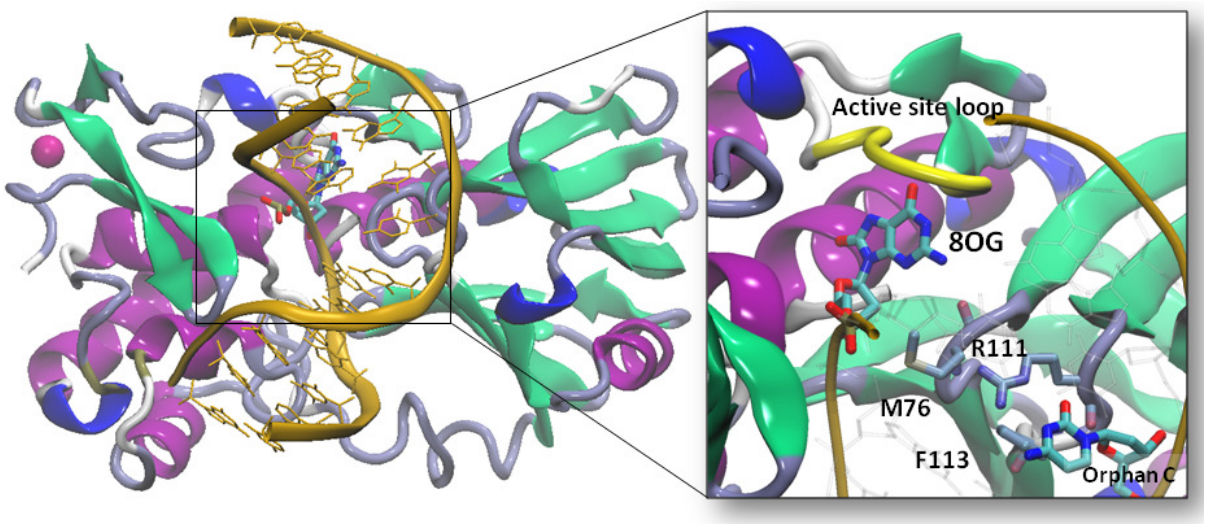

Figure1-6 Structure of Fpg with 8OG base trapped in active site, PDB entry 1R2Y. The DNA is shown in orange, with the everted 8OG base colored by atom name. Zinc from zinc finger DNA binding domain is colored pink and is in the upper left-hand corner. Right: Close-up of the active site bound 8OG. Catalytic loop is in yellow, intercalating wedge residues M76/R111/F113 and orphaned cytosine, which previously base paired with 8OG, are labeled.

Table 1-1 Summary of available crystal structures of Fpg/MutM

Structure	Method	Target base position	Binding loop	Other important features	PDB Reference
Apo Fpg *Thermus thermophilus*	--	--	Ordered	Helix-two-turn-helix and zinc finger DNA binding domains identified	1EE8[20]
Fpg bound A:T *Bacillus stearothermophilus*	Disulfide crosslink	Intrahelical	Disordered	F113 intercalated into helix, breaks base pair planarity	2F5P[21]
Fpg bound G:C *Bacillus stearothermophilus*	Disulfide crosslink	Intrahelical	Disordered	F113 intercalated into helix, breaks base pair planarity	2F5O[21]
Fpg bound 8OG:C *Bacillus stearothermophilus*	Disulfide crosslink and deleted active site binding loop	Intrahelical	--	F113 intercalated into helix, breaks base pair planarity	3GPU[22]
Fpg bound 8OG:C *Bacillus stearothermophilus*	E3Q catalytically inactive mutant	8OG everted and bound in active site	Ordered; amide groups form h-bonds to O6 of 8OG base	F113, M76, R111 intercalating triad fills gap left by everted base	1R2Y[19]
Fpg bound Abasic:C,T,G *Bacillus stearothermophilus*	$NaBH_4$ reduced abasic site	Post-excision	Disordered	Extensive phosphate contacts at lesion site (R263)	1L1T, 1L2C, 1L2D[23]
Fpg bound Schiff Base intermediate *E. Coli*	$NaBH_4$ trapped complex	Post-excision	Disordered	F113, M76, R111 intercalating triad,	1K82[18]

Two of the most relevant structures for this study summarize the differences in Fpg binding to an undamaged G (2F5O[21]) versus binding to 8OG damage (1R2Y[19]). The 8OG complex was captured using a catalytically inactive mutant (E3Q) to trap the 8OG base in an extrahelical, active site bound conformation. This can be considered the pre-catalytic conformation of the protein bound to its substrate. The βF-α10 active site loop is ordered here and orients the 8OG base in the binding pocket. Intercalating residues Met76, Phe113, and Arg111 invade the duplex, with Arg111 specifically forming hydrogen bonds to the orphaned

cytosine, the former Watson-Crick base pairing partner of 8OG. The G complex was crystallized using a disulfide crosslink tethering the protein (Q166C) to the phosphate group of the DNA 1-2 base pairs away from the target G:C base pair. In this structure the βF-α10 loop is disordered. Met76 and Arg111 are not intercalating the helix, but Phe113 intercalates into the duplex and interrogates the target G:C base pair, forcing it to buckle and inducing an overall bend in the DNA.

Prior to catalysis, Fpg must recognize its substrate 8OG:C base pair from an excess amount of undamaged DNA. The current model, based on stopped-flow fluorescence measurements of the intrinsic tryptophan fluorescence of Fpg, identify six reversible and irreversible steps, each with different reaction rates. Initial binding of Fpg shows no discrimination for the opposite base, indicating this step does not contribute to recognition.[24] Different substrates diverge early in the pathway at the second measured rate. This step is attributed to base eversion. The 8OG:C base pair is distinguished by a fast forward rate at this step and backwards rate two orders of magnitude lower. Subsequent rates are attributed to additional conformational rearrangements which aid in filtering out incorrect substrates.[25] Additionally, stopped-flow fluorescence measurements of 2-aminopurine labeled DNA correlated helix distortion with the initial fast rate, attributing it to base eversion. Further DNA distortion correlates with slower protein reaction rates, and is attributed to plugging of the helix with the hydrophobic wedge residues (Met/Arg/Phe).[24] It is clear from these kinetics experiments that many conformational rearrangements occur, which play a major role in the repair pathway of 8OG by Fpg.

The current kinetic model involves a) binding to a non-specific DNA sequence, b) scanning the DNA (sliding) via facilitated diffusion, c) damage recognition, d) formation of a

transient Fpg-DNA complex via stabilizing protein-DNA interactions, e) eversion of the 8OG base from the helix and f) binding the 8OG base in the active site.[26] The irreversible rate limiting chemical steps of glycosylation and β and δ elimination, are then performed.[25] Experimental understanding of this complex mechanism lacks an atomistic view of the transient high energy low populated states during conformational rearrangement of the Fpg-DNA complex. These states are difficult to experimentally probe, but have long been considered important intermediates along recognition and base eversion paths. In this work molecular dynamics simulations are used to explain these aspects of damage recognition at the atomic level, where experiment has proved limiting.

1.2 Molecular Dynamics

Molecular dynamics (MD) simulations are widely used to study various phenomena in biology. Briefly, MD simulations solve Newton's equations of motion for a system where the interactions between particles are described by a force field. The resulting trajectory is a compilation of structural changes over time. Traditionally, the amount of time covered in a molecular dynamics trajectory is on the order of picoseconds to nanoseconds; currently microsecond and millisecond trajectories are attainable for small systems.[27] However, advanced methods can approximate an ensemble in a time-independent manner. Simulations have been used to qualitatively and quantitatively analyze a wide variety of biomolecular systems.[28]

The first paper in which MD simulations were performed was by Alder and Wainwright in 1957 to calculate the phase diagram for a hard sphere system.[29] A simulation on liquid water was carried out in 1974 by Rahman and Stillinger, which reproduced thermodynamic properties

and yielded accurate molecular structure.[30] The first biomolecular simulation was performed in 1977 on the bovine pancreatic tripsin inhibitor protein (BPTI) by McCammon et al.[31] This nine picosecond simulation performed *in vacuo* demonstrated that dynamics are an important aspect of biomolecular structure and function. Currently, MD simulations are widely used in the study of protein structure and function, nucleic acid and lipid systems, as well as in the refinement of experimental x-ray crystallography and NMR structure determination.

1.2.1 The Ergodic hypothesis and Basic Physics

The approximation that is assumed in MD simulation is called the Ergodic hypothesis, and it allows for the comparison of simulation data to bulk experimental data. It states that the Boltzmann distribution of conformations for a system is sampled over a long enough period of time. By looking at the conformations that one structure samples, eventually this one structure will sample all possible conformations. The average of all conformations sampled in a long enough MD trajectory will be equivalent to the conformations present in bulk solution, equating the microscopic MD simulation with the macroscopic experiment.

MD simulations solve Newton's second law of motion for the atoms in a particular system. Specifically, if F_i is the force exerted on each atom *i*, m is its mass and a_i is the acceleration of particle *i*,

$$\vec{F_i} = m_i \cdot \vec{a_i} \tag{1-1}$$

the derivative of *a* with respect to time (*t*) yields

$$\vec{F_i} = m_i \cdot \frac{d\vec{v_i}}{dt} \tag{1-2}$$

where *v* is the velocity.

The second derivative of *a* yields

$$\vec{F_i} = m_i \cdot \frac{d^2\vec{x}_i}{dt^2} \tag{1-3}$$

where x_i is the coordinate position of the atoms in the system. The force acting on the *i*th atom at a given time can also be expressed as the gradient of the interatomic potential energy,

$$\vec{F_i} = -\vec{\nabla}_i V \tag{1-4}$$

which is a function of the positions of all the atoms. Combining these equations yields,

$$-\vec{\nabla}_i V = m_i \cdot \frac{d^2\vec{x}_i}{dt^2} \tag{1-5}$$

where the derivative of the potential energy is related to the changes in atomic position as a function of time.

The MD method is deterministic; all subsequent positions depend on the initial position, which is usually taken from experimentally solved x-ray crystallography structures or NMR structures. While the MD method is rigorous in that the equations are numerically solved, the approximation of the forces, calculated from the gradient of the potential energy function, determines in part how comparable the simulation is to real life. A simulation can only be as realistic as the potential energy function used to reproduce the behavior of the system.

1.2.2 Force Fields

The force field defines the potential energy function used to describe the chemical and physical interactions of a system. Pairwise additive force fields rely on modular, additive parameters to describe the chemistry of a system using physics based potentials. The original all

atom Amber force field for proteins and nucleic acids undertook parameterization of torsion and angle parameters, deriving them from experimental data.[32] This was a precursor to the current equation used to describe the potential energy, introduced in Cornell et al. and shown in Figure 1-7. [33]

$$V\left(r^N\right)=\sum_{\text{bonds}}\frac{1}{2}k_b\left(l-l_0\right)^2+\sum_{\text{angles}}\frac{1}{2}k_a\left(\theta-\theta_0\right)^2+$$

$$\sum_{\text{torsions}}\frac{1}{2}V_n\left[1+\cos\left(n\omega-\gamma\right)\right]+$$

$$\sum_{j=1}^{N-1}\sum_{i=j+1}^{N}\left\{\varepsilon_{i,j}\left[\left(\frac{r_{0_{ij}}}{r_{ij}}\right)^{12}-2\left(\frac{r_{0_{ij}}}{r_{ij}}\right)^{6}\right]+\frac{q_iq_j}{4\pi\varepsilon_0 r_{ij}}\right\}$$

Figure 1-7 The current form of the AMBER additive force field.

The contributions of all atoms in a system can be described by bonded (bond, angle, dihedral) and non-bonded (electrostatic and van der Waals) terms. The above equation defines the potential energy of the system (V) based on the positions of the atoms (r^N). The first term uses a harmonic potential to describe bond stretching, where the energy increases as the bond length (l) deviates from the reference length (l_0). The second term describes angle bending in a similar manner, using a harmonic potential. The third term models the torsional potential, how the energy changes as you rotate around a bond. The fourth term represents the non-bonded energy between all atom pairs. The first summation uses the Lennard-Jones 6-12 potential to describe the van der Waals interaction, where r_{0ij} is the separation at which the energy is at a minimum and ε is the well depth. It combines an attractive (r_{ij}^{-6}) and repulsive (r_{ij}^{-12}) part. The

second summation describes the electrostatic, or charged, interactions. It uses Coulomb's law to calculate the sum of interactions between pairs of point charges (q_i and q_j). Restrained electrostatic potential (RESP) fitting has been used to determine the point charge associated with each atom.[34; 35]

Subsequent revisions to the original ff94 parameters have primarily addressed the dihedral terms[36] or charge models[37] of amino acid residues. Current reviews and comparisons have emphasized the excellent performance of ff99SB, developed in the Simmelring lab, in reproducing experimental data. [38; 39; 40] Nucleic acid force fields have evolved as well, with better descriptions of sugar pucker and chi torsional parameters in ff99.[41] The most recent improvements to the Amber force field for nucleic acids are the parmbsc0 (or Barcelona) corrections which refit the alpha and gamma backbone dihedrals using high level QM calculations.[42]

Nucleic acid force field development has traditionally lagged behind protein force field development. However, current force fields prove adept at handling general structures for nucleic acids in different solvent environments.[43; 44] The conformational stability and variability of DNA in a canonical B form duplex has been evaluated over a 1 μs simulation using ff99 and the parmbsc0 corrections.[45] Timescales on the range of 50ns to 100ns have also been considered in analyzing the conformational effects of nearest-neighbor stacking, indicating convergence of conformational properties on this timescale.[46]

1.3 Advanced Simulation Methods

One of the biggest challenges in MD simulation is reconciling the limitations of timescales accessible in simulation (femtosecond to microsecond) with the timescales of biologically interesting problems (femtosecond to seconds/minutes).[47] For example, the process of interest in this work, Fpg-induced eversion of 8OG damage into the active site, occurs on a millisecond timescale.[48] Many advanced simulation methods have been developed to approximate the ensemble of structures for a given system. These techniques typically exploit multi-processor computers by distributing parts of the calculation, since a converged ensemble and not a liner trajectory is desired. Two such methods are the Nudged Elastic Band (NEB) method and Umbrella Sampling.

1.3.1 Nudged Elastic Band Simulation

Acknowledgements

The material presented in this chapter contains direct excerpts from *An Overview of String-Based Path Sampling Methods* by Christina Bergonzo and Carlos Simmerling, published in the Annual Reports in Computational Chemistry in 2011, Volume 7, Chapter 4, pages 89-97. The manuscript was written by Christina Bergonzo with suggestions and revisions by Carlos Simmerling.

One of the most relevant problems currently in computational chemistry is how to study conformational rearrangements between energy minima which occur along a lowest energy path. This path is referred to as the minimum energy path (MEP) of a conformational transition.

Mathematically, it is defined as the path (φ) connecting two states where the potential normal to the path, $(\nabla V)^{\perp}$ is equal to zero (Equation 1-6).

$$(\nabla V)^{\perp}(\varphi) = 0 \qquad (1\text{-}6)$$

The energy maximum along this path is the transition state or saddle point, an important value for defining the activation energy barrier. The MEP, once found, helps define reaction paths of both chemical reactions and complex conformational changes.

Approaches to finding minimum energy transitions and locating saddle points can generally be grouped into two families: surface walking algorithms and interpolation algorithms. Surface walking methods involve beginning in one energy minimum and proceeding stepwise along the slowest ascent path of the surrounding energy landscape.[49] These methods only require knowledge of reactants, but must be serialized which limits their performance for complex systems. Additionally, when the conformational change is known, an ensemble of paths can be generated to calculate reaction rates by transition state path sampling.[50; 51]

A solution to generating the minimum energy path (MEP) of a transition without a priori knowledge of the path is to use an interpolating method. In such methods, the reactant and product configurations are required and used as fixed endpoints of the path. Geometric configurations of the system are then generated in an interpolating path between these endpoints. The entire path is optimized at once.

Finding the MEP using a chain of connecting images was originated by Elber & Karplus in the plain elastic band method.[52] Restrained endpoint structures in energy minima were copied and linked along an interpolating path. First-derivatives were used to optimize the path between structures. Spring forces were added between the images which force each image to remain at an average separation between its partner images along the current path. The plain elastic band

method is, however, highly dependent on the initial path chosen. This leads to local rather than global minimum energy pathways. Elber and Karplus proposed using simulated annealing as an optimization technique to find the global minimum energy path.[52]

The plain elastic band's spring forces, used to keep the images in the pathway evenly spaced, interfere with the energy of each independent image, resulting in poor transition state geometries and incorrect energies along the path. For a rigid spring constant, the images along the path over-estimate the energies in the saddle point region, leading to corner cutting and preventing the path from resolving saddle point structures. For a weak spring constant, the forces on each image from the force field are not modulated and images slide down the path back toward the minima, and do not resolve the saddle points. [53]

The nudged elastic band (NEB) method was developed to address these problems. [53; 54; 55]

$$\vec{F}_i = \vec{F}_i^{\perp} + \vec{F}_i^{\parallel} \tag{1-7}$$

$$\vec{F}_i^{\perp} = -\vec{\nabla}V(\vec{P}_i) + \left(\left(-\vec{\nabla}V(\vec{P}_i)\right)\cdot\vec{\tau}\right)\vec{\tau} \tag{1-8}$$

$$\vec{F}_i^{\parallel} = \left[\left(k_{i+1}\left(\vec{P}_{i+1}-\vec{P}_i\right) - k_i\left(\vec{P}_i-\vec{P}_{i-1}\right)\right)\cdot\vec{\tau}\right]\vec{\tau} \tag{1-9}$$

The NEB approach, outlined in the above equations, calculates a tangent vector (τ) to the current path defined by the images, and is used to decompose the force described by the force field, $-\nabla V(P_i)$, and that described by the spring forces, k_i, into perpendicular ($F_i^{\perp}$) and parallel ($F_i^{\parallel}$) forces with respect to the path. [56] Endpoint images in local minima are restrained, and copies of these images are pulled into an interpolating path between the two endpoint structures and used as an initial guess at the pathway. Springs which serve to maintain spacing of images connect each image to its neighbor images along the path, and are only described by the parallel part of the force. The force field potential is then only applied orthogonal to the path tangent. That is, it

is projected out from each image and not along the path between images. Optimization is performed to minimize energy. [57; 58]

The climbing image is a modification to the NEB method which refines the transition state structure.[59] The highest energy image along the NEB path is determined, and the force at this image along the tangent is inverted. This image's energy is maximized along the path and minimized in all other directions, converging to an exact saddle point. Variable spring constants are used to increase parallel forces near the saddle point, helping to resolve this image.

A variation on the NEB method which seeks to improve the efficiency of saddle point optimization was addressed with the adaptive nudged elastic band.[60] Instead of increasing the number of images along the path to encourage better resolution of the saddle point, adaptive NEB performs iterations of NEB where the two images neighboring the highest energy image are reassigned as new endpoints for NEB. In this way, the saddle point energy is adaptively optimized.

Recently, a doubly nudged elastic band method was introduced to address the instability arising from complete removal of the spring gradient perpendicular to the path during optimization. The answer was to include some part of the spring forces perpendicular to the path to achieve stability during optimization.[61] In this way, the band resolves the high energy transition state regions first, as it is less elastic. Convergence of the path to the MEP solely using DNEB does not occur; rather, transition state candidates are further refined or a NEB description of the gradient is required. Coupling DNEB with discrete path sampling connects distant minima and transition states, and establishing a pathway for complex rearrangements. This method has been applied to examine folding in small peptides.[62]

Procedures for optimization need to be taken into account for the various chain-of-states methods presented here. First principles calculations and empirical force fields have both been used to optimize paths. In the NEB methods, force projections on each image must occur to optimize the interpolating path. This makes it difficult to define an objective function to minimize. Optimization procedures are used to minimize the NEB along these forces to the MEP. String methods use the same optimizers. Steepest descent, conjugate gradient, [63] and limited memory Broyden-Fletcher-Goldfarb-Shanno[64] optimizers have all been used.[57]

Path optimization can be performed using molecular dynamics based simulated annealing protocol, where final energy minimization is completed using a velocity-Verlet algorithm. [65] Recently, a super-linear minimization scheme based on the adopted basis Newton-Raphson method has been introduced, and has been shown to increase convergence to the MEP.[58] Combinations of NEB and second order parallel path optimizer have been used to refine QM/MM reaction paths. [66]

1.3.2 Umbrella Sampling

Umbrella sampling and the weighted histogram analysis method are used to derive the potential of mean force, or free energy, of a transition.[67; 68; 69; 70] Umbrella sampling is most useful when a transition is hindered by high energy barriers on the free energy surface. Typically, a simulation can spend its time sampling an energy minimum region of the surface, and transitions over a high energy barrier infrequently, if ever, occur. Thus, sampling the entire ensemble of possible structures is hindered. Umbrella sampling introduces a harmonic biasing potential along a reaction coordinate which confines the system to a small region which it can

efficiently sample. Many windows are run along a reaction coordinate with varying biasing potentials, from which the free energy difference is calculated. The potential of mean force (*A(x)*) for a specific state of a molecule is derived from the following equation:

$$A(x) = -k_B T \ln P'(x) - U'(x) \qquad (1\text{-}10)$$

k_B is the Boltzmann constant, and T is system temperature. *P'(x)* is the population sampled with a biased potential, and *U'(x)* is the biasing potential. Populations of neighboring windows must overlap so calculated free energies are continuous along the reaction coordinate. The Weighted Histogram Analysis Method is used to remove the harmonic biasing potential of each window to yield an unbiased probability distribution and self-consistently find the free energy.[70]

There are a few important caveats to successfully running an umbrella sampling simulation. The first is the choice of reaction coordinate. For known or simple transitions choice of reaction coordinate can be straightforward, for example the distance between two atoms forming a salt bridge.[71] Complicated, large scale conformational transitions pose a more complex problem for choosing reaction coordinates on which to project the free energy of the entire system. In some cases, it may be necessary to use multiple dimensions for complex transitions.

A second issue in umbrella sampling is the strength of the restraint force used in the biasing potential. A small force will result in structures migrating away from their expected values and lead to instability in the PMF calculation. Forces which are too high may cause system strain in parts which are not represented in the reaction coordinate, resulting in unrealistic energies. The strength of the restraint force used in the biasing potential is related to the spacing of windows along the reaction coordinate. Windows must overlap to provide continuous structures and energies along the reaction coordinate. Windows which do not overlap will cause discontinuities in the free energy surface, interpreted as high free energy barrier regions, thus

skewing the relative free energies along the entire reaction coordinate. In many cases, these parameters must be decided by trial and error.

Umbrella sampling has been used to study transitions in many systems, including examining base flipping in DNA.[22; 72; 73]

1.4 Overview of specific research interests

1.4.1 Duplex and Complex bound 8-oxoguanine structure

The conversion of guanine to 8-oxoguanine is the most common form of oxidative damage to DNA caused by reactive oxygen species. Formamidopyrimidine glycosylase is the bacterial enzyme that recognizes and initiates repair of 8OG:C pairs in DNA by removing the 8OG base. However, the recognition mechanisms of this enzyme remain elusive. Molecular dynamics simulations of damaged and undamaged duplexes in the presence of explicit solvent reveal fluctuations in backbone dihedral angle values between G and 8OG residues. These changes correlate with increased flexibility of distances between the O8 and deoxyribose oxygen of the damaged base. The simulations indicate that the oxidized base prefers a set of phosphate backbone angles in the BII substate range, while its undamaged counterpart adopts canonical B-DNA values for its backbone dihedral angles. In the Fpg-DNA complex simulations, both the lesion and undamaged G adopt similar conformations, with backbone angles comparable to those sampled by 8OG in the free duplex. Overall, this suggests that the repair enzyme imposes conformational constraints on the interrogated DNA that permit damaged DNA to remain in its preferred state while forcing the undamaged DNA into a less favorable state. The lower energetic

barrier for Fpg to bind 8OG damaged DNA is evidence that the enzyme may promote selective processing.

1.4.2 Development of the Partial NEB Method

Although the nudged elastic band method can be used to find a minimum energy path between two given starting structures, in its original implementation it calculated NEB forces between all atoms of a system, limiting its application to small systems with implicit solvent. In order to study the base eversion pathway of Fpg, a novel implementation of this method will be discussed, in which the nudged elastic band force decoupling is applied to only a specific, user-defined subset of atoms in a particular system. This is shown to produce results and minimum energy pathways comparable to the standard implementation for an alanine dipeptide test system. The new implementation, called partial NEB (PNEB), allows incorporation of explicit solvent with simulated systems, which may be preferred in many cases to an implicit solvent model. From a computational standpoint, PNEB also reduces the communication overhead inside the code, resulting in better performance for the larger Fpg system studied here, and large systems in general.

1.4.3 Energetics of base eversion for 8-oxoguanine via major and minor groove pathways

Base eversion is a fundamental process in the biochemistry of nucleic acids, allowing proteins engaged in DNA repair and epigenetic modifications to access target bases in DNA.

Crystal structures reveal endpoints of these processes, but not the pathways involved in the dynamic process of base recognition. To elucidate the pathway of 8OG during base excision repair by Fpg, we calculated free energy surfaces during eversion of the damaged base through the major and minor grooves. The minor groove pathway and free energy barrier (6-7 kcal/mol) are consistent with previously reported results.[22] However, eversion of 8OG through the major groove encounters a significantly lower barrier (3-4 kcal/mol) more consistent with experimentally determined rates of enzymatic sliding during lesion search.[74] Major groove eversion has been suggested for other glycosylases, suggesting that in addition to function, dynamics of base eversion may also be conserved.

1.4.4 Mutant studies to examine substrate recognition

One advantage to performing simulations is the ease with which mutations can be modeled to test the effect of specific residues on the base eversion pathway. Specific mutations have been made to address the role of protein side chains which intercalate into the DNA helix prior to eversion and during eversion. Residues F113 and R111 have been well characterized by crystallography, but the dynamic role they play in the base eversion pathway has not. The F113 wedge intercalates between the damaged base pair and its neighbor, buckling the 8OG:C base pair and facilitating eversion in the bent DNA. In the F113A mutant, the free energy of the initial eversion stages is about 5 kcal/mol higher than the WT, indicating this wedge plays an active role in destabilizing the base pair. To examine the role R111 plays in recognizing the orphaned C, two mutants were created: R111A and WT-8OG:A. R111 not only recognizes the orphaned base, but plays a role in stabilizing a critical intermediate in the base eversion pathway. The

R111A mutant shows a complete loss of stabilizing interactions along the transition state of the base eversion path, correlating well with experimentally derived kinetics data.

2 DUPLEX AND COMPLEX BOUND 8-OXOGUANINE STRUCTURE

2.1 Introduction

DNA continuously encounters reactive oxygen species through cellular respiration and inflammatory processes, causing deoxynucleotides to undergo chemical and structural modifications. One of the most common and widely studied types of oxidative damage to G is 8-oxoguanine (8OG) (Figure 2-1), which can pair with adenine during DNA replication by forming an alternate Hoogsteen base pair in the active site of DNA polymerases. Thus, 8OG promotes guanine to thymine mutations with loss of the original DNA sequence. Base excision repair (BER), an enzymatic pathway present in both eukaryote and prokaryote cells, prevents mutations caused by 8OG and other oxidative lesions. In prokaryotes, formamidopyrimidine glycosylase (Fpg), also known as MutM, initiates BER of 8OG by cleaving the lesion when paired to cytosine. MutY complements Fpg action by removing dA from mutagenic 8OG:dA pairs while MutT clears 8OG from the deoxynucleotide pool. These three enzymes are part of the 'GO cycle', which protects DNA against mutations induced by 8OG.[15]

Location of a modified base occurs in a vast excess of undamaged DNA and it is presently unclear how and when repair enzymes initially differentiate damaged from normal residues. There is ongoing debate as to whether the recognition mechanism can involve DNA glycosylases moving along the DNA helix via fast sliding or a hopping method.[75] In the first mechanism the enzyme slides along DNA, probing the integrity of each base pair until it finds a

lesion, implying that the damaged residue induces local structural or energetic features that differentiate it from normal bases.[74] A variation of this mechanism has the enzyme partially extruding into and probing all bases in its preincision site, while sliding along the helix.[74; 76] In contrast, the second mechanism implies that the enzyme continuously dissociates from and reattaches to the DNA, perhaps causing an increase of the helix breathing rate and probing the partially extruded bases in its preincision site where lesions are detected.[76] Stopped-flow kinetics studies with 2-aminopurine-labeled DNA have identified the rate constants of conformational transitions of DNA along the pathway of 8OG:C recognition and excision by Fpg.[48] These rate constants indicate that DNA destabilization and complex formation occur concurrently, suggesting that the enzyme aids in the destabilization of an already altered duplex structure. As important, perhaps, is the finding that the enzyme rejects poor substrates before everting them into its active site, suggesting an indirect readout by Fpg that uses a combination of structural and chemical features of the lesion.[48]

Structural features of the 8OG lesion have been studied via NMR and crystallography. An early NMR study of 8OG-damaged DNA in the context of the Dickerson dodecamer concluded that the overall structure is almost identical to that of the unmodified duplex, with small changes localized at the lesion site.[77] One of the few published X-ray crystal structures of 8OG-containing DNA has two 8OG:C pairs in the same molecule and shows close similarity to the structure of the unmodified cognate duplex.[78] The authors report only a slight shift in the glycosidic torsion of one of the 8OG bases and a shortening of the Watson-Crick hydrogen bonds between the 8OG and its cytosine partner.

An important recent study has shown that the conformational variability of damaged nucleic acid structure in protein-DNA complexes generally match the conformations seen via

NMR spectroscopy of damaged duplexes in solution.[79] These similarities suggest that molecular dynamics simulations of damaged duplexes could provide relevant insight into the effect of damaged residues on duplex structure and identify protein-binding preferences. In addition, molecular dynamics could identify fast conformational exchange in the duplex that occur on a time scale generally inaccessible to experiments.

Computational studies which capture the flexibility of the DNA backbone suggest dynamics could facilitate molecular recognition processes. Previous molecular dynamics simulations have shown conformational backbone variation and general structural transitions of nucleic acids occur at computationally accessible timescales.[80] Transitions between BI and BII conformational substates of B-DNA, specifically defined by local variations of the epsilon and zeta dihedral angle values, have been studied via molecular dynamics simulations. [81] These simulations examined local deformations during BI to BII transitions, noting that specific structural features of BII DNA are different from canonical B-DNA (Figure 2-1). It has more recently been proposed that the same conformational changes in the epsilon/zeta backbone dihedral angles which cause shifts in BI/BII substates may play a main role in base flipping[82] and in perturbation of base stacking interactions.[83] Characteristics such as these, which alter the canonical DNA structure, have been implicated in damage recognition.[84] Previous studies of the 8OG residue of an 8OG:A mismatch show that the damage is very flexible, undergoing spontaneous *anti* to *syn* transitions during simulations.[85] Steric hindrance and unfavorable electrostatic interactions of 8OG were associated with changes in the epsilon and zeta dihedral angles, causing destabilization of the *anti* orientation and inducing glycosidic angle transitions.

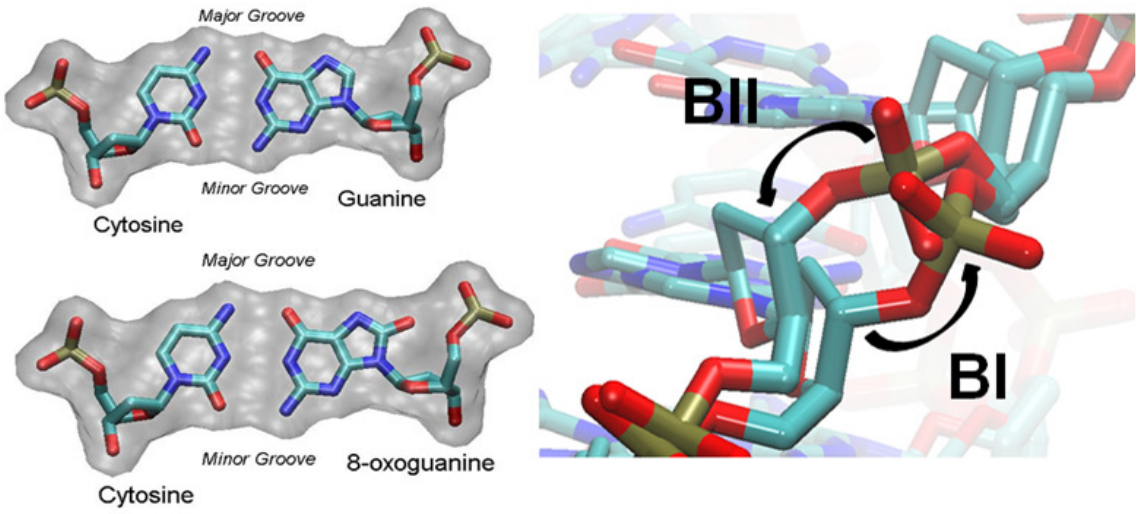

Figure 2-1 Left: Structure of guanine : cytosine (G:C), and 8OG : cytosine (8OG:C) in canonical Watson-Crick base pairing. Right: BI and BII conformational substates identified by diverging phosphate backbone angles.

Taken together, these studies suggest that backbone dynamics may also play a major role in damage recognition by Fpg. To study this motion with atomic level detail, molecular dynamics simulations were performed in explicit solvent of DNA containing an 8OG lesion or a guanine base, in an otherwise identical sequence context, using the recent improvements to the Amber nucleic acid force field.[42] The results suggest that the 8OG residue samples a flexible range of epsilon and zeta, rendering it significantly more flexible inside a B-form DNA duplex than its undamaged guanine counterpart. Destabilization of a BI backbone structure via population shift to a more favorable BII structure upon protein binding is seen in simulations of the Fpg protein complexed to damaged DNA, and while G is also restricted to BII in the complex, this involves significantly more perturbation to conformational preferences of the free duplex. This conformational capture may help the enzyme to recognize undamaged bases while maintaining the ability to extrude and cleave damaged bases.[86]

2.2 Methods

2.2.1 Systems and setup, minimization

The DNA sequences were taken from the 2F5O crystal structure[19; 86] of Fpg cross linked to DNA. 15-mer duplexes with sequence context 5'-(GGTAGATCCGGACGC)-3' and 5'-(GCGTCCXGATCTACC)-3' where X = G or 8OG were built in canonical B-form using the NUCGEN module of the Amber (version 9) suite of programs.[87] The duplex with an undamaged guanine will be referred to as G duplex and the duplex with an 8OG lesion will be referred to as 8OG duplex. The same 2F5O crystal structure of Fpg crosslinked to DNA interrogating an undamaged intrahelical guanine base was used for the starting structures of the protein complexed to DNA containing an intrahelical guanine or 8OG base. Although the bulk of the crystal structure was not changed, this structure was modified by adding coordinates for the catalytic loop (residues 217 to 237 inclusive, missing in the electron density map) based on the 1R2Y crystal structure[19]. The disulfide crosslink tethering the protein to the DNA in the experiment was not included in the 2F5O structure or in these simulations. For the simulations of protein complexed to 8OG damaged DNA, the protein and DNA positions were taken from the same 2F5O crystal structure. The 8OG residue at the position of the interrogation site was modeled based on the position of the guanine. Parameters for the 8OG residue were obtained from Miller et al.[88] The Fpg-DNA system is referred to as G complex, with a guanine base, or 8OG complex when including an 8OG base.

All molecular dynamics simulations were carried out with the SANDER module of the Amber (version 9) suite of programs.[87] For all simulations the ff99SB force field [89] with

parmbsc0 corrections for alpha and gamma nucleic acid torsion angles was used.[42] Structures were solvated in TIP3P water[90] in a truncated octahedron periodic box, with a minimum initial distance of 8Å between the solute and the boundary box. A total of ~6100 water molecules were used for DNA duplexes and ~10000 water molecules were used for the protein-DNA complexes.

The duplex system was minimized by 500 steps of steepest descent minimization followed by 500 steps of conjugate gradient minimization, both with 500 kcal mol^{-1} $Å^{-2}$ positional restraints on the DNA. A second round of 1000 steps of steepest descent minimization followed by 1500 steps conjugate gradient minimization without any restraints on the system were performed. Molecular dynamics was run for 100ps with 10 kcal mol^{-1} $Å^{-2}$ positional restraints on the DNA to heat the system from 0K to the target temperature of 330K, which reflects the thermophilic nature of the *Bacillus stearothermophilus* sequence that was used for crystallization.[21] This was followed by 2ns of molecular dynamics at 330K with no restraints which were discarded before the final production run.

The protein-DNA complex was minimized by two rounds of 200 steps steepest descent followed by 200 steps conjugate gradient minimization, with 5 kcal mol^{-1} $Å^{-2}$ positional restraints on the heavy atoms. The complex was then solvated and heated to 330K with positional restraints on the protein and DNA atoms, allowing only the water to move. This was followed by five 1000-step cycles of minimization in which the side chain restraints were gradually decreased, and four 5000 step cycles of MD simulations with decreasing restraints on the protein and DNA backbone. A subsequent 5000 step MD simulation was carried out without restraints.

Production dynamics for each of the four systems (G duplex. 8OG duplex, G complex, 8OG complex) were performed with the PMEMD module of Amber 9. Berendsen thermostat and barostat with 1ps relaxation constants were used to perform simulations in the NPT ensemble.[91]

SHAKE was used to constrain bonds to hydrogen, and a timestep of 2 femtoseconds was employed.[92] The production runs for each system were given different initial velocities in order to randomize trajectories. 10 simulations for each system were carried out, with each simulation totaling 15ns. This provided 150ns total of structure information for each system studied.

2.2.2 Analysis

Analysis was carried out with the PTRAJ module of Amber (version 9). MM-GBSA energy decomposition was used to calculate the interaction energies between two groups of atoms, designated "receptor" and "ligand." Free energy surfaces were calculated using a 2-dimensional histogram program which uses populations to calculate Gibbs free energy as follows:

$$\Delta G_i = -RT \ln\left(\frac{N_i}{N_0}\right) \qquad (2\text{-}1)$$

where N_i is the population of a particular histogram bin, and N_0 is the population of the most populated histogram bin (which results in a value of 0 for the lowest free energy state).

Kinetics analysis was performed to assess the rates of transition between the BI and BII DNA substates. A first order reaction was assumed with the form

$$\frac{A_n}{A_0} = e^{-k\Delta t} \qquad (2\text{-}2)$$

where A_n/A_0 is the fraction of population remaining with respect to the amount at the start, k is the rate constant, and Δt is the time the base remains in a single substate. Substates were defined using a cutoff of zeta > 210° for BI, and zeta < 210° for BII. A plot of fraction population versus

time spent in either the BI or BII substate was fit with a single exponential from which the rates were estimated.

Cluster analysis was performed on the DNA in both the duplex and complex simulations.[93] Distance similarity clustering using the average linkage method was used.[93] Clusters were formed with a bottom-up approach using a similarity cutoff of 1.25Å. Clustering was done on the heavy atoms of the central three base pairs.

2.3 Results and Discussion

2.3.1 Free energy profiles of backbone dihedral angle transitions

The relative populations of BI and BII conformations were compared between systems by calculating 2D free energy landscapes using epsilon and zeta angles (Figure 2-2). As expected, these dihedrals show significant correlation, and there is a distinct shift in the free energy minimum when the G duplex is compared to the 8OG duplex, favoring a BI conformation for the G duplex and a BII conformation for the 8OG duplex. The relatively low free energy barrier (0.2 kcal/mol) between the BI and BII populations for the guanine base indicates that it is relatively easy to accomplish the backbone rearrangement. The 8OG duplex simulation shows a free energy change of -0.6 kcal/mol to achieve the structural transition from BI to BII, consistent with the more favorable BII minimum for 8OG discussed above.

Transitions between the BI and BII conformations also occur in the protein-DNA complex simulations, but for both systems, the BII conformation becomes the global free energy minimum, consistent with G complex crystal structure epsilon and zeta angles (corresponding to

the star in Figure 2-2). For the G complex, this represents a modest shift in equilibrium away from BI structures, which are no longer energetically favorable. This implies that Fpg binding has some impact on the local behavior of the G duplex, and insertion of the Fpg wedge and significant overall kinking of the DNA does not permit G to remain in its solution state free energy minimum. The 8OG in the complex simulation is shifted to the BII conformation as well, a result of the imposition of a BII preference by Fpg. However, 8OG maintains the flexibility seen in the duplex simulations, and samples a BI conformation with a greater preference than G. As a result, G undergoes a much greater shift in equilibrium upon binding than does 8OG, with a correspondingly greater free energy penalty. Localization of the interrogated base in a BII state for the complex relative to the BI preferred in the duplex ($\Delta\Delta G_{Complex-Duplex}$) generates 2.9 kcal/mol of free energy strain in the case of the undamaged guanine base, but only 1.2 kcal/mol in the case of the lesion.

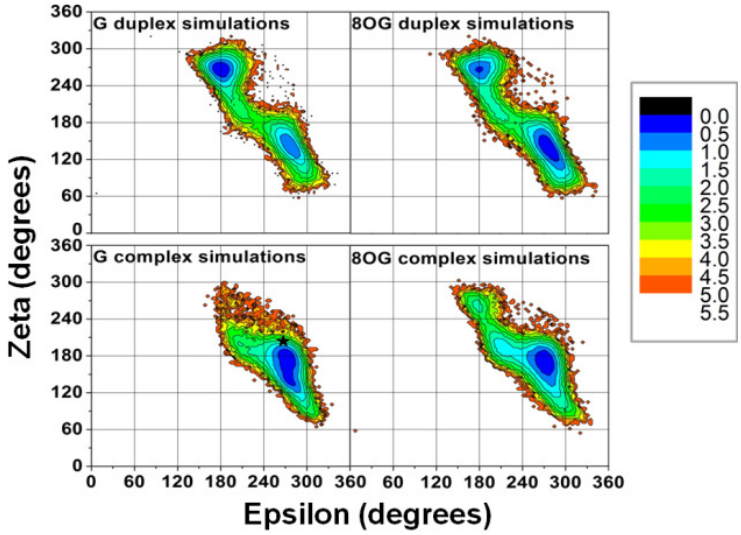

Figure 2-2 Free energy (in kcal/mol) surfaces of target base epsilon-zeta correlations in duplex (top) and complex (bottom) simulations, where the target base is guanine (left) or 8OG (right). The star "★" indicates the value in crystal structure 2F5O, used for initial coordinates in this simulation.

2.3.2 Kinetics analysis of duplex backbone transitions

The relative rates of transitioning between BI and BII substates were calculated, where the transition from BI to BII was considered k^1 and the transition from BII to BI was considered k^{-1}(see Figure 2-3 and Table 2-1). Initially, all data was fit using a single exponential, indicating a two- state model. The G base forward rate is slower than its reverse rate, indicating that the equilibrium is shifted towards sampling the BI structure. This is consistent with the thermodynamic preference of the G for the BI substate. The 8OG k^1 can be fit using a single exponential; indicating this forward transition from BI to BII is a two-state process. The 8OG k^1 is also the fastest rate calculated in all of the systems, indicating the preference of the 8OG base to move away from BI and to BII quickly. The 8OG k^{-1} was the slowest of all calculated rates, indicating the reverse process of BII to BI transition is slow, corresponding to the shift in equilibrium favoring the BII structure.

The 8OG k^{-1}, however, had a poor correlation when fit with a single exponential, and was refit using a double exponential. This implies there is an additional state other than the BI substate to which the 8OG, when in BII conformation, can transition. The double exponential was fit using a combination of fast and slow rates, with the ratio being 60:40 (slow:fast). This indicates there is a slower transition occurring, which is assumed to be BII to BI since it is comparable to the G BII to BI transition rate. There is an additional fast transition which is sampled from the BII substate. This transition to an alternate structure has been observed in some simulations and will be termed the BIII substate. Since the single exponential fit in the forward direction is in excellent agreement with the data, it seems as if this transition to a BIII state is only accessible from the BII state.

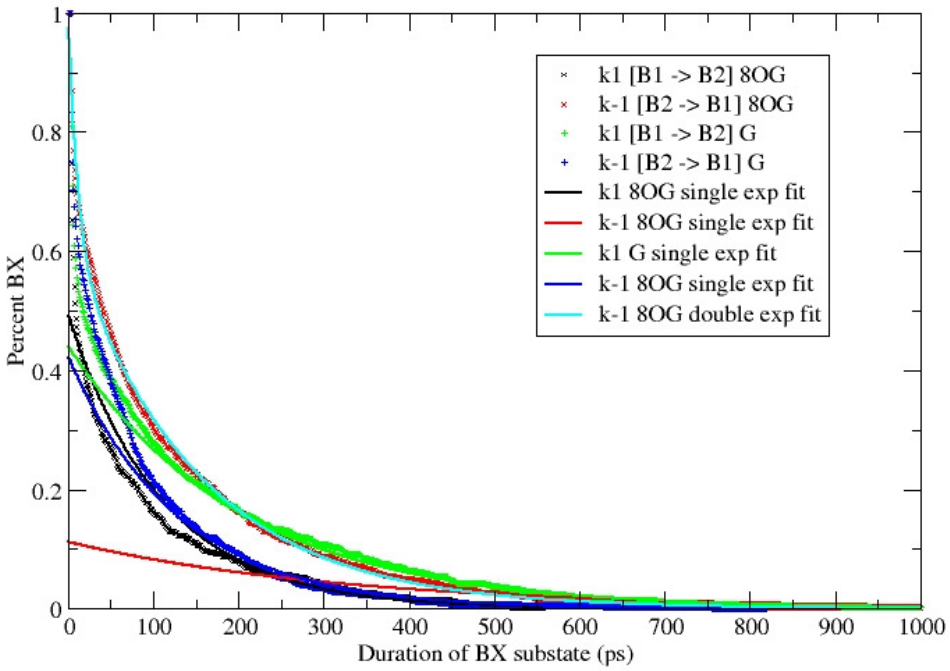

Figure 2-3 Calculated percent of BI or BII substate vs. duration of substate. Raw data (symbols) was fit using single or double exponentials (solid lines).

Table 2-1 Calculated rates of transition between BI and BII substates for G duplex and 8OG duplex simulations.

System	**Rate (ps^{-1})**	**Rate (ps)**	**Correlation**
G BI to BII, k^{1}	$5.00x10^{-3}$	199.9	0.99
G BII to BI, k^{-1}	$7.77x10^{-3}$	130.1	0.99
8OG BI to BII, k^{1}	$9.03x10^{-3}$	110.76	0.97
8OG BII to BI, k^{-1}	$3.07x10^{-3}$	326.13	0.88
Double exp fit: slow k^{-1}	$6.57x10^{-3}$	151.60	0.99
Double exp fit: fast k^{-1}	$9.11x10^{-2}$	10.98	

2.3.3 Structural significance and energy decomposition shows repulsive interactions with O8 of 8OG shift equilibrium to the BII structure

MMGBSA energy decomposition analysis was used to characterize the nonbonded interactions responsible for the shift in equilibrium between the BI and BII substates of the G and 8OG bases, as well as the shift between the bound and unbound forms. The stacking interactions between the target base and its 3'G was measured in the duplex and complex systems (Figure 2-4). In duplex form, 8OG has a more stable stacking interaction; this energy is much more similar to the values sampled by the complex. The G duplex values are stabilized upon binding in the complex, but must be perturbed from its solution state minimum to the values seen in the complex. Structural comparison of the preferred conformations of each duplex target base shows the 8OG (in a BII conformation) is more shifted towards the major groove than the G (in a BI conformation). The data shown here (as well as in Figure 2-2) indicates that the BII conformation is stabilized in the complex, indicating the complex prefers an increase in base pair shift towards the major groove. Interestingly, this shift towards the major groove correlates with previous work indicating a major groove base eversion pathway is used by Fpg to flip damage into the active site of the enzyme.[94]

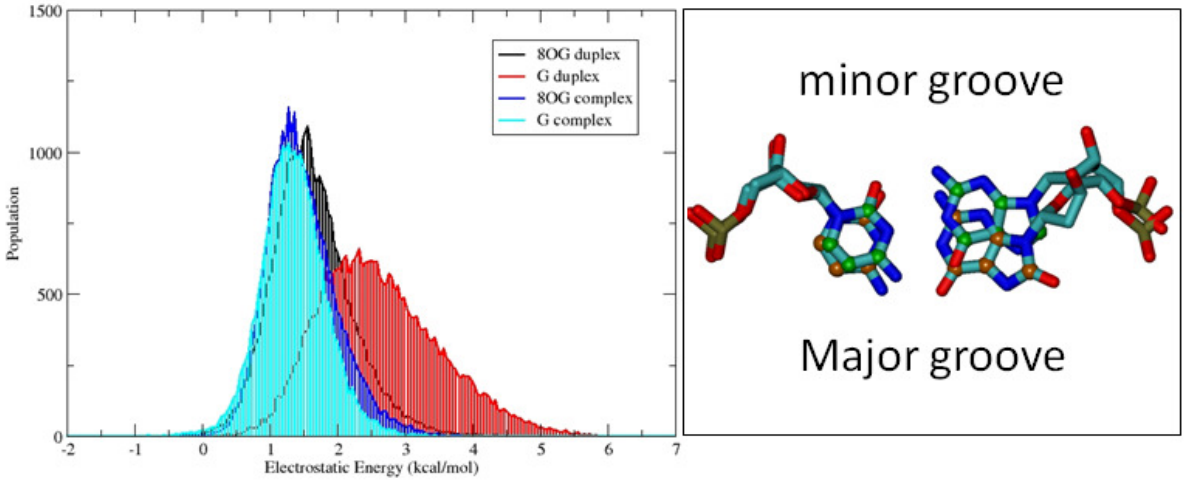

Figure 2-4 Electrostatic interaction between target base and 3'G. 8OG in duplex has a more stable stacking interaction than G, most likely due to extra partial charge. This energy is comparable to the vales in the complex. The structures show that the 8OG base (orange spheres) is positioned towards the major groove compared to the G (green spheres). Structures are overlapped on all heavy atoms of the central three base pairs.

Displacement towards the major groove appears to optimize the distance between the O8 atom and the DNA backbone. Two interactions were measured to analyze whether this affects the preference of 8OG in adopting a BII substate. Figure 2-5 shows the electrostatic interaction energy of a two atom decomposition. The two atoms are the O4' of the deoxyribose ring and atom 8 of the target base, corresponding to O in 8OG and H in G. The electrostatic energy is much less favorable for 8OG than for the undamaged G, which is expected since a repulsive interaction occurs between the O8 and O4' in the 8OG systems. Interestingly, there is a shift upon binding. For the undamaged G, when the protein binds the population shifts to a less favorable energy. For the 8OG damage, the reverse is true – the population shifts to a more favorable energy upon protein binding. Analysis shown in Figure 2-5 b and c indicates the

population shift correlates with the shift in the B-DNA substate, with the BII conformation being more favorable for 8OG than it is for G.

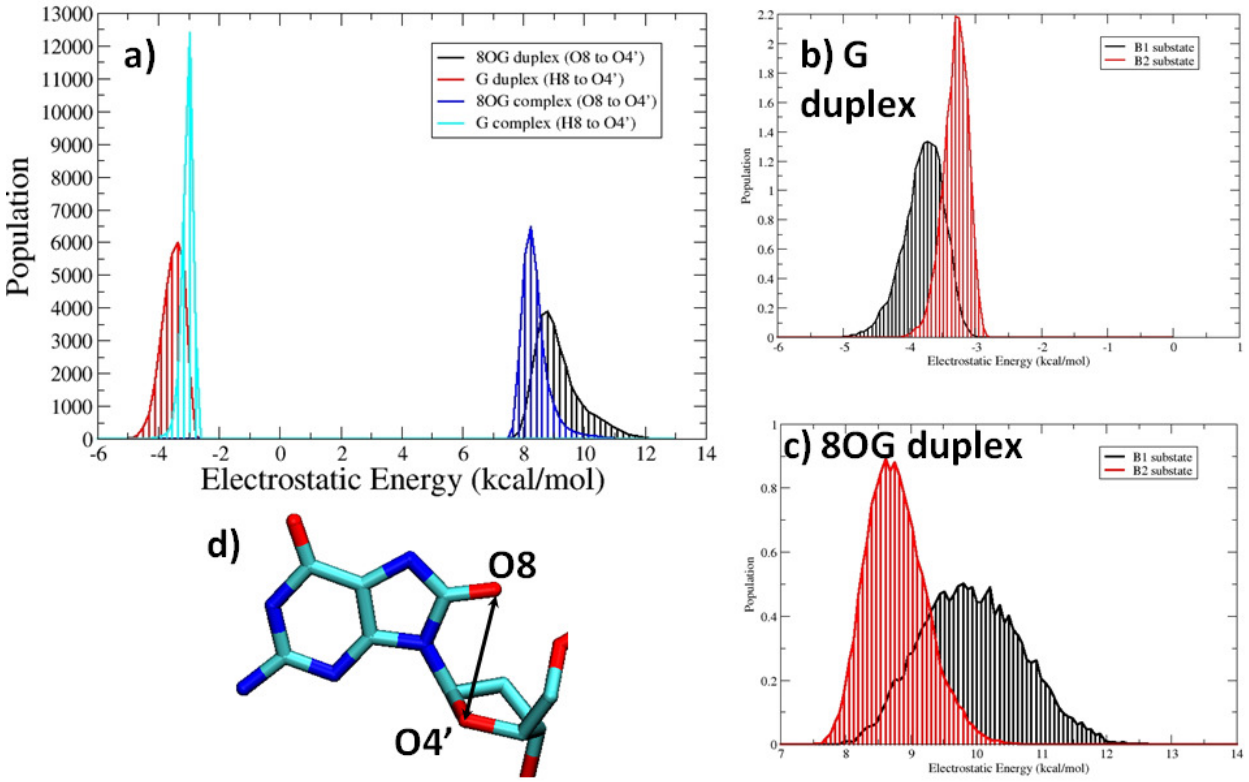

Figure 2-5 a) Electrostatic interaction energy of O8 to O4' in 8OG systems (duplex in black and complex in dark blue) and of H8 to O4' in G systems (duplex in red and complex in cyan). b) Electrostatic interaction energy of H8 to O4' in G duplex in B1 conformation (black) and B2 conformation (red). c) Electrostatic interaction energy of O8 to O4' in B1 conformation (black) and B2 conformation (red). d) Structural representation of O8 to O4' distance.

Additionally, the electrostatic interaction between the O8/H8 and O2P of the target base phosphate group was measured. Figure 2-6 shows this energy is more stable for G than for 8OG, where we expect a repulsive interaction from the two oxygen atoms. As in the interaction with

O4', the 8OG shifts to more favorable structures when bound in the complex (Figure 2-6a, black and blue). The G is shifted to less energetically favorable structures when bound in the complex (Figure 2-6a, red and cyan). This can be explained by the energy decomposition performed on each substate and seen in Figure 2-6b and c, where the preferred BI structure of the G duplex is more favorable than the BII conformation. For the 8OG base the reverse is true, and the BII substate has a more stable electrostatic energy than the BI substate.

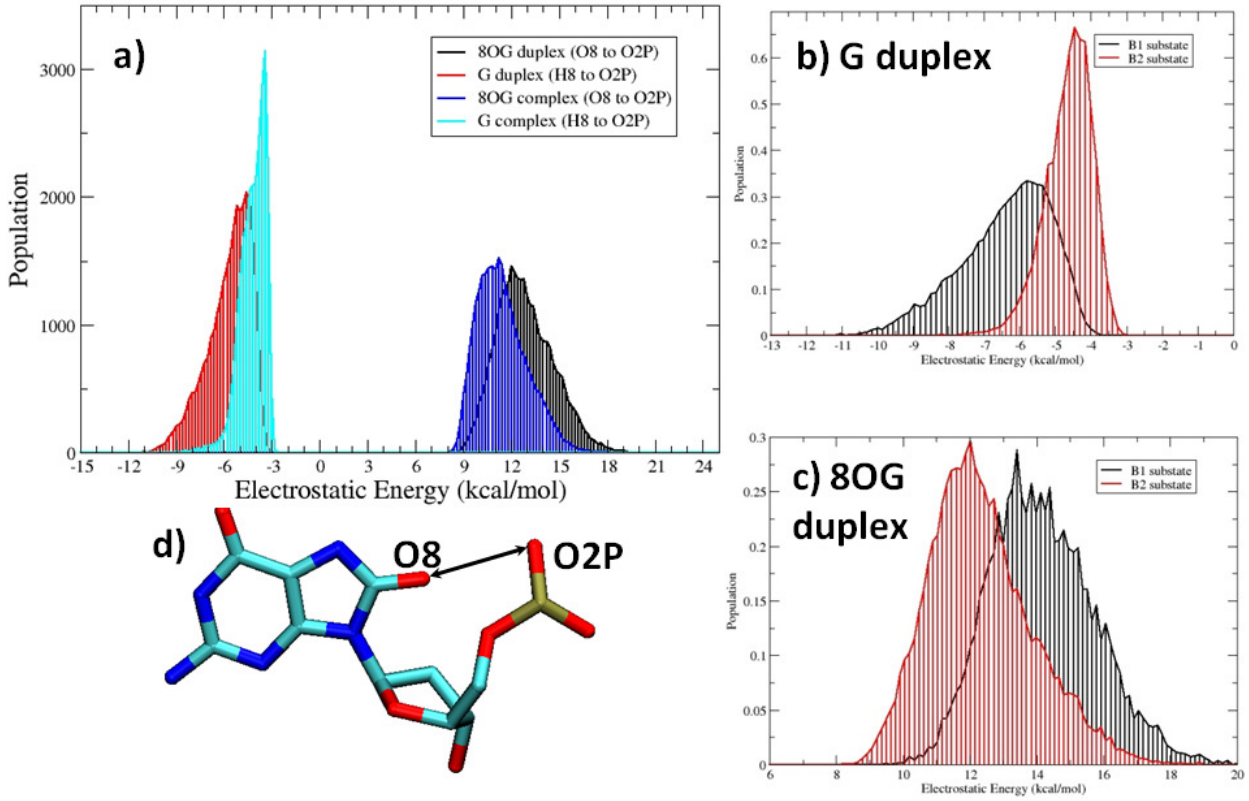

Figure 2-6 a) Electrostatic interaction energy of O8 to O2P in 8OG systems (duplex in black and complex in dk. blue) and of H8 to O2P in G systems (duplex in red and complex in cyan). b) Electrostatic interaction energy of H8 to O2P in G duplex in B1 conformation (black) and B2 conformation (red). c) Electrostatic interaction energy of O8 to O2P in B1 conformation (black) and B2 conformation (red). d) Structural representation of O8 to O2P distance.

A summary of these interactions is shown in Figure 2-7, where a structure comparison has been done based on the central three base pairs in the duplex. The representative structures from the most populated clusters from cluster analysis are shown overlapped in Figure 2-7a, with the 8OG structure shown in orange and the G structure shown in green. Figure 2-7b shows the distances between the G base's H8 and O4' atoms (3.35Å) and H8 and O2P atoms (4.5Å), which are representative values for the BI DNA substate. The 8OG duplex in BII form in Figure 2-7c shows the distances between the 8OG base's O8 and O4' atoms (3.87Å) and O8 and O2P atoms (6.42Å) which are longer than the G duplex values and due to the repulsive interaction of the two oxygen atoms. These longer values are preferred for the 8OG base, since during the transition to a BI structure these distances are shorter and less energetically favorable. This observation agrees with the calculated rate constants (Table 2-1), which indicate the BII to BI transition is at least 50% slower than the BI to BII transition for 8OG.

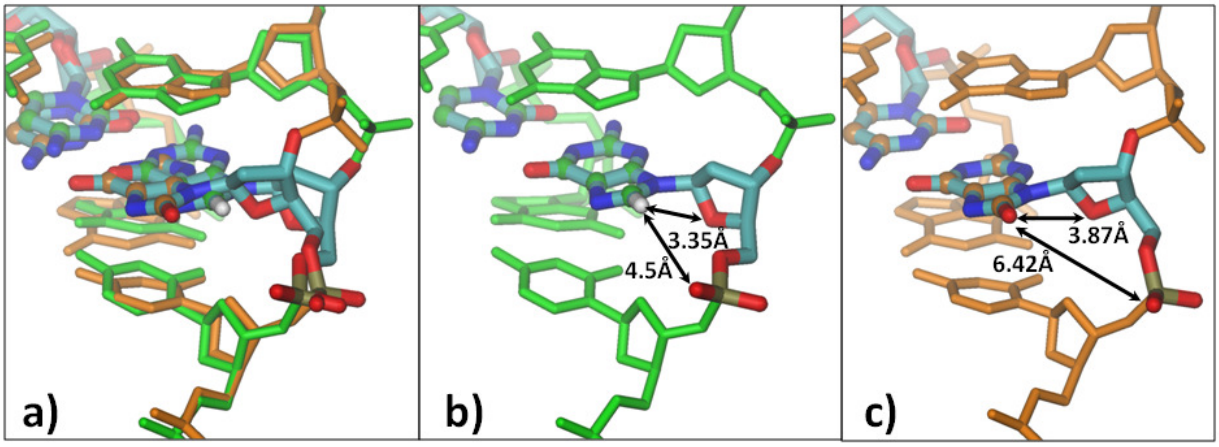

Figure 2-7 Structure comparison of representative conformations of 8OG (orange) and G (green). a) Overlap of structures based on heavy atoms of central three base pairs. b) G in majority conformation (BI). Distance between H8 and O4' (3.35 Å) and H8 and O1P (4.5 Å) shown. c) 8OG in majority conformation (BII). Distance between O8 and O4' (3.87 Å) and O8 and O1P (6.42Å) shown.

The BII substate is preferred in both sets of complex simulations. The representative structures from cluster analysis show many similarities in protein-DNA interactions, which is expected since the majority conformations of the 8OG bound complex and the G bound complex are the same. An interesting difference can be seen in the electrostatic interaction of Arg263 to the target base. This is illustrated in Figure 2-8, below. In the intrahelical BII conformation the interaction energy of the target nucleobase and the Arg263 side chain is more favorable for the 8OG complex than it is for the G complex. The negative partial charge of the O8 has a more favorable interaction with the Arg263 guanidinium group. This influences the orientation of the Arg263 side chain, as seen in Figure 2-9. The hydrogen bonding pattern differs between the two systems: in the 8OG complex, the hydrogen from the terminal amine hydrogen bonds to the 3' phosphate group, while in the G complex the HE hydrogen bonds to the 3' phosphate group.

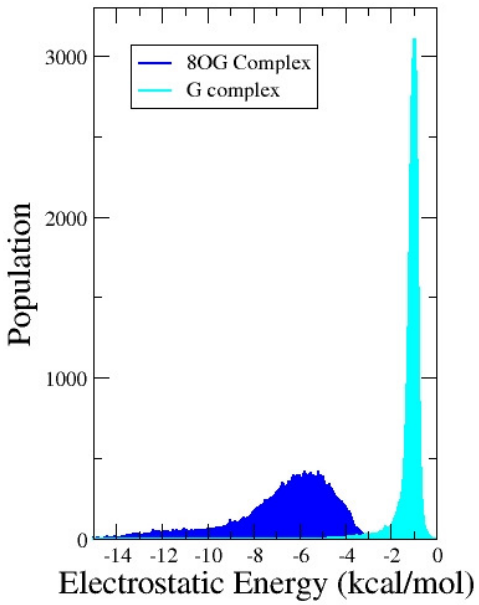

Figure 2-8 Electrostatic interaction energy between target base (8OG in blue, G in cyan) and Arg263 side chain.

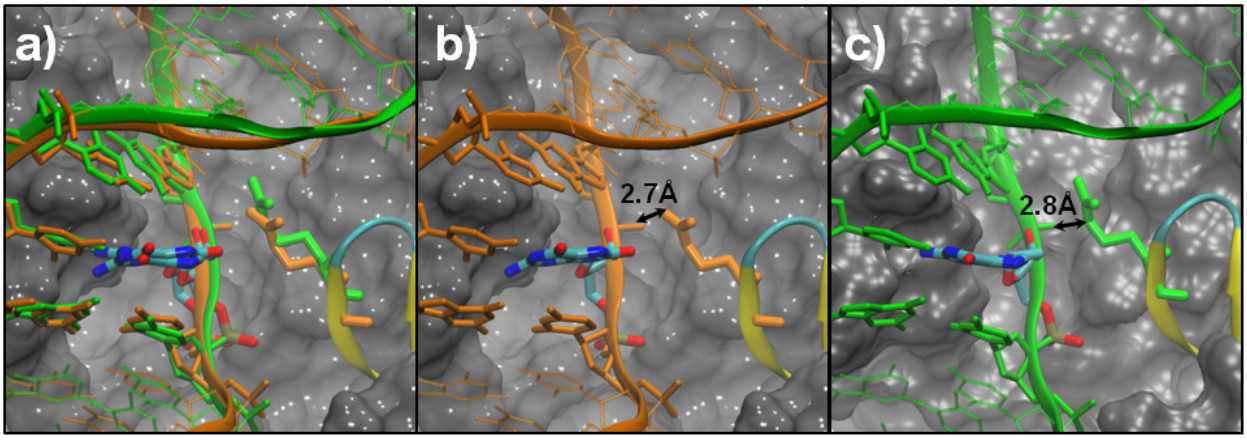

Figure 2-9 a) Overlap of best representative structures from the most populated cluser for 8OG complex (orange) and G complex (green) simulations. b) 8OG complex shows interaction of Arg263 NH1 with 8OG base 3'O2P. c) G complex shows interaction of Arg263 NE with 8OG base 3'O2P.

2.4 Conclusion

Damage recognition is a critical step in the repair pathway of 8OG. Despite the structural similarity between the 8OG lesion and the undamaged G, our simulations show a distinct difference in the dynamics of the bases is the context of their solution state duplexes. G prefers a canonical distribution of structures in BI and BII DNA substates, with the overall energy minimum populating a BI state. For 8OG this equilibrium is shifted to the BII state. The free energy cost for Fpg binding a G is twice as high as it is for the 8OG lesion, indicating the protein is able to discriminate between the 8OG and G pre-eversion. Additionally, energy decomposition and structure analysis show that repulsive interactions specific to the 8OG damage cause the shift in population to the BII structure, as this allows for optimal distance between the O8 atom and

the oxygen of the sugar ring and phosphate backbone. 8OG is shifted towards the major groove, which is the proposed direction of base eversion into the active site where catalysis occurs.

3 DEVELOPMENT OF THE PARTIAL NEB METHOD

Acknowledgements

The material presented in this chapter contains direct excerpts from the manuscript *A Partial Nudged Elastic Band Implementation for Use With Large or Explicitly Solvated Systems* by Christina Bergonzo, Arthur J. Campbell, Ross C. Walker and Carlos Simmerling, published in the International Journal of Quantum Chemistry in 2009, volume 109, pages 3781-3790. The manuscript was written by Christina Bergonzo with suggestions and revisions by Arthur J. Campbell, Ross C. Walker, and Carlos Simmerling. NEB simulations and analysis were performed by Christina Bergonzo and umbrella sampling simulations were performed by Arthur J. Campbell.

3.1 Introduction and Theoretical Background

One of the most relevant problems currently in molecular dynamics simulations is how to study conformational changes which occur along a defined path. Many approaches to finding minimum energy transitions involve beginning in one energy minimum and proceeding stepwise along the slowest ascent path of the surrounding energy landscape.[49] These methods are inherently serial which limits their performance and thus applicability to complex systems. From a computational perspective arguably the best approach to solving this problem is to use a method which can take advantage of current massively parallel computing resources by distributing multiple simulations among many processors. An example of such an approach is umbrella sampling.[67] In such methods a reaction coordinate is typically chosen along which the

system is forced to move. In practice, however, these methods can lead to discontinuous paths and require a prior definition of a reaction coordinate along which to bias the simulation. Such prior definition of a reaction coordinate is often impractical in systems of interest which typically involve many degrees of freedom.

An alternative solution to generating the minimum energy path of a conformational change is to use a chain-of-states method. In such methods, two images are used as endpoint conformations, additional images are then generated between these two end points, and all are optimized at once. For biological systems, the chain-of-states method was pioneered by Elber & Karplus[52], whose plain elastic band method used first-derivatives to optimize the path between endpoint structures subject to restraints. Spring forces were added between the images which force each image to remain at an average separation between its partner images along the current path. The plain elastic band method is, however, highly dependent on the initial path chosen. This leads to minimum energy pathways that are local rather than global minima. Elber & Karplus proposed using simulated annealing as an optimization technique to find the global minimum energy path.

The plain elastic band method falls short in determining the transition state geometries and energies of the path because the spring forces which keep the images evenly spaced interfere with the energy of each independent image. For too rigid a spring constant, the images over-estimate the energies in the saddle point region, causing corner cutting and preventing the path from resolving saddle point structures. For too weak a spring constant, the forces on each image from the force field are too prominent, and images slide down the path back toward the minima and do not resolve the saddle points.[53]

To combat these problems, the nudged elastic band (NEB) method[53; 54; 55] was developed as a chain-of-states method which finds a minimum energy path of a conformational transition given only two initial structures to use as the energy minimized endpoints. The NEB approach implements a tangent definition, which is the tangent of the current path defined by the images. The tangent is calculated at every step in the simulation, and is used to decompose the force described by the force field and the spring forces into perpendicular and parallel forces with respect to the path.[56] The two endpoint images are fixed in space, and copies of these images can be used as initial seed structures for the pathway calculation. Additional images may be used to seed the starting path, but these are not exempt from the NEB force calculation as the endpoint structures are. Each image is connected to its neighbor images by the virtual springs along the path which contribute only to the parallel part of the force. The springs serve to maintain spacing of the images along the path. The force described by the force field is then only applied orthogonal to the path tangent, that is, projected out from each image and not along the bath between images. From their initial coordinates, the images are pulled into an interpolating path between the two endpoint structures. This initial path is then optimized[57; 58] to minimize the energy. The result is a minimum potential energy path that represents the conformational change between the two endpoint structures, independent of timescale.

The NEB method uses multiple simulations of the system, connected by springs, to map conformational changes along a path (Equation 3-1), except for the two endpoint structures which are fixed in energy and space. The tangent vector to the path is used to decouple the perpendicular component of the force, described only by the standard force field, and the parallel component, described only by the springs which keep the images evenly spaced along the path, shown in Equations 3-2 and 3-3, below. This way, the optimization of each discrete image is not

affected by the springs holding it evenly spaced along the path, and the force field does not cause images to move along the path. The tangent τ can be based on the position of the neighboring images or their energies, the latter yielding a more stable definition of the path.[56]

$$\vec{F}_i = \vec{F}_i^{\perp} + \vec{F}_i^{\parallel} \tag{3-1}$$

$$\vec{F}_i^{\perp} = -\vec{\nabla}V(\vec{P}_i) + \left(\left(-\vec{\nabla}V(\vec{P}_i)\right) \cdot \vec{\tau}\right)\vec{\tau} \tag{3-2}$$

$$\vec{F}_i^{\parallel} = \left[\left(k_{i+1}\left(\vec{P}_{i+1} - \vec{P}_i\right) - k_i\left(\vec{P}_i - \vec{P}_{i-1}\right)\right) \cdot \vec{\tau}\right]\vec{\tau} \tag{3-3}$$

In the above equations, the total force F on each image *i* is decoupled (Equation 3-1) to a perpendicular force and a parallel force by the tangent vector. The tangent vector defines the path between the two endpoint conformations at every image on the path, and when the dot product is taken with the potential defined by the force field for each image, represents the contribution of the force field for that image along the path. As shown in Equation 3-2, this is subtracted from the force field description (V) of each image (P_i) to remove any force contribution along the path from each image's individual potential. This component is termed the perpendicular force, since it represents each image as it moves in potential energy space normal to the path, fulfilling the requirement for the minimum energy path calculation.

The parallel component of the force accounts for the artificial springs linking each image together. In Equation 3-3, k_i is equal to the spring constant between images P_i and P_{i-1}, and P is the 3N-dimensional position vector of image *i*. Again, the tangent definition is used to subtract out the spring forces which act normal to the path. This way, the spring force which keeps the images evenly spaced along the path does not affect the relaxation of the individual images. The discrete images are maintained at specific intervals as to resolve saddle points; otherwise, they would fall back down to the local minima where they began.

3.2 Implementation

Despite the ability of NEB to find minimum energy paths, the previous implementation, within Amber, inherently imposes limitations in the choice of systems to be simulated. One issue is the requirement of applying NEB forces to all atoms of the system of interest, meaning system sizes in general had to be small to be computationally efficient to simulate. Applying NEB forces to all atoms necessarily restricted systems to implicit solvent. If water molecules exist, they could cause simulation artifacts since they would be forced to move between their locations at the endpoints, perhaps through the system itself, and not allowed to equilibrate in response to the conformational change. Since the goal of this work is to link biologically relevant structures of the Fpg-DNA complex in a MEP, further development of this method was necessary to accommodate large periodic systems, large scale conformational changes, and explicit solvent.

In the previous implementation of the NEB method in Amber9, any system used had to be set up inside the *addles* functionality of Amber, where all images distributed along the path were treated as one simulation. In that implementation one coordinate file was used, which grouped together all the coordinates of every image, and then split them into discrete structures at the end of the NEB simulation. The parameter file contained a copy of the system for each bead. The nonbonded exclusion list, which was the same for each image and not copied, was used to keep the images from interacting. Using particle mesh Ewald (PME)[95; 96] was impractical because of the large size (#beads * $(\#atoms)^2$) of the nonbonded exclusion list, each of which would need a distance calculation with PME to remove their interaction in reciprocal space. [97] This essentially limits the potential applications to small systems. Furthermore, the entire set of beads uses a single reciprocal space grid (which typically limits scaling due to fast Fourier transform calculations).

To address these issues, a new implementation of the nudged elastic band method to be released as part of Amber11 is discussed in this paper. In this new implementation the user may decide which parts of a system to apply the NEB forces to, for example, just the solute and not the solvent atoms. Rearrangement and rewriting of the code itself also allows for greater parallel efficiency, as well as accommodation of periodic systems, since the coordinates used for the rms fit can be defined as a subset of the image's atoms. The multisander functionality in the Amber MD engine (SANDER) runs multiple SANDER jobs simultaneously under a single MPI program, and has been applied to the NEB method, allowing each image to remain a discrete simulation, and allowing PME with no exclusion list beyond that for the standard system, and a separate reciprocal space calculation for each bead. This means each image writes output and trajectory information during its own MD simulation. This new implementation allows smaller systems to be simulated faster and larger systems to be more easily accommodated. The new method, termed partial nudged elastic band (PNEB), was validated using an alanine dipeptide model, for which the potential and free energy landscapes can be readily calculated using standard MD and minimization approaches thus allowing for a direct comparison. The PNEB method reproduces the minimum potential energy pathway of alanine dipeptide phi/psi isomerization in both implicit and explicit solvent. To convert the structural information along the path to free energies, umbrella sampling based on the PNEB path coordinates was performed with restraints, and 2D WHAM post processing of the umbrella sampling trajectories yielded free energies.

In the PNEB implementation, communication and distribution of the decoupled force calculation was changed from the previous version of the code. Many images are linked together and simulated simultaneously using the multisander functionality in Amber. In this way, each

image is simulated independently, and each can be distributed to multiple processors. An MPI communicator spanning each of these individual image simulations allows the transfer of the information necessary to perform the NEB force calculation. Each simulation along the chain is assigned a MPI master process; these masters perform the NEB force calculation. The force calculation routine calls the NEB force subroutine at each time step. For each image in the chain, with the exception of the first and last which are fixed in space, the coordinates of its two neighbor images are sent to each image's master process using MPI point-to-point communication. The master then performs an RMS fit of the neighbor coordinates to the image's self coordinates to remove rotational and translational motions for atoms in a user-specified mask. In this manner each image remains an individual continuous simulation where there is no change in the coordinates due to NEB imposed RMS fitting for each image. This is important for simulations with periodic boundary conditions. The master process gathers potential energies and determines spring constants, which it then uses to calculate and normalize tangent vectors depending on energy differences as described elsewhere[65]. Following this, the spring forces are calculated, and from all of this data the NEB forces are compiled and stored. After returning to the force calculation the NEB forces are broadcast to the other MPI threads for this image and added to the main force array for use in the parallelized coordinate/velocity integration step.

In the Amber11 implementation of NEB, an increase in communication timings when compared to a multisander simulation on the same system without the NEB force calculation can be expected (see Figure 3-1 and

Table 3-1). This is attributed to the master process requiring forces for all atoms in a NEB calculation, which is not needed in a standard parallel MD calculation. Additionally, at every time step, an MPI master must send and receive the coordinates of the neighbor images for each image and potential energies of every bead before performing the additional work of tangent normalization and calculating the spring forces. The NEB forces must be broadcast via MPI from each image to the main force array. All of these processes are in addition to calculating the dynamics of each image along the path, and yields increased timings when compared to standard multisander.

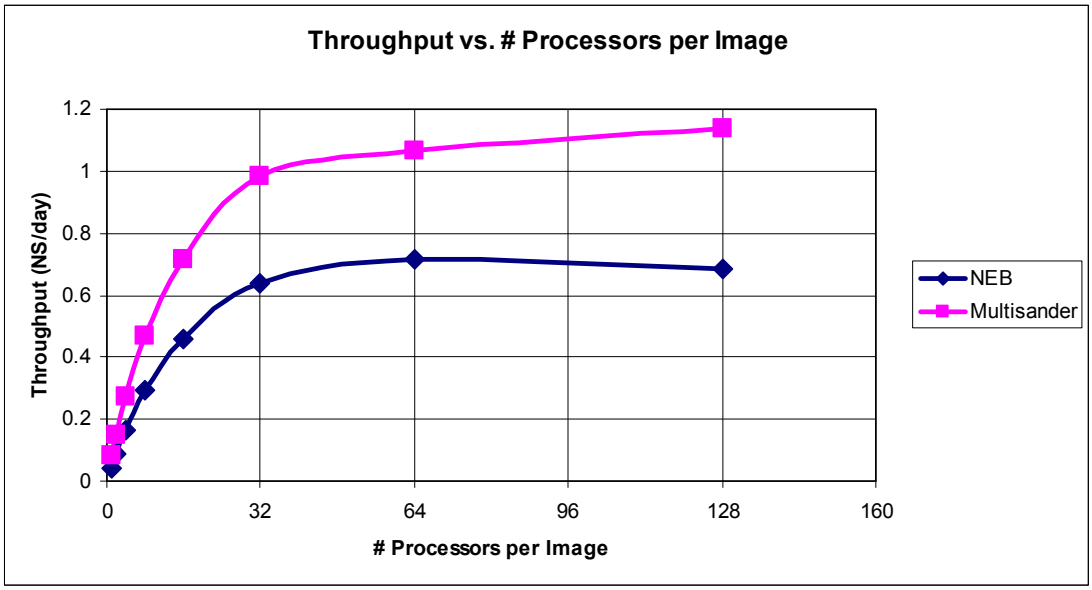

Figure 3-1 Throughput vs. processors per image for NEB (blue) vs. Multisander (pink).

Table 3-1 Benchmark comparison for NEB calculation (includes force decoupling) vs. Multisander calculation in Amber11. 60,000 atom system, 1000 steps MD, dt=0.001, 30 total images.

NEB	# Processors per image	Time (seconds)	Speedup (vs. 1 proc)	Efficiency (speedup/# proc)	% efficiency	Steps/Sec	NS/sec	NS/day
	1	2137	1.00	1.00	100.00	0.47	4.68E-07	0.04
	2	982	2.18	1.09	108.81	1.02	1.02E-06	0.09
	4	530	4.03	1.01	100.80	1.89	1.89E-06	0.16
	8	295	7.24	0.91	90.55	3.39	3.39E-06	0.29
	16	188	11.37	0.71	71.04	5.32	5.32E-06	0.46
	32	135	15.83	0.49	49.47	7.41	7.41E-06	0.64
	64	121	17.66	0.28	27.60	8.26	8.26E-06	0.71
	128	126	16.96	0.13	13.25	7.94	7.94E-06	0.69
Multisander	# Processors per image	Time (seconds)	Speedup (vs. 1 proc)	Efficiency (speedup/# proc)	% efficiency	Steps/Sec	NS/sec	NS/day
	1	1058	1.00	1.00	100.00	0.95	9.45E-07	0.08
	2	582	1.82	0.91	90.89	1.72	1.72E-06	0.15
	4	316	3.35	0.84	83.70	3.16	3.16E-06	0.27
	8	184	5.75	0.72	71.88	5.43	5.43E-06	0.47
	16	121	8.74	0.55	54.65	8.26	8.26E-06	0.71
	32	88	12.02	0.38	37.57	11.36	1.14E-05	0.98
	64	81	13.06	0.20	20.41	12.35	1.23E-05	1.07
	128	76	13.92	0.11	10.88	13.16	1.32E-05	1.14

An important additional feature of this new implementation allows the NEB forces to be applied to specific subsets of atoms. Masks in the input files designate specific atoms to apply NEB forces to as well as specific atoms to be broadcast for rms overlapping to remove rotational and translational motion. The atom selection uses the standard mask notation described in the Amber manual. In a SANDER input file with the ineb flag =1, the tgtrms mask denotes which atoms to apply NEB forces to, and the tgtfit mask denotes atoms to be broadcast for overlap of partner images to the self images. In this way, the NEB force decoupling is performed for only part of the system of interest, while the remainder is simulated with standard MD.

The tgtfit and tgtrms designations allow incorporation of explicit solvent, where the solvent atoms are not specified by either the tgtfit or tgtrms flags. This means that the NEB forces are not calculated for solvent molecules, nor are they used to overlap structures when removing the rotational and translational motion. The latter proves impossible, due to diffusion of water molecules which precludes overlapping between structures. Eliminating solvent from the NEB force calculation is important because this allows the solvent environment to react independently to the conformational change that is being studied. Additionally, communicating only part of the coordinates can dramatically improve parallel scaling. However, this approach necessarily means that more familiarity with the system of interest is required, and optimization methods must be carefully monitored to assure sufficient exploration of conformational space for the NEB part of the system while maintaining the integrity of the non-NEB part. Careful attention must be paid to annealing temperatures if this method of optimization is chosen otherwise non-physical structural deformations can be induced in the sections of the system that are not subject to NEB restraints.

Since only part of the system is simulated with NEB forces, it is important to monitor convergence of the non-NEB part of the system. The NEB forces ensure spacing of the NEB atoms along the minimum energy path. The non-NEB part of the system must be allowed more time to respond to this conformational change, since it essentially acts as a standard MD simulation. Since NEB is a timescale independent method, many of the conformational transitions this method could be applied to occur over long biological timescales. The NEB part of the system is simulated in a timescale independent manner, but the non-NEB part is not. However, in many cases the transitions in the non-NEB part of the system are expected to respond to that part of the system which is moving in a timescale independent manner, and

converge faster than they would if no NEB driving forces were applied to the region of interest. Testing this point could be done by comparing the continuity of any structural differences in the endpoints that were not included in the mask of atoms to which NEB spring were applied.

3.3 Method

3.3.1 Generating endpoints

This work was done using the Amber9 and Amber10 suite of programs. The alanine dipeptide (N-acetylalanyl-N-methyl-amide) was built from the sequence ACE ALA NME in the Leap module of the Amber9 suite of programs. The impose command was used to generate two structures with phi [C,N,CT,C] and psi [N,CT,C,N] angles of 66^{o} and 9^{o}, and -82^{o} and -6^{o}, respectively. The Amber ff99SB force field was used.[36] The coordinates were subjected to 400 steps of conjugate gradient minimization, with 5 kcal mol^{-1} $Å^{-2}$ restraints on all atoms except hydrogen. The implicit solvent model used in these calculations was the GB-HCT model (igb=1 in the SANDER module) with the mbondi radii set.[43; 98; 99]

Well minimized endpoints are critical for NEB simulations, since it is assumed that the transition endpoints are low energy structures. For explicit solvent simulations, the GB minimized coordinates were then solvated with 378 TIP3P[100] water molecules and further minimized and equilibrated. This round of minimization began with 1000 steps of steepest descent minimization with restraints on the dipeptide. The next step was 25ps of restrained MD to increase the system temperature from 0K to 300K. Two subsequent rounds of steepest descent minimization were performed, in which the restraints were reduced to 0.1 kcal mol^{-1} $Å^{-2}$. This was followed by 25ps each of equilibration with restraints on all atoms except water and

hydrogen at 1.0 kcal mol^{-1} $Å^{-2}$ which were reduced to 0.5 kcal mol^{-1} $Å^{-2}$. The last steps of equilibration were to reduce the restraints to just the backbone atoms, in two rounds of 25ps with restraints of 0.5 kcal mol^{-1} $Å^{-2}$ decreasing to unrestrained MD for 50ps. This same equilibration was performed for both endpoint structures. After the initial 400 step minimization mentioned above, the structures were subjected to 10ns of unrestrained dynamics.

3.3.2 Using PNEB to generate multiple paths using two solvent models and different atom mask selections

Simulated annealing was used to optimize the alanine dipeptide isomerization pathway. This protocol was adapted from Mathews & Case 2006.[65] The longer simulated annealing protocol described was used here, with the exception of the high temperature range which was taken from the shorter protocol description. Here, a 760ps protocol was used, and the communication of atoms was specified by the tgtfit mask only in the case of PNEB runs. For the first 40ps the system was heated to 300K with a Langevin[101] collision frequency of 1000 ps^{-1} and pulled into an interpolating path using 10 kcal mol^{-1} $Å^{-2}$ spring forces. The system equilibrated along this path for 100ps after increasing the spring forces to 50 kcal mol^{-1} $Å^{-2}$ where they remained for the rest of the NEB simulation. Heating of the system from 300K to 500K and subsequent cooling back to 300K was performed stepwise over 300ps. Further cooling to 0K to remove kinetic energy occurred over 120ps, and quenched MD was run for 200ps. SHAKE was used to constrain bonds to hydrogen.[92] Structure distribution of images along the minimum potential energy path at each step in this optimization procedure is shown in Figure 3-2. The same protocol was used for all three systems, regardless of solvent environment, to examine

consistency of the method. A lower collision frequency is more likely relevant when solvent is present, as it is related to the viscosity of the system.

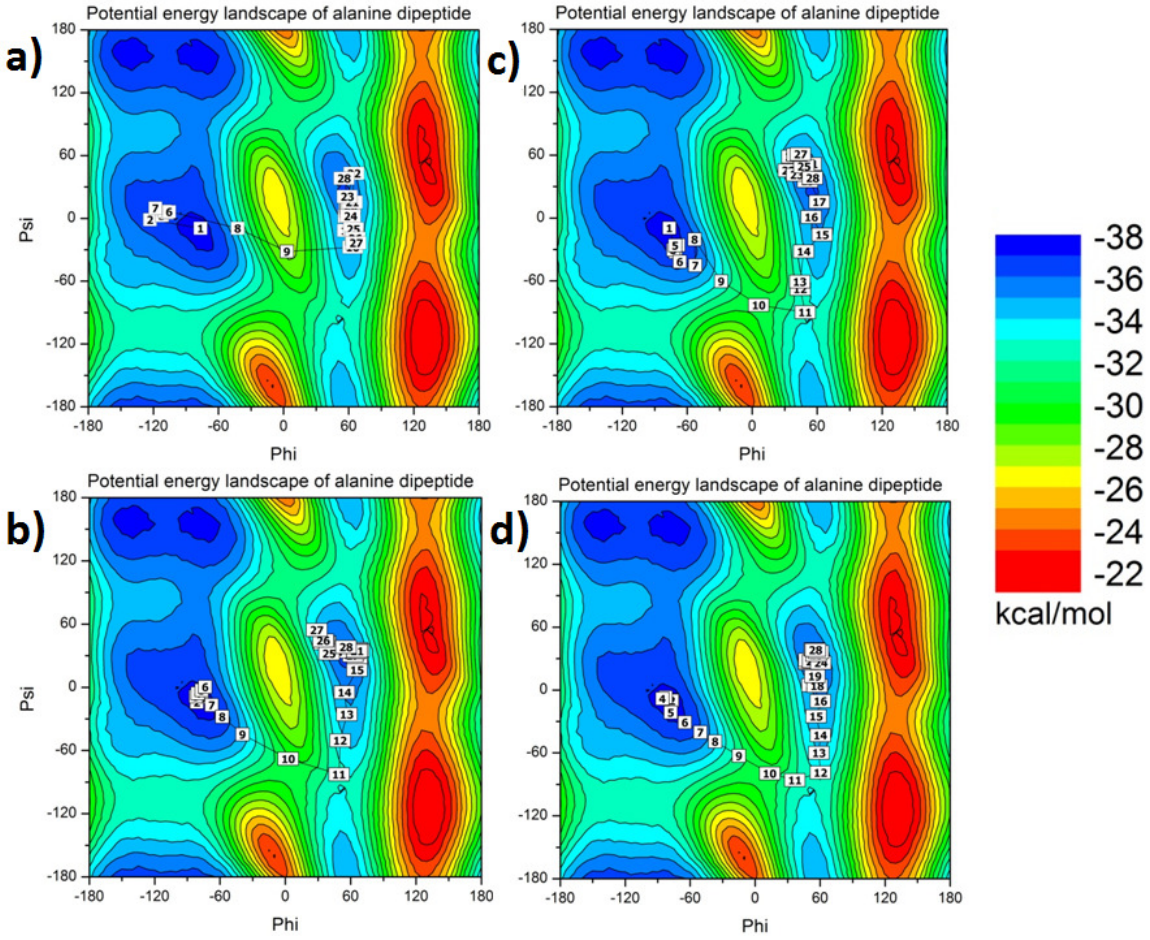

Figure 3-2 Structure distribution along the path at different levels of optimization in PNEB simulated annealing protocol, plotted on the 2D potential energy surface of alanine dipeptide. a) Initial interpolating path between left handed helix and right handed helix minima, b) After equilibration of initial path, c) Images after simulated annealing steps, d) Images in final minimum potential energy path distribution.

The potential energy surface was calculated by restrained energy minimization of alanine dipeptide at 2^{o} intervals over the entire phi/psi plane. 50 kcal mol^{-1} rad^{-2} restraint forces on the

phi and psi angles were used, and 500 steps of steepest descent minimization were followed by 500 steps conjugate gradient minimization.

3.3.3 Generating 2D free energy surface

A common way to calculate free energy using Molecular Dynamics is to use the umbrella sampling technique. The free energy surface of the alanine dipeptide system was calculated using umbrella sampling by first building and minimizing independent structures where the phi and psi were rotated every 5^o to cover the complete periods (-180^o to 180^o) of both backbone dihedrals, totaling 1296 simulations. During the minimization process the phi and psi dihedrals were restrained with a force constant of 500 kcal mol^{-1} rad^{-2} and the minimization was run for 1000 steps. Each minimized phi/psi structure was used as the starting structure of an independent umbrella simulation, yielding 1296 windows. The force constant for the production dynamics was 75 kcal mol^{-1} rad^{-2} on the phi and psi dihedrals, and each window was simulated for 10ps at 300K. The phi and psi values were saved at every time step. Data was postprocessed into free energy profiles using the Weighted Histogram Analysis Method (WHAM)[68; 69; 70], using a program provided by Alan Grossfield, (freely available at http://membrane.urmc.rochester.edu/Software/WHAM/WHAM.html).

The umbrella sampling and subsequent free energy calculation for the PNEB path were performed using starting structures from the last step of the PNEB simulated annealing protocol, the 200ps quenched MD simulation. A grid was mapped along the path at every 10^o of phi and psi using a 20^o buffer in each direction along the PNEB path. The grid indicating initial phi/psi angles of each window, overlapped with the PNEB path, is shown in Figure 3-3. Structures closest to these values from each bead's 200ps trajectory were restrained to the angles described

by the grid with a 75 kcal mol^{-1} rad^{-2} restraint force. 25ps of dynamics at 300K were run for each of the 147 windows. Overlapping populations of each window are shown in Figure 3-3. 2D WHAM analysis was used to calculate free energies.[102]

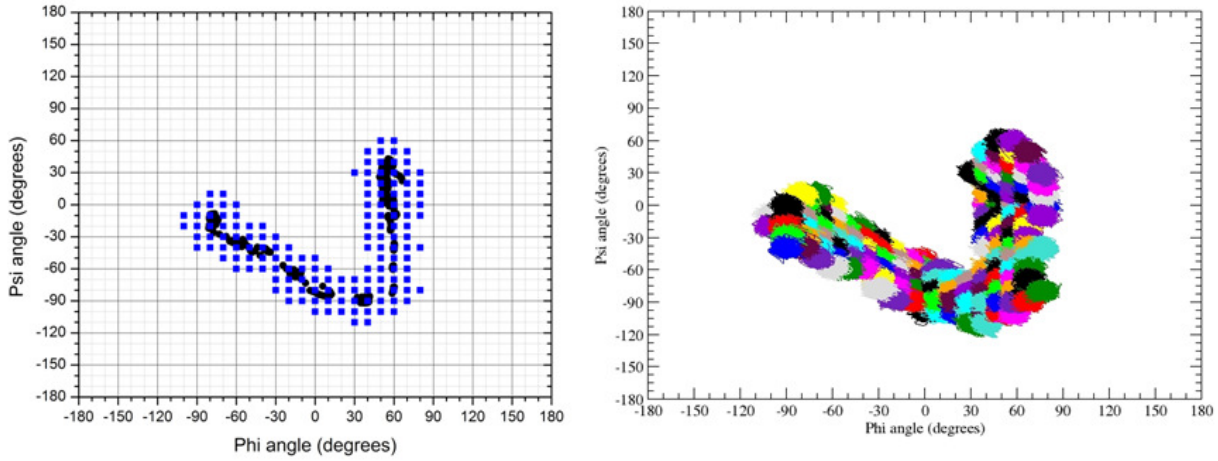

Figure 3-3 Left: Initial distribution (shown in blue squares) of umbrella sampling starting structures from PNEB path (shown in black dots). Right: Population distribution of each umbrella sampling window used for 2D WHAM analysis.

3.4 Results and Discussion

3.4.1 Calculation of MEP

The PNEB method was tested on the isomerization of an alanine dipeptide **(**Figure 3-4**)**. This system was chosen because of its small size and the relative ease of determining its underlying energy surface in both implicit and explicit solvent. The energy surface of the alanine dipeptide is dominated by rotations around the phi and psi angles, and can be represented topographically as seen in Figure 3-5, where restrained minimizations at 2^{o} intervals give

potential energy over the entire plane. The starting conformations of the alanine dipeptide were chosen from the right handed helix (negative phi and psi angles) and left handed helix (positive phi and psi angles) basins. 15 copies of each image were used as the initial structure distribution along the path (30 images total).

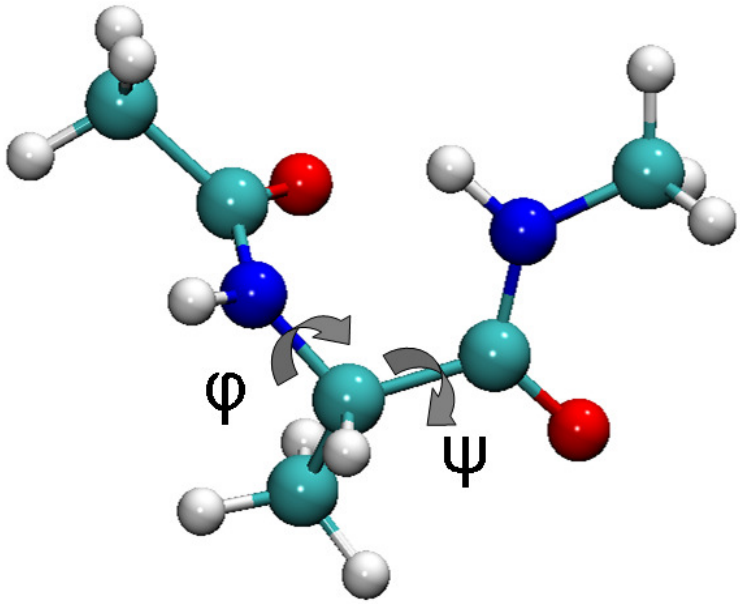

Figure 3-4 The structure of alanine dipeptide. Rotation around the phi (φ) and psi (ψ) dihedral angles dominates the energy landscape.

The purpose of performing alanine dipeptide simulations in implicit solvent was to verify that changes made to the force decoupling and communication would not change the behavior of the method in finding the minimum energy pathway since full NEB could also be used. The minimum energy path reported in Figure 3-5 shows that the standard NEB implementation, where forces are applied to every atom in the system, returns a minimum energy path consistent with previously published results.[58; 65] The path proceeds perpendicular to the contour lines on

the underlying potential energy landscape, and passes through the saddle point, indicating it is a minimum energy path.

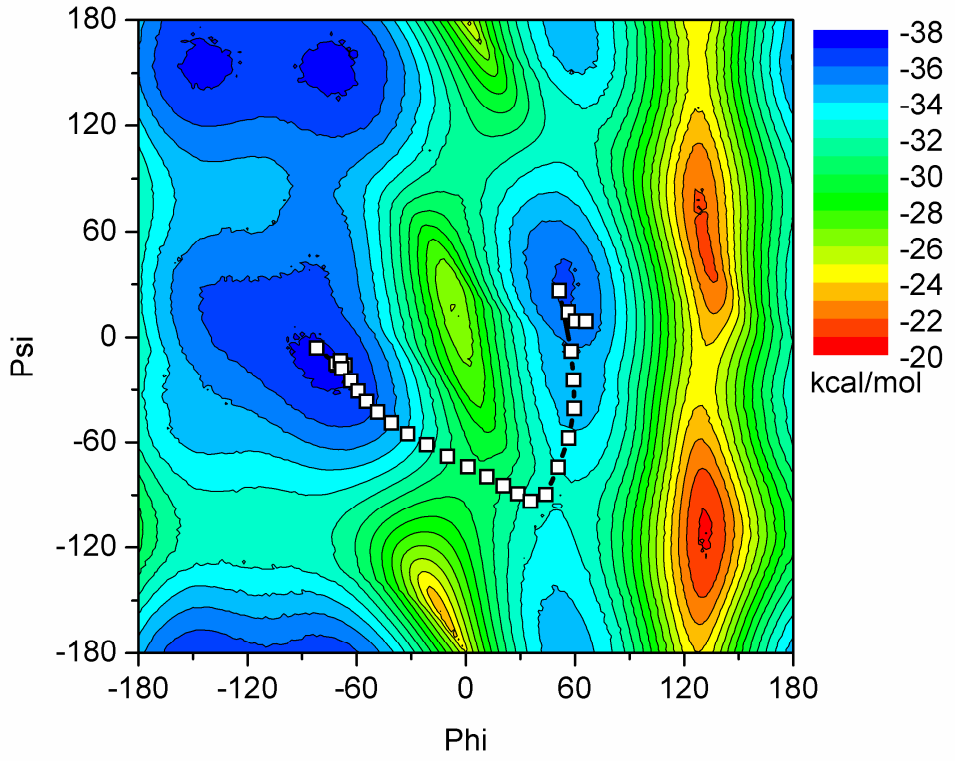

Figure 3-5 The potential energy landscape for alanine dipeptide isomerization around the phi/psi dihedral angles. Minimum energy path for standard NEB implementation is shown using squares for each bead. Contours are shown at 2 kcal/mol intervals. The surface was generated by performing restrained minimizations at every 2^{o} of the phi and psi angles.

3.4.2 Comparison between two solvent models and different atom mask selections

In the standard NEB test performed on the alanine dipeptide system, NEB forces were applied to every atom, shown colored by atom name in Figure 3-6a. To test the implementation of the PNEB method, and assure it returned the same minimum energy path as the standard NEB implementation, two test systems were utilized where the NEB forces were applied to a subset of the atoms of the alanine dipeptide. The PNEB implementation was verified for non-periodic and periodic systems by performing two tests. In the first test, NEB forces were applied to only the backbone C and N atoms of an implicitly solvated alanine dipeptide, as specified by the orange atoms in Figure 3-6b. The second test of the method was to apply NEB forces to the same backbone atoms of an explicitly solvated alanine dipeptide, shown in Figure 3-6c. Rotational and translational motion between images was effectively removed before the tangent calculation by RMS fitting on every atom of the alanine dipeptide for both of these systems.

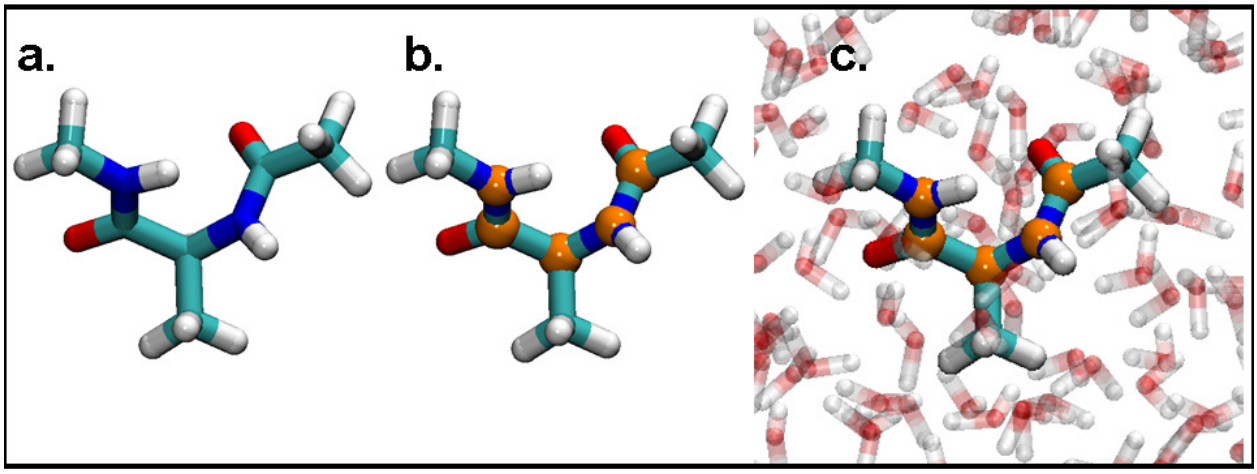

Figure 3-6 Alanine dipeptide test systems: Alanine dipeptide in implicit solvent with a) NEB forces applied to all atoms b) NEB forces applied to atoms in orange spheres. c) Alanine dipeptide in explicit solvent with NEB forces applied to atoms in orange spheres.

In Figure 3-7 it is shown that the test systems for PNEB return the same minimum energy path between the endpoint conformations. In all three cases, the minimum energy transition follows a path from the right handed helix (negative values of phi and psi) to a saddle point of about (50^{o},-100^{o}) and from there to positive phi/psi values in the left handed helix.

Even though all three systems reproduce the same overall minimum energy path, the paths differ slightly. The endpoint structures undergo a conformational motion that is not represented by the phi/psi plane. For the standard NEB in implicit solvent, shown in Figure 3-7a, the movement in the images close to the endpoint structures is attributed to rotation of the end methyl groups, which is a conformational motion not represented by the phi/psi angles. The PNEB motion would not account for movement in the hydrogen atoms, since these are not included in the part of the molecule for which we calculate NEB forces. What occurs in Figure 3-7b is the methyl group of the C terminus rotates from the right handed helix endpoint structure,

which is at an imperfect angle, to a slightly more stable structure where the angle along the C-N bond becomes completely anti conformer with regards to the rest of the structure. The N terminus methyl group also rotates its hydrogen away from the carbonyl oxygen. In Figure 3-7c, when explicit solvent PNEB is used, there is a larger population of structures which remain near the ending points than in the other two simulations. This is due to a few images' trajectories along the path forming a hydrogen bond between the NH group of the N terminus and the oxygen of the C terminus, which is energetically unfavorable to break. This hydrogen bond forms more readily due to the inclusion of explicit solvent, and it is less favorable to form during implicit solvent simulations.[103] Since we are using PNEB, however, the NEB forces applied to the backbone atoms ultimately force the images along the minimum energy path to sample the transition state region.

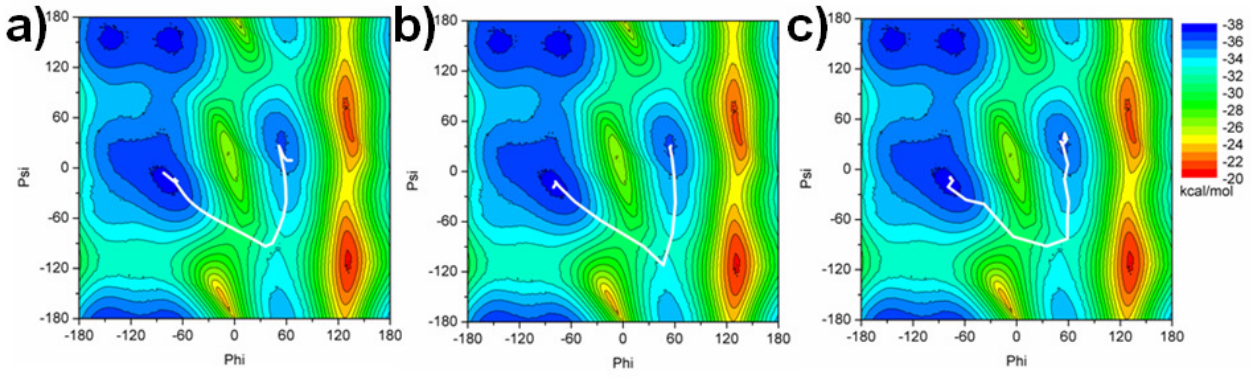

Figure 3-7 Potential energy surface of alanine dipeptide with minimum energy paths determined by a) standard NEB, b) PNEB in implicit solvent, and c) PNEB in explicit solvent. The minimum energy path is reproducible between all three systems.

3.4.3 Comparison to 2D free energy surface

Although the NEB method provides the user with a minimum potential energy path of a conformational transition, in some cases it may be desirable to determine free energies along this path. 2D-WHAM analysis was used to generate free energies along the PNEB derived minimum energy path. For many systems, the problem with applying umbrella sampling is twofold; the path cannot be easily defined by a small number of reaction coordinates that can be used as restraints, or the transition that is forced to occur along the reaction coordinate introduces artifacts along the transition path. In the alanine dipeptide example, the definition of a two dimensional reaction coordinate along which the isomerization occurs is clearly described by the phi and psi angles. For larger systems with more degrees of freedom, PNEB itself can define important reaction coordinates upon analysis of the resulting path. As shown in Figure 3-8a, the 2D WHAM procedure which uses PNEB derived structures as a starting path returns reasonable free energies. Figure 3-8b shows that the free energy along the NEB path is almost identical to that from the full free energy surface calculated independently. In this way we can generate free energies along the minimum potential energy path calculated with NEB.

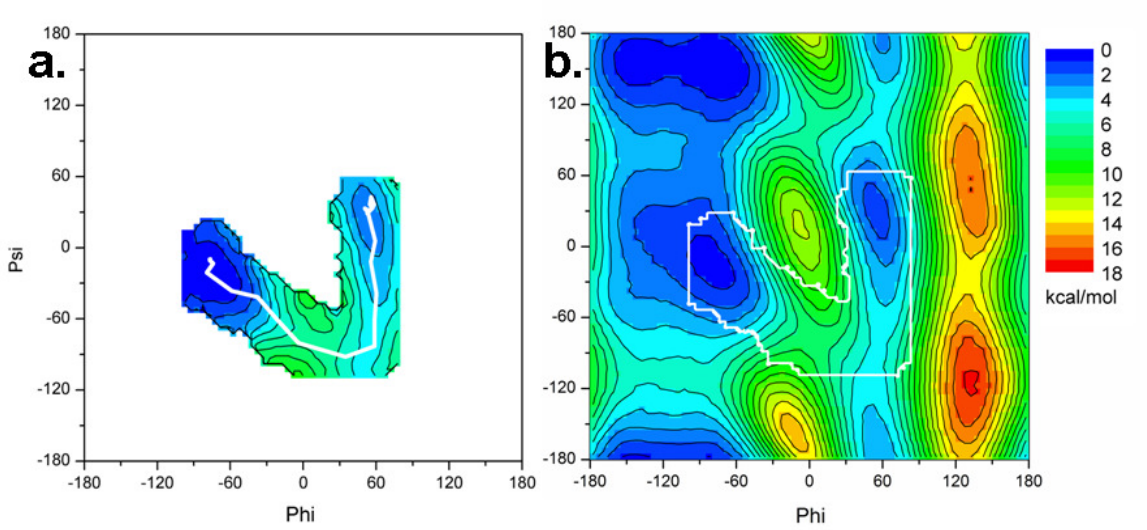

Figure 3-8 a) Free energy profile in the region of the transition calculated using the PNEB path as a starting point. b) Full free energy surface of alanine dipeptide calculated by restrained umbrella sampling followed by 2D-WHAM. Boundary corresponding to a) is outlined in white to show similarity between calculated free energy using only the PNEB path and that calculated for the entire surface.

3.5 Conclusion

The NEB method available in the Amber10 suite of programs has been updated to a partial nudged elastic band method where the NEB calculation is performed on a subset of the system, defined by user specified masks. Each image along the path is an independent MD simulation coupled to the others in a coarse grain manner through an MPI communicator. Testing on the alanine dipeptide model system, where the underlying potential energy surface can easily be determined, shows that the partial NEB method returns the same minimum energy pathway as the standard NEB method, even when the system is explicitly solvated. This verifies

the method performs the same function, namely, finding the minimum energy pathway of a conformational change, even when only a subset of the system's atoms are specified for the NEB force decoupling.

It is worth noting, however, that the alanine dipeptide system is only a model, chosen because the "answer," in this case, the potential energy surface, is straightforward to calculate. It has been shown that dipeptide phi/psi preferences are different from phi/psi behavior in larger protein systems.[104] This presents a unique challenge in adapting this method to the larger Fpg-DNA protein system: as the number of degrees of freedom in a system increase, the conformational changes become more difficult to quantify.

The Fpg-DNA complex in particular has many challenges. There are not only specific, local conformational changes due to the everting 8OG base, but larger global changes in DNA bending and protein domain motion. Additionally, the presence of solvent which makes up the bulk of the interactions precludes analysis of potential energy alone as indicator of the preferred pathway for the transition. The next chapter details how PNEB is used in this system to generate and optimize MEPs, and subsequent umbrella sampling is used to validate the global minimum free energy path.

4 ENERGETIC PREFERENCE OF 8OG EVERSION PATHWAYS IN FPG

Acknowledgements

The material presented in this chapter contains direct excerpts from the manuscript *Energetic Preference of 8-oxoG Eversion Pathways in a DNA Glycosylase* by Christina Bergonzo, Arthur J. Campbell, Carlos de los Santos, Arthur P. Grollman, and Carlos Simmerling, published as a Communication in the Journal of the American Chemical Society in 2011, volume 133, pages 14504-14506. The manuscript was written by Christina Bergonzo with suggestions and revisions by Carlos de los Santos, Arthur P. Grollman, and Carlos Simmerling. NEB simulations, umbrella sampling simulations and analysis were performed by Christina Bergonzo and umbrella sampling simulations were performed by Arthur J. Campbell.

4.1 Introduction

Base eversion is a fundamental process in the biochemistry of nucleic acids, allowing proteins engaged in DNA repair and epigenetic modifications to access target bases in DNA. Crystal structures reveal endpoints of these processes, but not the pathways involved in the dynamic process of base recognition. To elucidate the pathway taken by 8-oxoguanine during base excision repair by Fpg, free energy surfaces were calculated of the damaged base during eversion through the major and minor grooves. The minor groove pathway and free energy

barrier (6-7 kcal/mol) are consistent with previously reported results.[22] However, eversion of 8OG through the major groove encounters a significantly lower barrier (3-4 kcal/mol) more consistent with experimentally determined rates of enzymatic sliding during lesion search.[74] Major groove eversion has been suggested for other glycosylases, suggesting that in addition to function, dynamics of base eversion may also be conserved.

8-oxoguanine (8OG), one of the most abundant products of oxidative DNA damage in cells, arises from oxidation of guanine by endogenous and exogenous reactive oxygen species.[105] 8OG readily pairs with adenine in a Hoogsteen orientation; during DNA replication, this conformation results in a transversion mutation.[15] Base excision repair (BER) pathways in eukaryote and prokaryote cells excise 8OG to preserve genomic integrity. In prokaryotes, formamidopyrimidine-DNA glycosylase (Fpg), also known as MutM, excises 8OG when the oxidized base is paired to cytosine.[106] Human oxoguanine glycosylase (hOGG1) is a functional homologue of Fpg, but the two enzymes share little structural homology. [107]

The base eversion mechanism is not limited to positioning damaged bases for enzymatic catalysis, and also plays a role in determining selectivity of substrates. Eversion of bases from the helix facilitates recognition of substrates. For example, studies of uracil DNA glycosylase (UDG) using imino proton exchange NMR indicate that UDG has a passive role in trapping spontaneously opened bases, the rates of which are increased for the substrate.[108] However, the relatively low barrier to sliding of hOGG1 and Fpg (about 2 kcal/mol per base pair), as determined by single-molecule fluorescence studies[74], suggests these glycosylases do not extrude every base into an extrahelical site, and that substrate recognition takes place at an earlier stage.

4.2 Methods

4.2.1 Endpoint structure homology modeling

Fpg-DNA systems were built using the 2F5O[86] crystal structure of Fpg crosslinked to DNA interrogating an undamaged intrahelical guanine base for the intrahelical endpoint, and the 1R2Y[19] crystal structure of a catalytically inactive mutant trapping 8OG in the active site for the extrahelical endpoint. Disulfide crosslinks used to trap the complexes were not included in these simulations. The DNA sequence of the 1R2Y structure was standardized to match that of the 2F5O structure by manually deleting the nucleobase from the pdb and building in the standard (2F5O) sequence nucleobases using the Leap module of AMBER.[109] The standard sequence is as follows: 5'-(AGGTAGATCCGGACGC)-3' and 5'-(TGCGTCC**8OG**GATCTACC)-3'. Additionally, the intrahelical endpoint derived from the 2F5O structure required modification to add coordinates for the catalytic loop (residues 217 to 237 inclusive, missing in the electron density map) based on the ordered 1R2Y structure. Residue 2 of the 1R2Y structure was mutated back to its wild type sequence of E from Q. The N-terminal proline was modeled as neutral.[110] Parameters for 8OG were taken from Miller et al.[88] Zinc was modeled using the Stote non-bonded model.[111] The 2F5O structure is referred to as "intrahelical" and the 1R2Y structure as "everted."

4.2.2 Generating a MEP using PNEB

Molecular dynamics (MD) simulations were carried out with the Sander module of the Amber10 suite of programs.[109] The Amber ff99SB force field[89] with the parmbsc0[42] corrections for nucleic acid backbone torsions was used. During minimization and equilibration constant pressure (1atm) and temperature (330K) were maintained using the weak-coupling algorithm with a coupling time constant of 0.5ps. [91] SHAKE was used to constrain bonds to hydrogen.[92] A time step of 1 fs was employed. The particle mesh Ewald method was used for calculating electrostatic energy with a nonbonded cutoff of 8Å on direct space interactions.[95; 96]

The structures were initially minimized using the GB-OBC model and mbondi2 radii set.[112] Coordinates were restrained with 5 kcal mol^{-1} positional restraints on all heavy atoms and minimized using 200 steps of steepest descent minimization, followed by 200 steps of conjugate gradient minimization. The minimized structures were solvated using ~8,000 TIP3P[90] explicit solvent molecules in a truncated octahedron box with 8Å distance between the solute and boundary. They were heated linearly over 100ps to a target temperature of 330K, using positional restraints on all atoms of the protein and DNA, allowing only solvent to move. This was followed by five 1000-step steepest descent minimization rounds in which restraints on side chains and bases were gradually decreased from 10.0, 5.0, 2.5, 1.0 and 0.0 kcal mol^{-1} A^{-2}. Four 5000-step cycles of MD with decreasing restraints on the protein and DNA backbone were then performed. A 20ns NVT unrestrained dynamics simulation was carried out using a Langevin thermostat with a collision frequency of 1.0 ps^{-1} and a timestep of 2fs.[101]

PNEB path calculations for the minor groove path were performed by linking 16 copies of the intrahelical structure and 16 copies of the everted structure, to total 32 structures along the

path. The major groove path needed to be seeded with an intermediate structure. Targeted MD was performed over 30ps using the intrahelical structure, with a restraint value of 0.0305 kcal mol^{-1} deg^{-2} and a target value of 195° of the eversion dihedral.[73] The restraint mask was applied to heavy atoms of the entire protein and DNA backbone for targeted MD. 30 frames from the targeted MD trajectory were used to seed the PNEB major groove pathway, in addition to the intrahelical and everted endpoint structures, totaling 32 structures along the starting PNEB path.

For PNEB and umbrella sampling simulations, the NVT ensemble was used. NEB forces were calculated for all atoms of the solute and not the solvent. A simulated annealing path optimization procedure was adapted from Mathews and Case 2006.[65] Initial spring forces of 10 kcal mol^{-1} $Å^{-2}$ and an initial collision frequency of 100 ps^{-1} were used. The first stage of path optimization for 80ps at 330K was followed by equilibration of all structures at this temperature for 500ps. Heating of the path to 386K took place over 60ps, with an increased spring constant of 25 kcal mol^{-1} $Å^{-2}$ and decreased collision frequency of 75 ps^{-1}. Equilibration at 386K for 220ps was followed by annealing to 330K over 100ps. Final equilibration at 330K to generate populations for each structure was performed over 500ps (Figure 4-1).

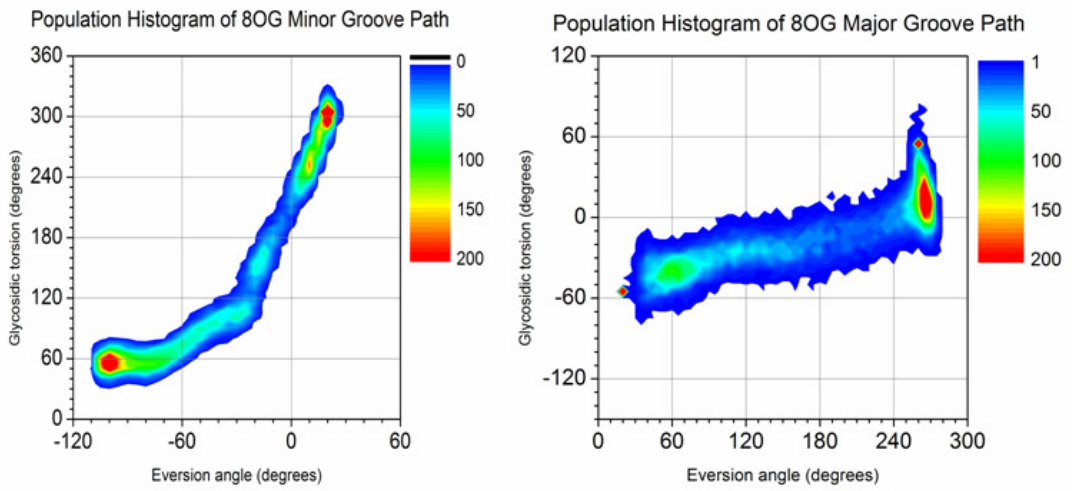

Figure 4-1 Population histogram of PNEB data during final equilibration at 330K over 500ps. Scale is given in number of structures. Left: Minor groove path. The intrahelical structure is located at (20,300) and the everted structure is located at (-100,60) and eversion proceeds from right to left. Right: Major groove path. The intrahelical structure is located at (20,-60) and the everted structure is located at (260, 60) and eversion proceeds from left to right.

4.2.3 Grid spacing and umbrella sampling

The process used to generate PMFs from PNEB structures using umbrella sampling was the same as described previously, and is illustrated in Figure 4-2.[113] Each window's starting structure was restrained with a force constant of 0.0609 kcal mol^{-1} deg^{-2} in both the eversion and glycosidic torsion angles and was simulated in the NVT ensemble for 250 ps. The temperature of each window was held constant at 330K using a Langevin thermostat with a collision frequency of 75.0 ps^{-1}.[101] For the minor groove simulations, a total of 201 and 236 windows were run with 10^{o} spacing for the glycosidic torsion and eversion angles. For the major groove simulations, a total of 576 and 575 windows were run with 5^{o} spacing for the eversion angle and 10^{o} spacing

for the glycosidic torsion. 2D WHAM analysis was used to calculate free energies of the combined minor and major groove pathways.[102]

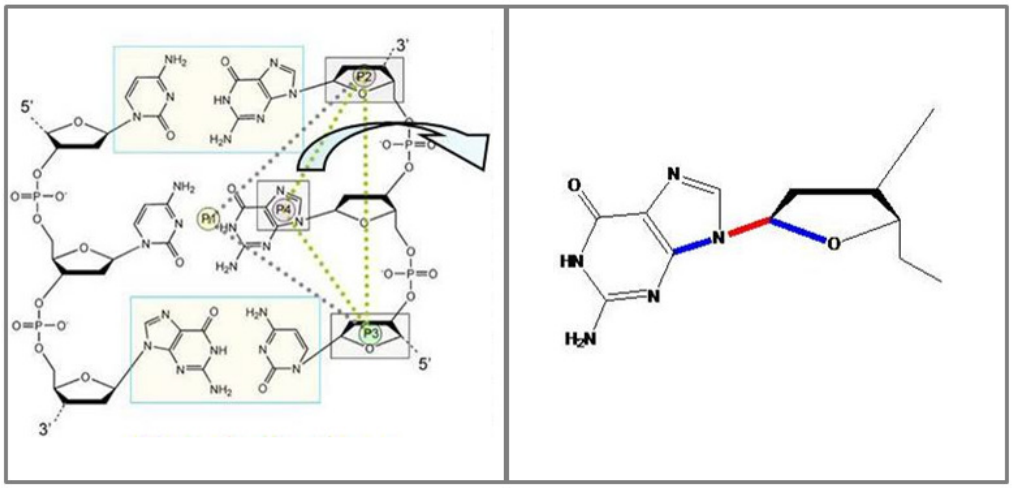

Figure 4-2 Left: Schematic representation of the eversion dihedral reaction coordinate based on the center of mass of the flanking base pairs as point one, the center of mass of the heavy atoms of the flanking bases' sugar rings as points two and three, and the heavy atoms of the imidazole ring as point four. Right: Representation of the glycosidic torsion used as a reaction coordinate measuring the local rotation of the base.

4.2.4 Convergence analysis of umbrella sampling protocol

To assess convergence, two entirely different sets of starting structures were taken from each PNEB pathway and the umbrella sampling procedure was performed independently for each set (Figure 4-3). The resulting free energy profiles are in agreement within 1-2 kcal/mol, indicating the method for calculating free energy converges for two different sets of starting structures. Additionally, the differences in free energy between the major and minor groove paths are outside this 1-2 kcal/mol margin of error.

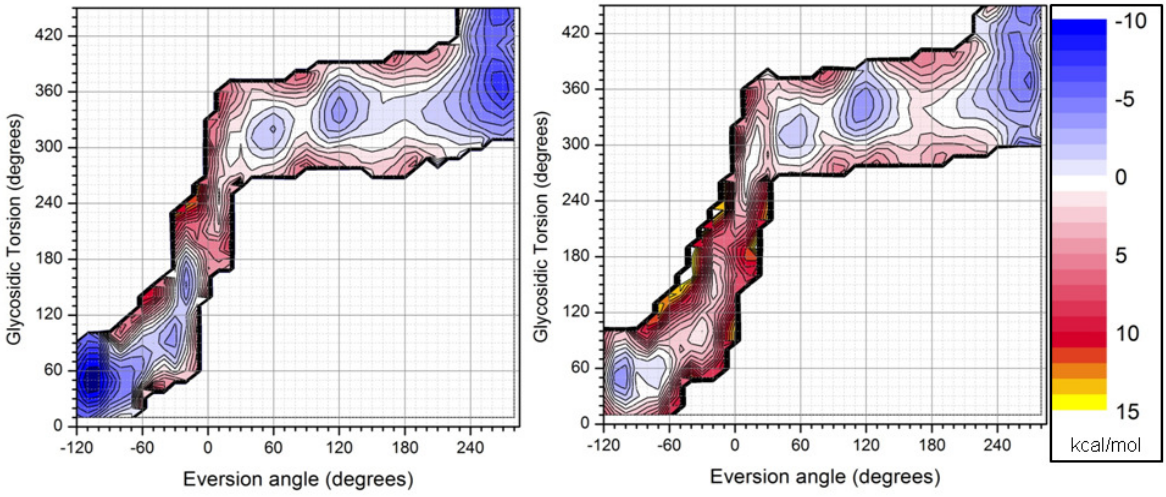

Figure 4-3 PMF profiles from independent initial structure sets.

4.3 Results and Discussion

The low energy states of BER DNA glycosylases in general, and Fpg and hOGG1 in particular, have been characterized, primarily by X-ray crystallography. Catalytically inactive Fpg mutants allow trapping of substrates containing 8OG in the active site.[19] Intermolecular crosslinking has been used to generate structures of putative intermediates in the catalytic pathway[18; 114] as well as an intrahelical structure, interpreted as initial binding of Fpg to 8OG damaged[22] and undamaged[21] DNA. Figure 4-4 illustrates the eversion pathways of 8OG through both the major and minor groove as determined by PNEB. Visual inspection suggests the minor groove pathway as the more direct route. However, crystal structures of hOGG1 complexed to DNA have been reported with bases in the major groove at putative intermediate locations along the eversion pathway.[115; 116] UDG crystallography also points to a major groove pathway to avoid unfavorable steric clashes with protein side chains in the minor groove.[117]

4.3.1 Minor groove vs. Major groove pathway

Molecular dynamics (MD) provides an unparalleled ability to simultaneously connect structure, dynamics and energy for conformational changes. To generate a minimum potential energy path (MEP) from the intrahelical to the extrahelical endpoint, the partial nudged elastic band (PNEB)[113] implementation in Amber10[109] was used with all-atom simulations in explicit water. Both paths begin and end with the same intrahelical and extrahelical coordinates, and are continuous in all solute degrees of freedom. The optimized, time-independent paths are shown in Figure 4-4 (Population histograms along each PNEB path are provided in Figure 4-1). The

minor groove path rotates around the glycosidic torsion early in eversion to minimize unfavorable steric clashes and form favorable electrostatic interactions with the intercalating triad (M76/R111/F113, see Figure 4-4). The major groove path shows no steric hindrance for initial breaking of the intrahelical hydrogen bonds.

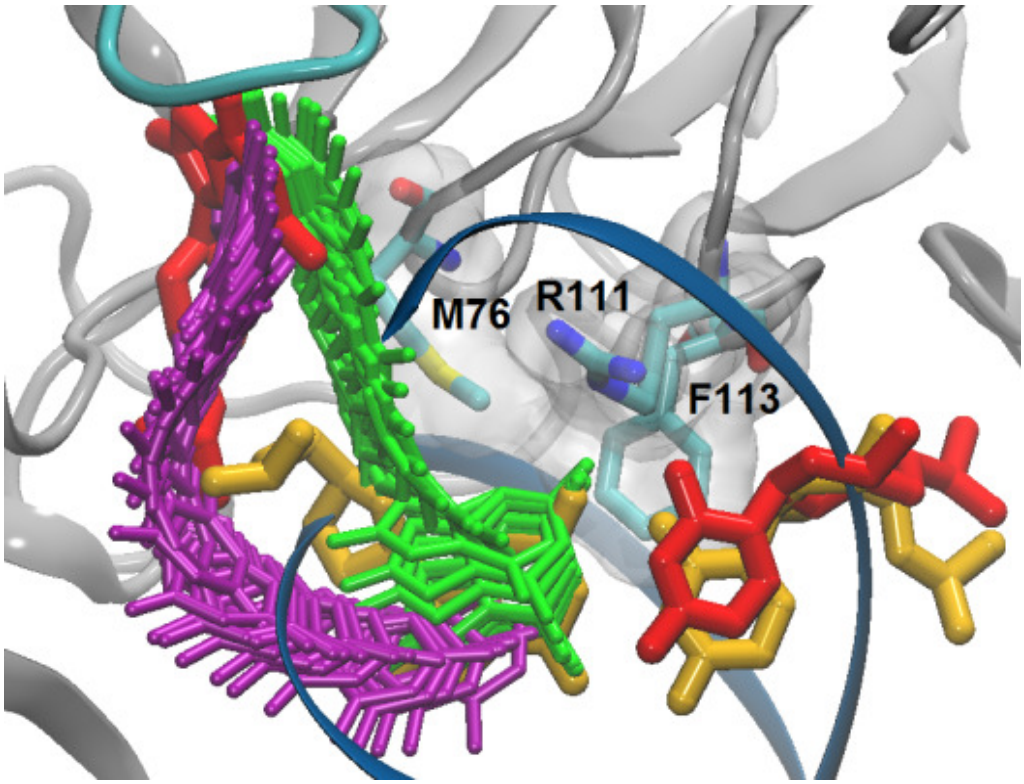

Figure 4-4 Base eversion paths from PNEB calculations. Paths are indicated by the last frame from each PNEB bead's independent trajectory, with major groove shown in purple and minor groove in green. The intrahelical 8OG:C base pair is shown in yellow, the extrahelical 8OG:C base pair is shown in red, and the catalytic loop is shown in cyan. Intercalating wedge residues M76/R111/F113 are shown colored by atom name and with transparent surfaces to illustrate the extent of their insertion into the minor groove. Positions of these residues are based on the intrahelical endpoint structure.

Two-dimensional umbrella sampling was performed to obtain free energy profiles along these paths, using two previously reported reaction coordinates that measure base eversion and the glycosidic torsion rotation.[73; 118] The resulting potential of mean force (PMF), shown in Figure 4-5 reveals free energy profiles for both pathways. Eversion through the minor groove encounters an energetic barrier of 6-7 kcal/mol; this observation is consistent with the ~10 kcal overall barrier to extrusion into the enzyme active site reported in Qi et al. (who studied only the minor groove path).[22] Although a small energy penalty is incurred upon breaking intrahelical hydrogen bonds, the main barrier to entering the active site corresponds to initial rotation of the glycosidic torsion.

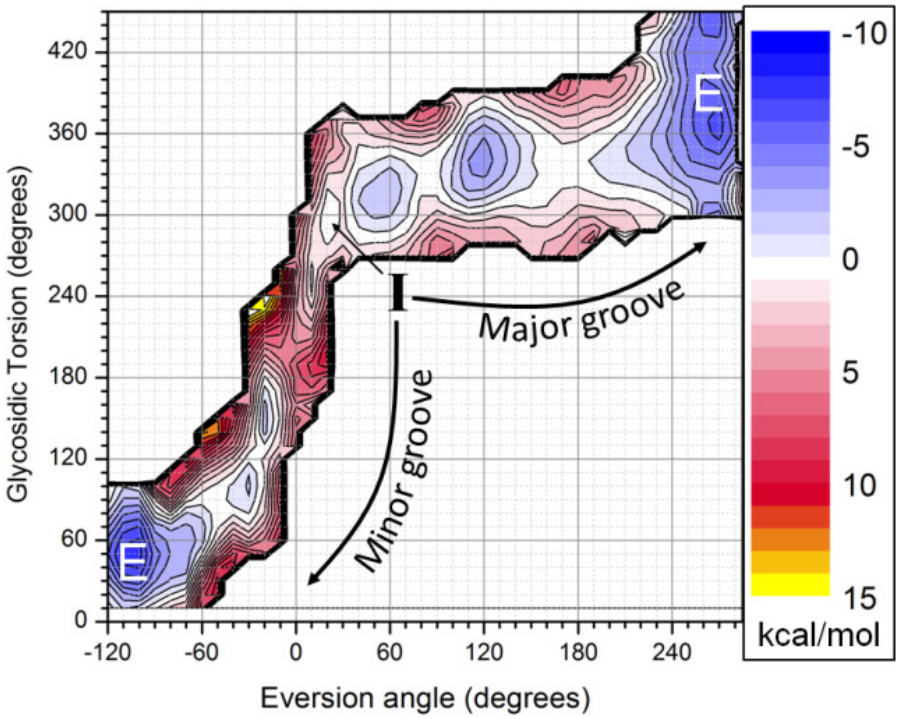

Figure 4-5 Free energy pathways of base eversion via the major (top) and minor (bottom left) grooves. The initial intrahelical state "I" is in the middle, and the final extrahelical state "E" is reached on both the right and left sides of the graph. Free energies are given in kcal/mol.

The major groove PMF indicates a maximum overall barrier of 3-4 kcal/mol, significantly less than through the minor groove. As suggested for UDG,[117] the major groove pathway avoids steric clashes, and instead the lower barrier arises from eversion past the β-hairpin of the zinc finger motif . 8OG partially rotates to form a stable intermediate (eversion angle = 120^{o}) that may have an important role in substrate recognition. Two minima in the active site (labeled "E" for extrahelical) have glycosyl torsion values of 360^{o} (0^{o}) and 430^{o} (70^{o}) and correlate with initial access of 8OG into the active site (0^{o}) and final positioning as described in the 1R2Y crystal structure (70^{o}). Previous simulations describe in detail rotation around the glycosidic torsion in the active site of Fpg and conclude that a torsion of 60^{o} for 8OG is preferred, agreeing with the crystal structure and the simulations presented here.[119]

4.3.2 Structural evaluation and comparison of the major and minor groove PNEB and umbrella sampling paths

By using dynamic simulations to find the minimum energy paths of eversion through the major and minor groove, individual interactions between protein side chains and the 8OG base can be measured. The simulations can link points on the free energy surface to specific structure interactions, yielding an explanation of the eversion process. Here we compare specific interactions along each path, and their respective values throughout the eversion process represented by the PNEB pathway and the umbrella sampling pathway. This assures important contacts seen in the PNEB path are not lost in the umbrella sampling, which samples structures close to the initial PNEB structures used as starting points for windows. For the umbrella

sampling simulations through each groove a "trajectory slice" was taken from the intrahelical structure to the extrahelical structure, combining adjacent windows in the eversion dihedral.

4.3.2.1 8OG recognition in the major groove path

The major groove interactions (shown in Figure 4-6), reading from top to bottom, progress through the major groove from an initial eversion dihedral value of 20^{o} at the intrahelical state to a value of 260 ^{o} at the extrahelical state. The direction of eversion from intrahelical to extrahelical is from left to right. The top row of plots shows the initial interaction of Arg263, located on the β hairpin of the zinc finger binding motif, with 8OG. A favorable electrostatic interaction between the Arg263 amine and the O8 of the 8OG base (an atom which distinguishes damaged 8OG from undamaged G) forms as the Watson-Crick hydrogen bonds break (black). A cation-pi stacking interaction between the Arg263 and 8OG follows, as the 8OG everts over the Arg263 further towards the active site (red). This is illustrated in Figure 4-6a, which shows a snapshot from the umbrella sampling pathway where the distance between the Arg263 guanidinium group and the 8OG is at a minimum value of 3.2Å.

The second row of plots show the interaction at the *exo* site intermediate discussed above (an eversion dihedral value of 120 ^{o}). A specific hydrogen bonding interaction between the amide group of the Asn173 side chain and the O8 of the 8OG (green) forms, while the backbone carbonyl of Gly264 forms another specific hydrogen bond with the protonated N7 of G (blue). Thus, 8OG procession is promoted by specific contacts with the damaged atom, and G is discriminated against by an interaction which disappears in the case of the 8OG damage. The *exo* site is illustrated in Figure 4-5b, where a minimum distance of 2.5Å between heavy atoms is

reported for the interaction of Asn173 and 8OG, and the Arg263 hydrogen bonds to the phosphate group of the 3'G after 8OG has passed over it along the eversion pathway. These interactions are critical, as they are specific to the atoms which distinguish 8OG from an undamaged G, and thus can discriminate in favor of 8OG and against G, where these interactions would be replaced by electrostatic repulsion.

After stabilization and potential discrimination in the *exo* site, the 8OG proceeds to the active site, overcoming its largest free energy barrier along the major groove path. The third row of graphs and the illustration in Figure 4-6c show critical interactions which stabilize this transition state. Arg111intercalates into the area vacated by the everting 8OG and hydrogen bonds to the Watson-Crick face of the orphaned C (pink). A minimum distance of 2.7Å is shown between the Arg111 amine and the C. At the same time, a bridging interaction is formed between the Arg111 and the phosphate group 3' to the 8OG base (cyan). These interactions are hypothesized to stabilize the DNA helix during the final stages of eversion. The active site loop orients around the O6 atom of the 8OG base, forming hydrogen bonds to the amide backbone groups of several residues in the loop, including Thr223 (orange).

Interactions between the everting 8OG base and Fpg side chains Arg263, Gly264, and Asn173 had previously been isolated to and their influence the on the major groove flipping pathway evaluated by energy decomposition.[120] However, the data shown here is critical to evaluating the fidelity of the umbrella sampling to the original structure information contained in the PNEB pathway.

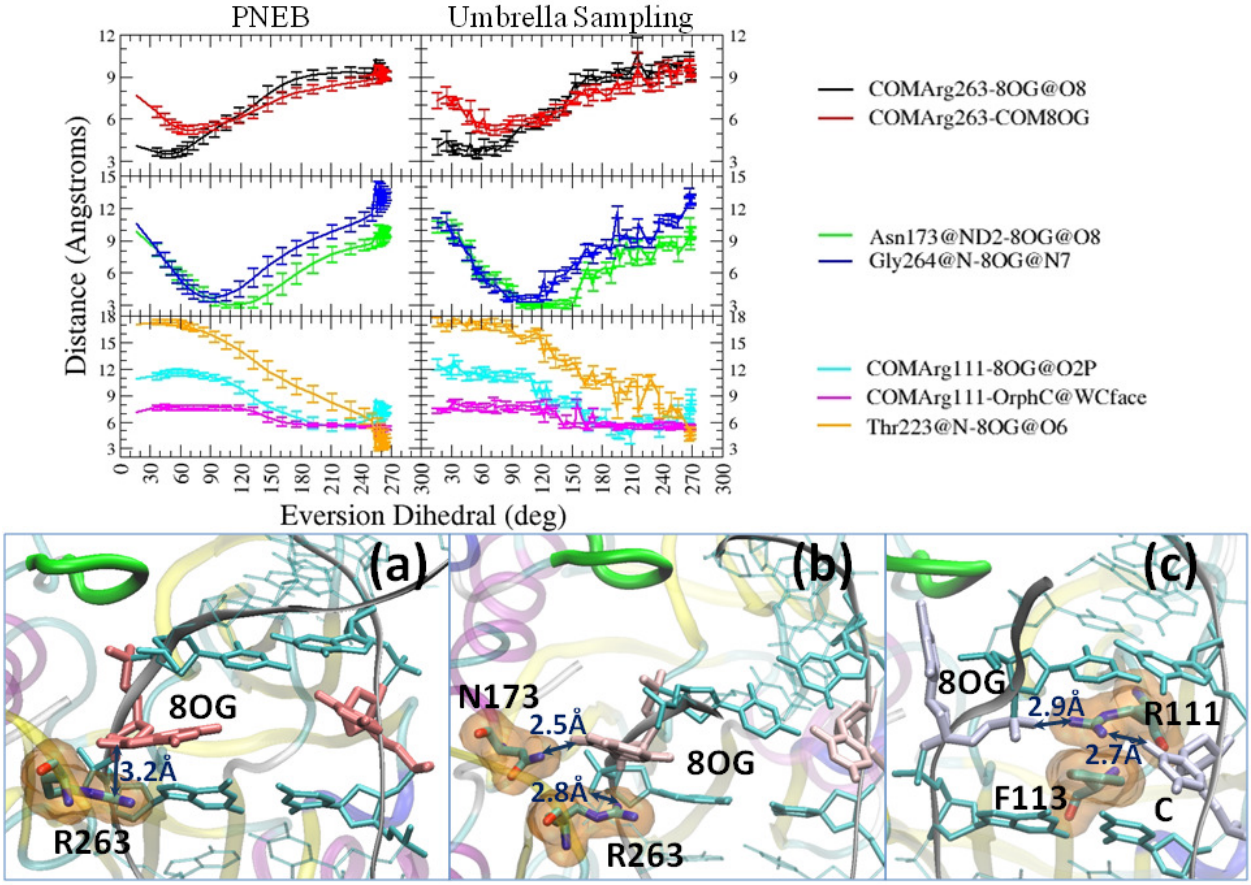

Figure 4-6 Top: Graphs of critical interactions in the base eversion pathway. Error bars are standard deviation in each bead (for PNEB, left column) or window (for umbrella sampling, right column). Bottom: Intermediates in the major groove eversion pathway. The 8OG:C base pair is colored by timestep of the base eversion trajectory from red (intrahelical) to blue (extrahelical). a) First minimum corresponding to eversion dihedral value of 60°. 8OG forms cation-pi stacking interactions with R263. b) Second minimum, termed *exo* site at an eversion dihedral value of 120°. 8OG hydrogen bonds to N173 side chain, and the R263 side chain hydrogen bonds to phosphate oxygen of 8OG. c) Transition state structure, corresponding to eversion dihedral value of 180°. R111 intercalates into helix, forming bridging hydrogen bonds to the orphaned cytosine and the phosphate group of the flipping 8OG base.

4.3.2.2 Non-specific contacts in the minor groove path

Interactions in the minor groove base eversion pathway are shown in Figure 4-7. The intrahelical endpoint has an eversion dihedral value of 20^{o}, and the extrahelical endpoint has a dihedral value of -110^{o}. Progression through the minor groove is read on the plots from right to left, with the earliest path interactions at the top. The top plots show the interaction between Arg263 and the O8 atom of 8OG, which is specific to the damaged base (black). As this favorable interaction forms, the O8 atom of the 8OG base shifts towards the oxygens in the phosphate backbone (red) and deoxyribose sugar (blue). This shift from longer optimum distances (refer to Chapter 2, Section 2.3.3) to distances closer to these oxygen atoms results in electrostatic repulsion, as the 8OG rotates around its glycosidic torsion to alleviate these unfavorable interactions. This is illustrated in Figure 4-7a, where the interactions with Arg263 is at a minimum distance of 2.6Å and the propeller twist of the 8OGC base pair is apparent. It is important to note that although isolated interactions can be favorable in energy, the overall steric clash resulting from rotating around the glycosydic bond while in an intrahelical position, and also in the presence of at least three intercalating amino acid side chains, results in a very unfavorable free energy, and it is unlikely that efficient processing of 8OG through the minor groove occurs.

If the 8OG progresses through the minor groove, interactions with Arg111 dominate the eversion pathway. During rotation around the glycosidic torsion, some stabilizing cation pi stacking interactions occur when 8OG is between Arg263 (green) and Arg111 (orange). After rotation from *anti* to *syn*, the O6 of 8OG is positioned to interact with Arg111 (purple). As illustrated in Figure 4-7b, hydrogen bonding interactions form between the guanidinium group of

Arg111 and the O6. Here a bifurcated hydrogen bonding pattern is shown; however, the basic side chain can form multiple hydrogen bonds.

The ability of Arg111 to adopt several hydrogen bonding orientations facilitates base eversion. In Figure 4-7c, the Arg111 remains hydrogen bonded to the O6 of 8OG, but forms interactions with the orphaned cytosine along its Watson-Crick hydrogen bonding face (cyan). After an eversion dihedral value of -70^{o}, the Arg111 interaction with the O6 atom of 8OG is replaced by an interaction with the amide backbone of Thr223 (pink), forming the contacts between the catalytic loop and the 8OG base in the active site. Computational analysis of the eversion pathway using targeted MD performed in Qi et al. indicate Arg111 plays the same role in 8OG eversion reported here.[22]

Other than the initial interaction with the O8 atom of 8OG, the interactions between Fpg side chains and the everting 8OG base are non-specific contacts with the O6 of 8OG, which is common to both the damaged and undamaged base. According to the calculated free energies of this process, it is unlikely that rotation around the glycosidic torsion occurs, allowing 8OG to evert via a minor groove pathway. Additionally, the lack of checkpoints in which the O8 or N7 of 8OG are recognized during the eversion process implies eversion through the minor groove is not specific enough to 8OG damage.

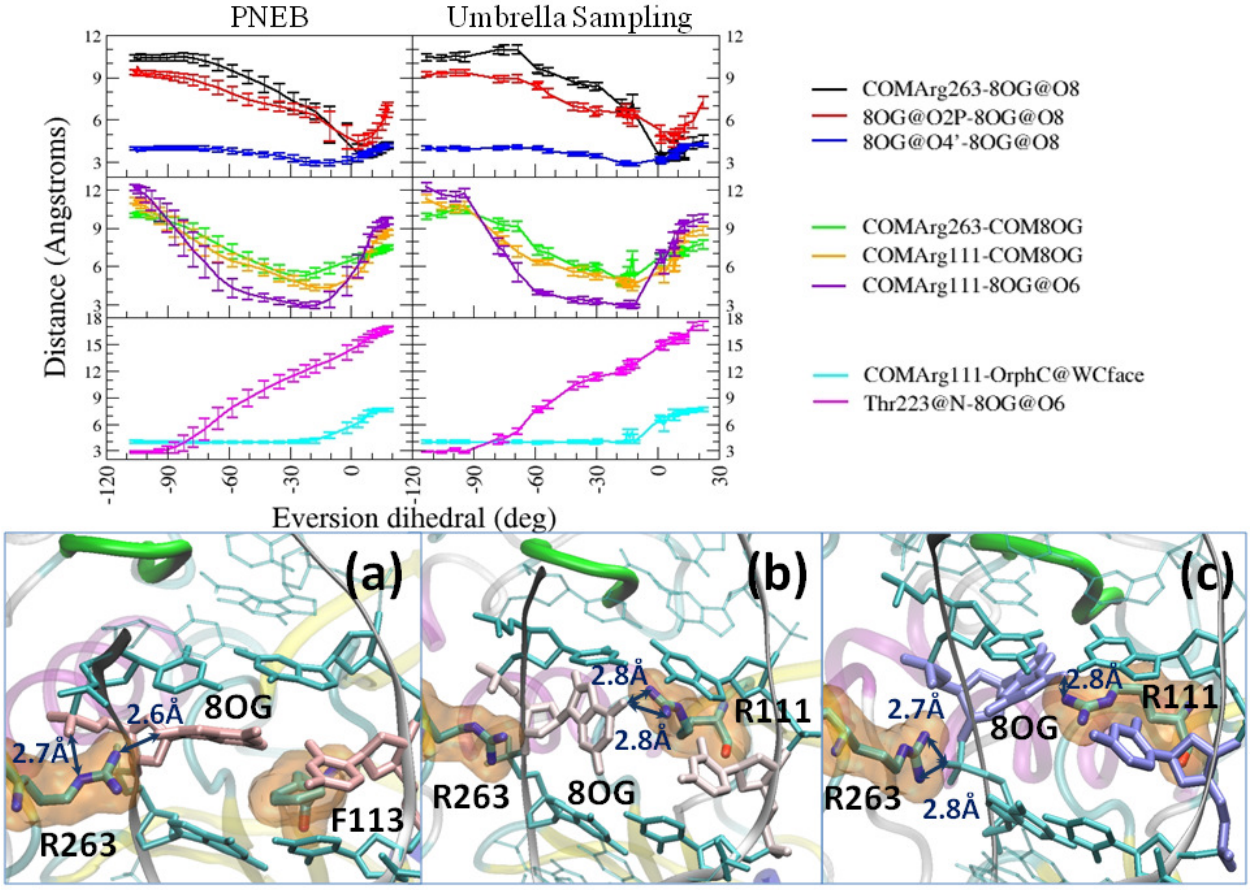

Figure 4-7 Top: Graphs of critical interactions in the base eversion pathway. Eversion occurs from right (intrahelical) to left (extrahelical). Error bars are standard deviation in each bead (for PNEB, left column) or window (for umbrella sampling, right column). Bottom: Intermediates in the minor groove eversion pathway. The 8OG:C base pair is colored by timestep of the base eversion trajectory from red to blue. a) First minimum at eversion dihedral value of 10^{o}. 8OG interacts with the R263 side chain, forming hydrogen bonds between the terminal amine and O8 and the HE and phosphate backbone. b) After rotating around its glycosidic torsion, R111 forms bifurcating hydrogen bonds between its amine groups and the O6 of the flipping base. c) 8OG continues on its eversion pathway, forming a hydrogen bond between its O6 atom and the R111 side chain. R263 hydrogen bonds to the 3' phosphate group of the flipping base.

4.4 Conclusion

In summary, the complete free energy profile of base eversion for Fpg with its 8OG substrate has been determined using PNEB and umbrella sampling. The minor groove PMF indicates a free energy barrier to flipping into the active site that is significantly lower for major groove eversion (corresponding to ~2-3 orders of magnitude faster eversion rate), indicating that this is the preferred path for lesion processing in Fpg. Interestingly, the human functional analog hOGG1 is also believed to involve base eversion via a major groove pathway, reinforcing the functional conservation between the prokaryotic and eukaryotic enzymes.[76; 121] Crystal structures of hOGG1 indicate a similar *exo* site where undamaged guanine is trapped to avoid processing.[76] The major groove intermediate may provide a similar readout mechanism by Fpg, allowing 8OG to continue along the path to the active site, while discriminating against G. The data suggest that Fpg not only shares functional similarity to hOGG1, but convergent evolution is also present in dynamic aspects of base recognition by DNA glycosylases.

5 MUTANT STUDIES TO EXAMINE FPG SUBSTRATE RECOGNITION

5.1 Introduction

Experimental and computational descriptions of the base eversion pathway of 8OG point to specific interactions between protein and DNA which facilitate this biological process. Mutation of critical residues and observing their effect on the pathway helps to resolve their specific functions as well as the overall chain of events in the eversion pathway. The first crystal structure of Fpg, in which the Schiff base intermediate is trapped, shows three protein residues intercalate into the void left after 8OG eversion.[18] These residues, Arg108/111, Met73/76, Phe110/113, form contacts to both the orphaned cytosine and the flipped base (numbering scheme for *E. coli* Fpg and *B. st.* Fpg, respectively). The phenylalanine is proposed to disrupt pi stacking between the orphaned cytosine and its neighbor, stabilizing the flipped conformation. When co-crystallized with DNA containing an abasic site, these three residues are still important in maintaining the integrity of the helix in its bent conformation by gap filling.[122] The goal of this study is to isolate the function of two intercalating residues in the major groove eversion pathway, specifically the phenylalanine and arginine residues from this intercalating triad.

The role of a wedge residue in general, and specifically Phe113 in Fpg, has been hypothesized to destabilize intrahelical Watson-Crick base pairs.[19; 21] An aromatic residue such as Phe113 can interrupt the pi stacking interactions between base pairs, thus interfering with an important stabilizing property of canonical B-form DNA structure. Intercalation of a bulky side chain into the helix induces a kink in the DNA, causing local deformation of the DNA but

retaining B-form like characteristics 3' and 5' to the kink. This can disrupt the canonical base pair orientation, which, along with facilitating the kinked and bent structure of the DNA, is hypothesized to lower the energy barrier to eversion, which is critical in systems which perform base flipping as part of their biochemical mechanism. The crystal structures of Fpg bound to an intrahelical G:C base pair and an intrahelical A:T base pair show binding results in unavoidable intercalation of the Phe113 wedge.[21] This supports the hypothesis here that the wedge plays an active role in the eversion pathway by transferring binding energy to destabilize the base pair. Stop-flow fluorescence experiments point to several discrete steps along the eversion pathway, and it is reasonable to think the wedge plays an active role in destabilizing the base pair to facilitate the initial steps of the eversion pathway.

Intercalating wedge residues are pervasive in DNA repair machinery.[123] The human Fpg analog hOGG1 has a tyrosine wedge proposed to act in the same capacity as the Phe wedge in Fpg.[124] The MutY glycosylase also has a tyrosine wedge, which intercalates into the helix 5' to the 8OG:A base pair.[125] UDG contains a tyrosine residue which intercalates into the helix, and can stack with the adenine of the U:A base pair, or rotate to form cation-pi interactions with Arg residues of the protein. Both of these functions in UDG serve to widen the minor groove for U recognition.[126] This early interaction along the eversion pathway in Fpg can be tested by computationally mutating the Phe residue to an alanine, thus removing the bulky side chain and pi stacking characteristics of the original residue. The focus of this work is to test whether the F113A mutant will stabilize the orientation of the 8OG:C base pair interrogated by Fpg by allowing the base pair to remain planar and thus in its optimal orientation for forming Watson-Crick hydrogen bonds and pi stacking interactions with the surrounding DNA. This will result in a higher free energy barrier to initial breaking of the hydrogen bonds and subsequent eversion.

Intercalating residues play an additional role in stabilizing the DNA helix during later intermediate steps in the eversion pathway. In the crystal structure of Fpg bound to an abasic site, an arginine residue hydrogen bonds with the Watson-Crick face of the orphaned cytosine, discriminating the cytosine opposite 8OG damage.[23] As part of the "GO" repair cycle, Fpg is responsible for recognizing 8OG damage paired opposite a C, everting and subsequently excising the 8OG base. Perhaps as important, however, is the rejection of the 8OG:A base pair, where removing the 8OG damage (the primary function of Fpg) would facilitate the loss of the original DNA sequence. Previous biochemical studies had determined a purine opposite 8OG damage is the worst substrate for Fpg function.[16]

Site directed mutagenesis was performed by our experimental collaborators on *E. coli* Fpg, mutating Arg108 to Ala, and bound to DNA containing 8OG opposite each of the four natural bases, in order to probe this residue's effect on complementary base recognition.[127] The mutation led to at least a 15-fold decrease in Fpg activity on all substrates, including a more drastic decrease in K_M for Fpg bound to 8OG:T and 8OG:G, thermodynamically unstable base pairs.[16; 128] Decreases of an order of magnitude in activity for both 8OG:C and 8OG:A were due to an effect on k_{cat}. When acting on a substrate which has no opposite base preference, dihydrouracil (DHU), R108A catalytic activity decreased by about 2 fold from the WT. Even when it is unimportant to specify the opposite base, deleting the Arg108 side chain has some effect on the catalytic activity.[127]

As discussed in Chapter 4, Section 4.3.2., *B. st.* Arg111 forms several contacts in both the major and minor groove paths. However, its common function in both the major and minor groove paths is to hydrogen bond to the Watson-Crick binding face of the orphaned cytosine after intercalation into the void left by the flipping 8OG base. In the major groove path, this

interaction plays a dynamic role during base eversion by forming additional contacts to the phosphate group of the flipping 8OG base, stabilizing the helix at an important transition state. Although it is not a local minimum on the free energy surface, we hypothesize this bridging hydrogen bond network is critical to lowering the transition state energy of the complex by stabilizing the DNA helix.

These late interactions in the eversion pathway can be tested by creating two mutant systems. The first system studies how recognition of the C base changes by mutating away the Arg111 residue. Arg111 invades the helix during flipping, forming hydrogen bonds to the Watson-Crick face of the orphaned C. In WT simulations, this residue also formed bridging hydrogen bonds to the phosphate group of the flipping base, thus stabilizing a non-canonical eversion intermediate. When the side chain is mutated away, as in the R111A mutant studied here, no intermediate can form these stabilizing interactions and a high barrier to eversion results. The second system acts as a negative control, where we have mutated the interrogated base from 8OG:C bound in a Watson-Crick orientation to 8OG:A, bound in a Hoogsteen orientation, as shown in Figure 5-1a and b. The Arg111 does not hydrogen bond to the orphaned A in the same manner as it did the orphaned C, because steric constraints inhibit access to the cavity between the adenine and the phosphate group of the flipping base. This system also serves as an example of Fpg binding to an 8OG opposite its least favorable substrate, where we can determine the effects *syn* 8OG and orphaned A have on the base eversion pathway.

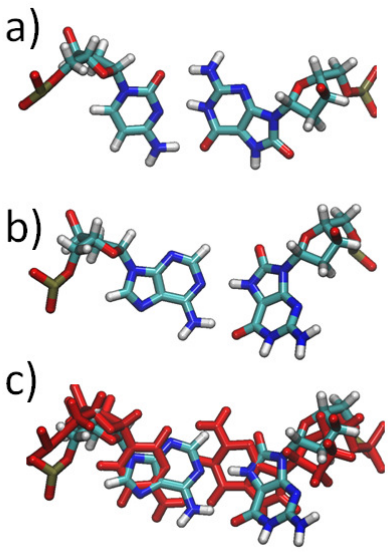

Figure 5-1 a) C:8OG base pair, with 8OG in *anti* conformation. b) A:8OG base pair with 8OG in *syn* conformation. c) Overlap of C:8OG base pair (red) and A:8OG base pair (colored by atom name).

5.2 Methods

Amber11[129] was used for all simulations. The ff99SB[36] force field for proteins and the ff99 force field for nucleic acids with the Barcelona parmbsc0 corrections[42] were used for simulation of these protein-DNA complexes. The complexes were solvated using TIP3P water in a truncated octahedral box with a 12Å buffer.[100] The Miller et al.[88] parameters were used for the 8OG residue and the Stote nonbonded model was used for the zinc in the zinc finger DNA binding motif.[111] A Langevin thermostat was used to regulate temperature to 330K for PNEB and umbrella sampling simulations with a collision frequency of 2 ps^{-1} unless otherwise specified.[101] SHAKE was used to constrain bonds to hydrogen.[92]

Except where noted, intermediate structures were minimized using the following procedure. Each mutated structure was simulated in the NVT ensemble for 100ps using a 1fs timestep, with 100 kcal mol^{-1} A^{-2} positional restraints on all heavy atoms of the solute, allowing the solvent and hydrogen atoms to equilibrate. A collision frequency of 2 ps^{-1} was used.

5.2.1 F113A mutation

The effect of removing the wedge on the global minimum free energy pathway was calculated by mutating wild type structures from the combined trajectory of each NEB bead's final path equilibration at 330K. Since the intrahelical region and initial steps in the eversion pathway are most affected by the intercalating Phe, a reaction coordinate more specific to the hydrogen bonding face of 8OG was used. From the subset of NEB structures used as initial structures for umbrella sampling windows, a drop line from the intrahelical structure (20^o, 310^o) to the *exo* site structure (120^o, 340^o) was assigned. Structures with values closest to this line were chosen as starting structures for 1D umbrella sampling for the WT system. These 40 starting structures were mutated using the leap module of Amber11 by removing all atoms of the Phe113 side chain except for the CB, and building the alanine methyl group based on the CB coordinates. Solvent box information was retained.

The earliest part of the eversion pathway through the major groove involves breaking the Watson-Crick hydrogen bonds between the 8OG:C base pair as the 8OG base begins to evert, until it is eventually stabilized by *exo* site contacts. This is the stage of eversion where the Phe wedge is hypothesized to play an active role in promoting eversion of the destabilized 8OG:C base pair. Since this section (eversion dihedral of 20^o to 120^o) of the 2D PMF shows little

variation in the glycosidic torsion, the problem can be described by a 1D reaction coordinate, which better describes the local change in hydrogen bonding between the flipping 8OG and its intrahelical position. This coordinate is summarized in Figure 5-2, and is the distance between the centers of mass of the 3' and 5' base pairs and the heavy atoms of the Watson-Crick hydrogen bonding face of the everting 8OG. It was adopted based on unpublished results from Haoquan Li for the human hOGG1 base eversion pathway, and is referred to as the "eversion distance."

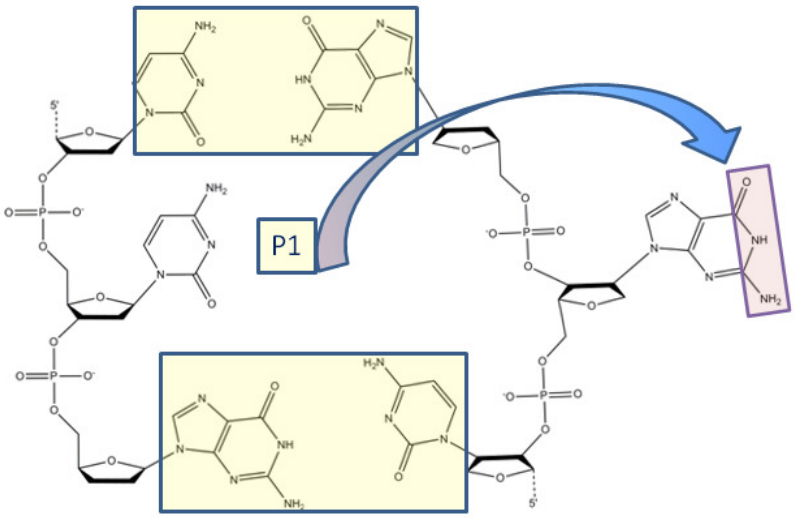

Figure 5-2 Schematic representation of the eversion distance reaction coordinate based on the distance between the heavy atoms of the Watson-Crick hydrogen bonding face of the everting 8OG and the center of mass of the flanking base pairs.

Umbrella sampling was performed using 40 windows spaced by 0.3Å intervals from 0.3 Å to 12.0 Å, for both the WT and F113A systems. Each window was simulated in the NVT ensemble for 500ps with a 2fs timestep, with 10 kcal mol^{-1} A^{-2} distance restraints along the

reaction coordinate described above. Convergence of the PMF calculation was tested using error bars based on the first and second 250ps of the simulation.

5.2.2 R111A mutation

New optimization using PNEB was performed to find the minimum potential energy path of the R111A mutant's transition along the base eversion path. This was not necessary for the F113A mutant since that mutation affects the thermodynamics of the intrahelical region only, and not the subsequent intermediates along the base eversion path. Here, the R111A mutant affects the middle of the base eversion pathway. The starting structures generated from NEB and used for the WT major groove umbrella sampling had Arg111 mutated to alanine by removing all atoms of the Arg111 side chain after the CB, and building in the methyl group for alanine using leap. These structures were equilibrated as described in Section 5.2. 2D umbrella sampling was performed using a procedure described for the WT simulation (Section 4.2.3). In order to ensure well-converged PMFs it was essential to reoptimize the path with PNEB, requiring well minimized R111A endpoints.

The intrahelical and extrahelical WT endpoints were both mutated (R111A) using Leap and retaining solvent box information. Coordinates were restrained with 100 kcal mol^{-1} A^{-2} positional restraints on all heavy atoms of solute and minimized using 10,000 steps of steepest descent minimization. They were heated linearly over 100ps to a target temperature of 330K, using positional restraints on all atoms of the protein and DNA, allowing only solvent to move. A timestep of 1 fs was used. Temperature was regulated with a Berendsen thermostat at constant volume.[91] SHAKE was used to constrain bonds to hydrogen.[92]

Two NPT cycles of MD with restraints of 100 kcal mol^{-1} A^{-2} and 10 kcal mol^{-1} A^{-2} on side chain and base atoms were performed for 100ps and 250ps, respectively, increasing the pressure relaxation time from 0.1ps to 0.5ps. Two more 100ps NPT cycles of MD with decreasing restraints from 10 kcal mol^{-1} A^{-2} to 1 kcal mol^{-1} A^{-2} on the protein and DNA backbone were then performed. Figure 5-3 shows the extrahelical endpoint structure of both the wild type and the R111A mutant.

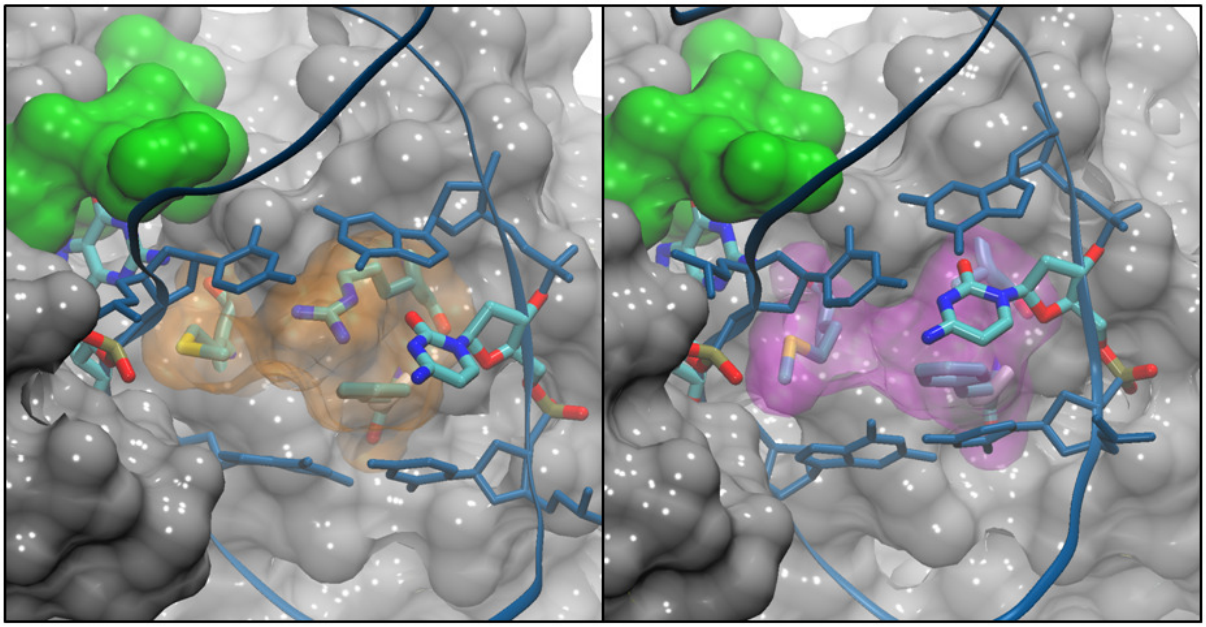

Figure 5-3 WT (left) and R111A (right) everted endpoints from NEB simulations. Residues Met76, Phe113 and Arg111 are shown colored by atom name and their surfaces are colored in orange in the WT (left) and pink in the R111A mutant system. The catalytic loop is in green, and the 8OG:C base pair is colored by atom name.

Intermediate structures for the PNEB pathway were selected from the R111A umbrella sampling pathway. Final structures from windows every 10^{o} in the eversion dihedral from 30^{o} to 270^{o} along a glycosidic torsion of -30^{o}, and windows every 10^{o} in the glycosidic torsion from -

30^o to 20^o were chosen as intermediates for the PNEB pathway. 30 frames from the 2D umbrella sampling windows were used to seed the PNEB major groove pathway, in addition to the intrahelical and everted endpoint structures, totaling 32 structures along the starting PNEB path.

For PNEB simulations, the NVT ensemble was used. The same procedure as the original WT PNEB pathway was followed. NEB forces were calculated for all atoms of the solute and not the solvent. A simulated annealing path optimization procedure was adapted from Mathews and Case 2006.[65] Initial spring forces of 10 kcal mol^{-1} $Å^{-2}$ and an initial collision frequency of 100 ps^{-1} were used. The first stage of path optimization for 80ps at 330K was followed by equilibration of all structures at this temperature for 500ps. Heating of the path to 386K took place over 60ps, with an increased spring constant of 25 kcal mol^{-1} $Å^{-2}$ and decreased collision frequency of 75 ps^{-1}. Equilibration at 386K for 220ps was followed by annealing to 330K over 100ps. Final equilibration at 330K to generate populations for each structure was performed over 500ps.

5.2.3 8OG:A mutation

The WT endpoints were mutated to 8OG:A endpoints using Leap. For the intrahelical structure, this consisted of removing the atoms of the 8OG and C nucleobases in the PDB, leaving the sugar and backbone atom positions unchanged. The positions of the C8 and C4 atoms were switched manually in the PDB to build the 8OG residue in the *syn* position which it adopts upon pairing with an adenine.[85] The N1 of the orphaned C was changed into an N9 manually to build the A residue. The same procedure was followed for the everted endpoint for the A. The 8OG was left unchanged in this endpoint since it is bound in the active site in a known

orientation. The everted endpoint followed the same minimization and equilibration protocol as the R111A mutants above.

There were significant problems with the initial minimization of the intrahelical structure. After mutating the 8OG:C base pair to an 8OG:A base pair, initial positional restraints on all heavy atoms of the solute were decreased. After switching to positional restraints on all backbone heavy atoms to encourage minimization of the DNA bases and protein side chains, the system violated these restraints due to the unfavorable energy of the 8OG:A base pair. In Figure 5-4 below, the hydrogen bond distances jump to values of 15Å before the simulation crashes. The difficulties in minimizing an intrahelical 8OG:A base pair reflects the biology of the enzyme; the 8OG:A base pair is the least favorable substrate for Fpg, with a dissociation constant about two orders of magnitude higher than the wild type.[127] Advanced minimization techniques were necessary to generate a stable structure with which to run NEB.

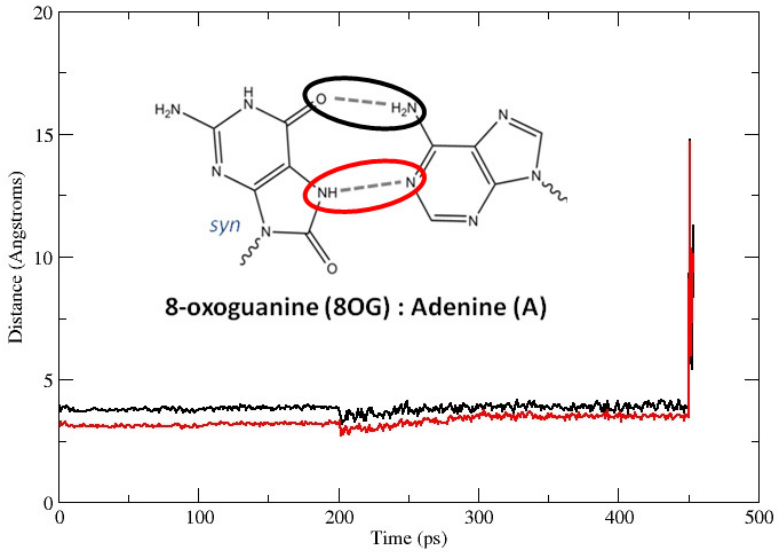

Figure 5-4 Hydrogen bond heavy atom distances during 8OG:A intrahelical structure minimization.

A conjugate gradient minimization method using a limited memory quasi-Newton preconditioner to help speed convergence, implemented in the xmin module in Amber11, was used in order to successfully minimize the intrahelical WT-Fpg with an 8OG:A base pair.[130] The minimization was performed using the HCT (gb=1) implicit solvent model,[98] used because it is able to reproducing DNA structure comparable to explicit water simulations.[131] The results, summarized in Table 5-1, show that initial minimization produces lower energies and a better structure than that initially built, represented by optimized Hoogsteen hydrogen bond distances.

Table 5-1 Xmin Results for Intrahelical WT-8OGA Endpoint

Test	**Iterations**	**Energy (kcal/mol)**	**Gradient**	**CPU time**	**H-bond dist (heavy atoms, Å)**
Initial Structure	0	-14729.79	19.031	---	3.15, 3.78
Test1	10	-19081.23	1.098	13 min	2.9, 3.35
Test2	100	-19504.52	0.025	3.5 hrs	2.9,2.8
Previous Min Scheme	0 (Single Point Energy)	-14549.98	18.661	19 sec	---

The xmin minimized intrahelical structure was solvated with TIP3P water using a 12Å buffer. Minimization and equilibration were then performed using the same procedure as the R111A endpoints. PNEB was performed using 16 copies of each endpoint structure as starting beads along the path, totaling 32 images. Optimization was performed using the same method applied to the R111A PNEB simulation and the WT PNEB simulation.

5.3 Results and Discussion

5.3.1 F113A mutant explains the role of wedge residue in destabilizing the intrahelical 8OG:C base pair

The PMFs for eversion distance of the WT and F113A systems are shown in Figure 5-5. Both the wedge present (WT, black) and wedge absent (F113A, blue) graphs are zeroed in the *exo* site at 11Å, since the Phe wedge is proposed to affect the intrahelical region and have no effect on the *exo* site stability. The difference plot shown in Figure 5-5c represents the F113A free energy subtracted from the WT free energy, and allows us to quantify there is a 5 kcal/mol stabilization in free energy when the wedge is absent. This is consistent with the hypothesis that F113 destabilizes the 8OG:C base pair.

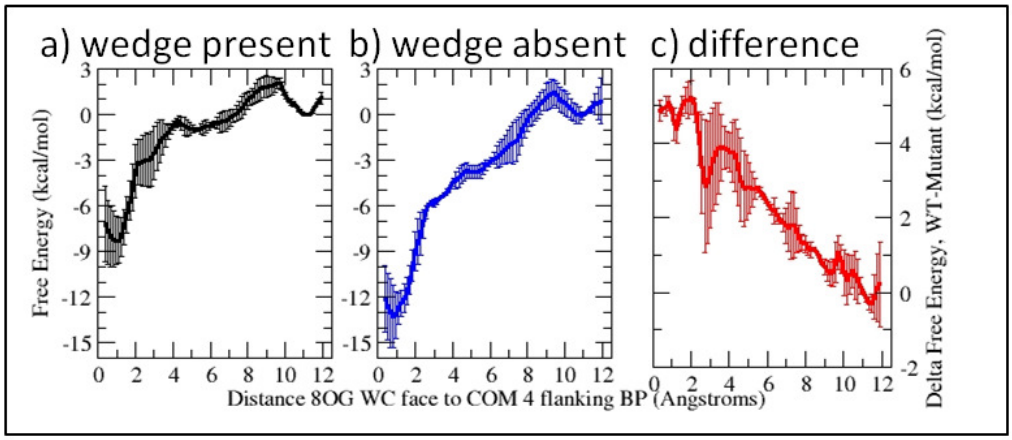

Figure 5-5 1D PMFs of WT, F113A, and the difference between the two (WT-F113A). PMFs are zeroed in the exo site at 11Å. There is a 5 kcal/mol stabilization in free energy when the wedge is absent. Error bars are calculated from the first 250ps and second 250ps of the simulation.

The intrahelical structures of the WT and F113A systems from umbrella sampling show structural differences at the minimum of each PMF. The most obvious difference is that once the wedge residue is lost, the base pair buckle is able to relax to a more planar geometry, as shown in Figure 5-6. The structures are aligned based on the positions of the 3' and 5' base pairs. The inclination of the WT system 8OG:C base pair is tilted towards the major groove in addition to a slight displacement towards that same groove. The intrahelical geometry of the F113A mutant has been optimized relative to the WT, and stacking with at least the 3' base pair is improved.

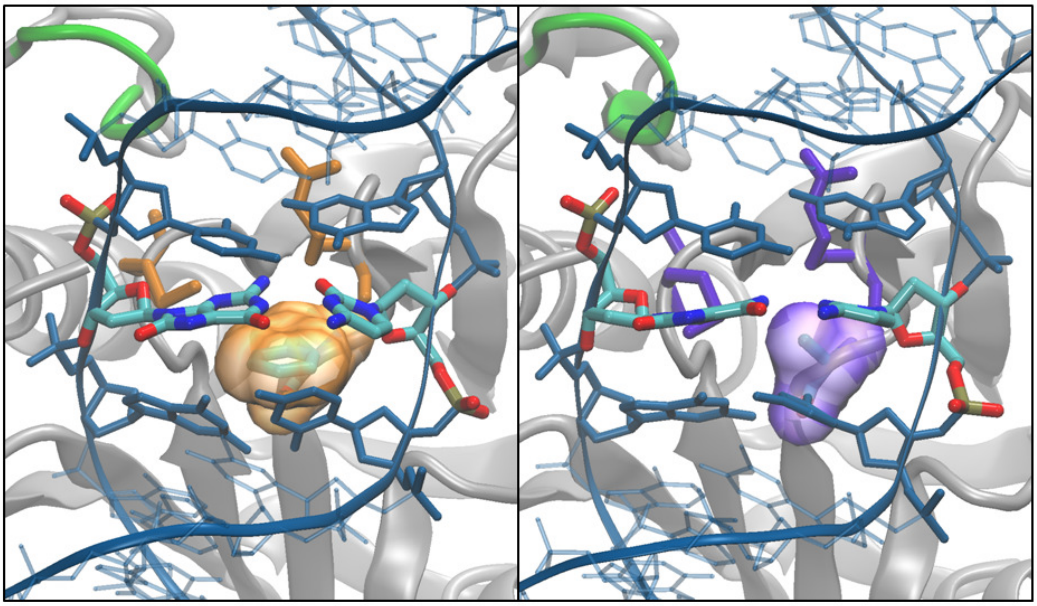

Figure 5-6 WT (left) and F113A (right) intrahelical structures after umbrella sampling, at their free energy minima (reaction coordinate distance = 1.2Å). Residue Phe113 is shown colored by atom name with surface in orange (WT, left) and residue Ala113 is shown colored by atom name with surface in purple (F113A, right). Residues Met76 and Arg111 are shown colored in orange in the WT (left) and purple in the F113A mutant system (right). The catalytic loop is in green, and the 8OG:C base pair is colored by atom name.

To quantify the difference in hydrogen bonding in the intrahelical region, the heavy atom distance of the Watson-Crick hydrogen bonding pairs was calculated. This compares the hydrogen bonding complementarity of the 8OG:C base pair when the wedge is present vs. when the wedge is absent. As the base begins to evert, these interactions must get weaker. Figure 5-7 shows the results for the WT and F113A systems. The WT system maintains the intrahelical hydrogen bonds up to an eversion distance of about 2Å. Additionally, the WT system is discontinuous where the base breaks its Watson Crick hydrogen bonds, indicating this is a rapid and not easily reversible event. The opposite can be seen in the F113A mutant, where an improved intrahelical geometry facilitates stretching of the hydrogen bonds past values sampled in the WT, as well as allows breaking and reforming of Watson Crick hydrogen bonds. The absence of the wedge allows for a more favorable backwards rate than the WT, which favors a forward rate and proceeding along the base eversion pathway.

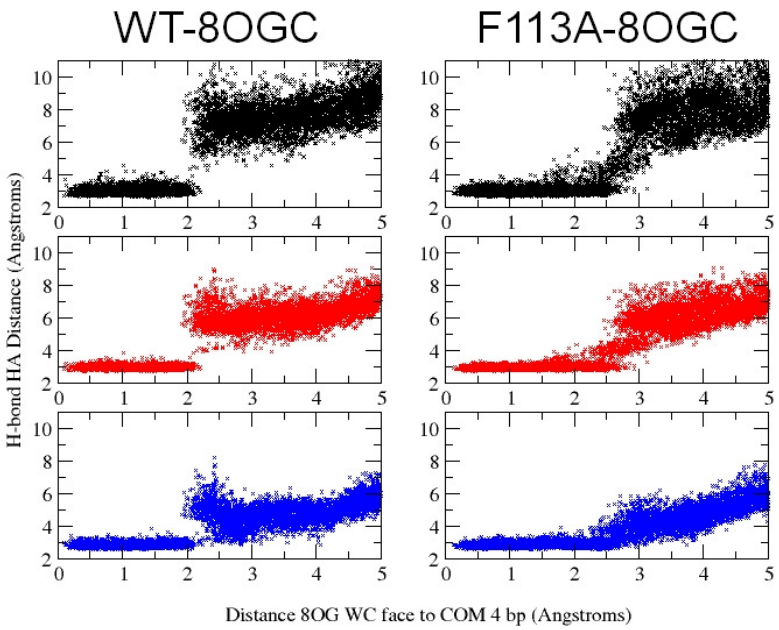

Figure 5-7 Watson-Crick hydrogen bond lengths for the 8OG:C base pair in the WT (left column) and F113A (right column) systems. The eversion distance is truncated at 5Å. The heavy atom distances measured are C@N4 to 8OG@O6 (top, black); C@N3 to 8OG@N1 (middle, red); C@O2 to 8OG@N2 (bottom, blue).

A structure comparison of the two systems is shown below in Figure 5-8. It is apparent that two different structures are accommodated at an eversion distance of 2.9Å in the reaction coordinate used for umbrella sampling. In the WT system, the hydrogen bonds are all broken, with distances between their heavy atom hydrogen bonding partners of 4.7 Å, 6.3 Å, and 8.3 Å. The F113A maintains a hydrogen bonded base pair orientation, with values of 2.7 Å, 2.9 Å, and 3.1 Å. The presence of the wedge in the WT promotes buckling of the 8OG:C base pair and stretching of the hydrogen bonds, destabilizing the 8OG:C pair and allowing the system to overcome the free energy barrier of 5kcal/mol.

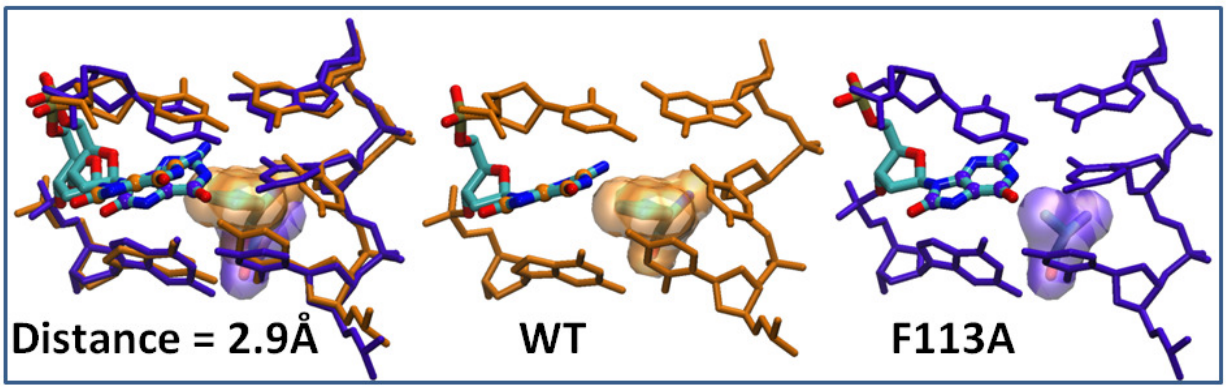

Figure 5-8 Structures from WT (orange) and F113A (purple) systems with a reaction coordinate distance of 2.9Å. In each, 8OG is colored by atom name. Left: Structures overlapped on all atoms of the three base pairs shown. Center: The Phe wedge of the WT system is colored by atom name and its surface shown in orange. Right: The Ala mutant in F113A is colored by atom name and its surface is shown in purple.

Unpublished experimental results from collaborators, in which a F110A point mutation in *Eco* Fpg was generated, resulted in a loss of glycosylase activity. Lyase activity on AP sites maintained activity in the mutant, indicating the Phe wedge does not affect reduction of AP sites, but rather plays an active role in 8OG base eversion, since without destabilization of the intrahelical 8OG:C base pair eversion and excision did not occur.[132] This agrees with the computational evidence presented here, outlining the active role in base eversion this Phe wedge plays.

5.3.2 Destabilization of a critical eversion intermediate

In Chapter 4, the role of Arg111 in the major groove pathway was discussed in Section 4.3.2.1. Specifically, it was noted that the conformation of Arg111 was such that it forms bridging hydrogen bonds between the orphaned cytosine and the phosphate group of the everting 8OG base at an eversion dihedral value of 180^{o}, corresponding to the major groove barrier region of the PMF. Here, we propose that this intermediate plays the dual role of discriminating against non-substrate 8OG:A base pairs and stabilizing the WT base eversion pathway.

We can compare properties of the DNA bound by Fpg to understand the deviation from WT for the R111A system. These properties should change if Arg111 does indeed affect the DNA conformations outside of passive recognition of the orphaned base, allowing us to understand how presence of a C affects 8OG recognition. Figure 5-9 shows the groove widths comparing the wild type and two mutant systems. The major groove width is similar throughout the eversion process, but the minor groove width fluctuates in the mutant systems. For R111A, this is a result of removing the stabilizing contact which Arg111 makes at the transition state of 180^{o} in the eversion dihedral between the phosphate group of the flipping base and the Watson-Crick hydrogen bonding face of the orphaned cytosine. No side chain exists to stabilize these two areas of the DNA, so the orphan C has little to no bad electrostatic or van der Waals contacts to worry about and it migrates towards the minor groove. In the 8OG:A system, the Arg111 remains outside the helix, unable to intercalate due to a steric clash between the Arg111 side chain and the bulky A base. In this system the phosphate group of the everting 8OG is not in an optimal orientation to form hydrogen bonds with the Arg111 side chain. This results in added

strain to the helix, increasing the transition state energies instead of lowering them as in the WT, where the pathway is energetically favorable.

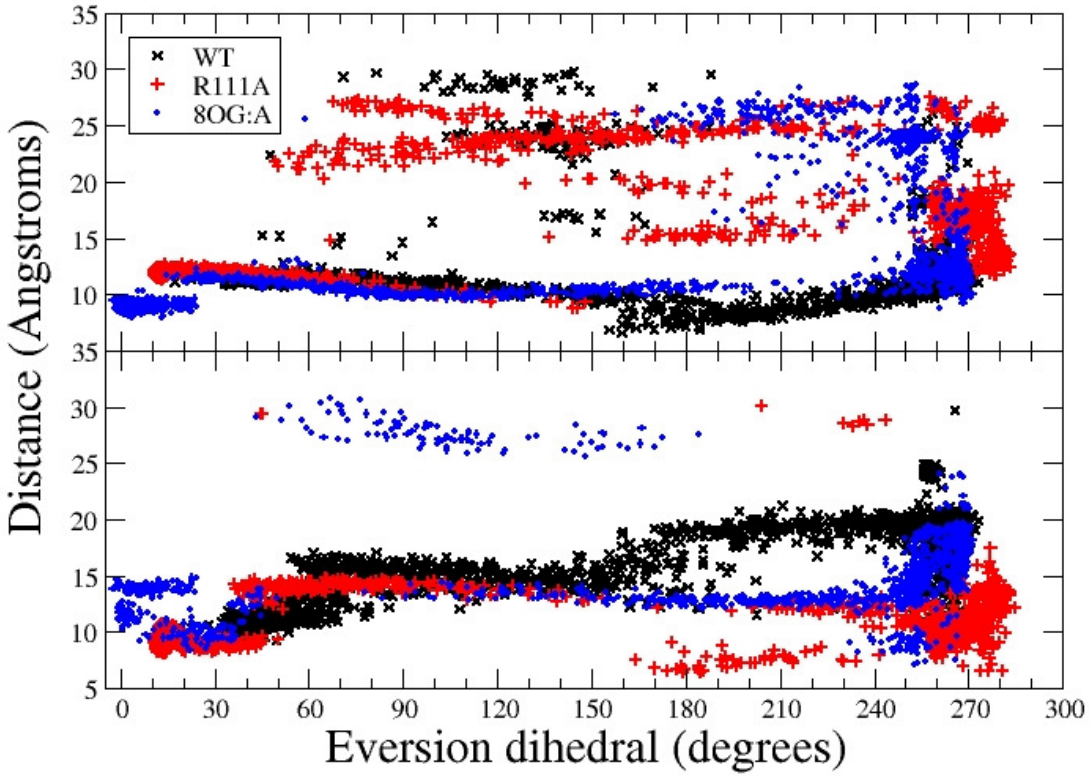

Figure 5-9 Top: Minor groove width vs. eversion dihedral. Bottom: Major groove width vs. eversion dihedral. In both graphs, WT is designated by a black x, R111A is designated by a red plus, and 8OG:A is designated by a blue circle.

Structures of each system at an eversion dihedral value of 195^{o}, which is the location of the WT transition state, are shown in Figure 5-10. These images show the orientation of residue 111 relative to the orphaned base in the WT, R11A, and 8OG:A systems. The WT system (Figure 5-10a) shows the Arg111 side chain adopts an orientation which can reasonably form bridging hydrogen bonds between the phosphate group of 8OG backbone and the orphaned C. In

the R111A system (Figure 5-10b), where Arg has been mutated to Ala, the orphaned C and phosphate group of 8OG are shifted further towards the minor groove since no intercalating residue is present to stabilize this orientation. When the orphaned base is an A (Figure 5-10c), there is a steric clash between the R111 side chain and the A, and the Arg111 does not intercalate into the helix. Visualization of these transition state structures show a structural check, performed for the correct substrate, results in a more favorable energy at the transition state. In the incorrect substrate this non-optimal structure has a reduced energy (8OG:A), and when the error-checking residue is eliminated (R111A), there is no interaction that can stabilize this transition state.

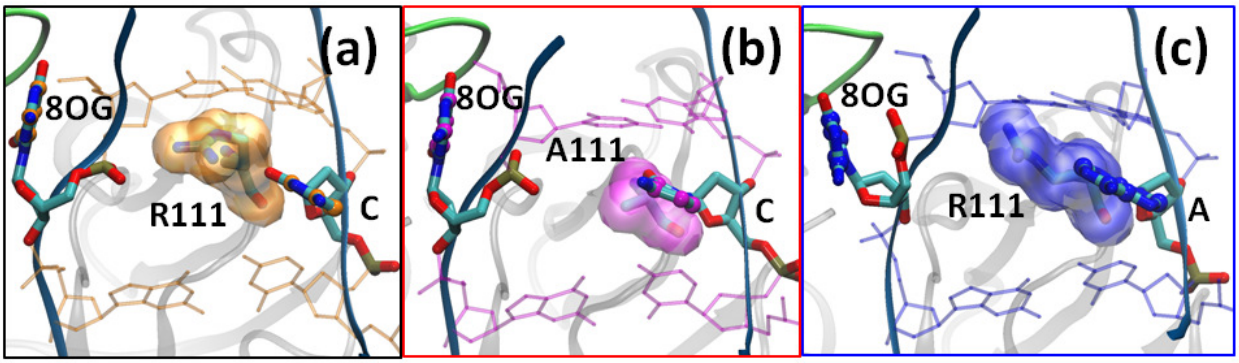

Figure 5-10 Structures from NEB simulations of a) WT, b) R111A, and c) 8OG:A systems. Structures have an eversion dihedral value of 195^{o} and are overlapped on their flanking base pairs. The catalytic loop is shown in green and the DNA backbone is shown as blue ribbons. The 8OG and its base pairing partner are colored by atom name. Residue 111 is highlighted using a surface representation colored orange, pink, and blue for the WT, R111A, and 8OG:A systems respectively.

Energy decomposition was performed quantify the difference in transition state structures seen in the PNEB simulations and shown above in Figure 5-10. To analyze the interaction energies between residue 111, the orphaned base and the backbone of 8OG we compare the difference in energy relative to the intrahelical endpoint. Figure 5-11 illustrates this comparison, where the nonbonded interaction energy can be decomposed into its electrostatic and van der Waals components. For the WT system (black), there is a clear stabilization in the interaction between Arg111 and the 8OG and orphaned C at an eversion dihedral between 180^{o} and 240^{o}, corresponding to the barrier region seen in the major groove PMF. This stabilization is dominated by the electrostatic interaction, but some additional favorable vdW energy contributes to the overall stability. As expected, this interaction disappears when the R side chain is mutated away to an A (red). These results correlate well with reported specificity constants for *E. coli* Fpg, where k_{sp} for WT-8OG:C > R108A-8OG:C.[127] Interestingly, the 8OGA system (blue) does show some stabilizing interaction between the R111, 8OG backbone, and orphaned A, however not to the extent seen in the WT system. This does not match the experimental trend in specificity constants for the three systems, since WT-8OG:C > R108A-8OG:C >> WT-8OG:A.[127] This indicates that the change in R111 interaction is not the determining step discriminating against the 8OG:A base pair; although the interaction is weakened, an earlier intermediate must prevent processing of 8OG:A pairs.

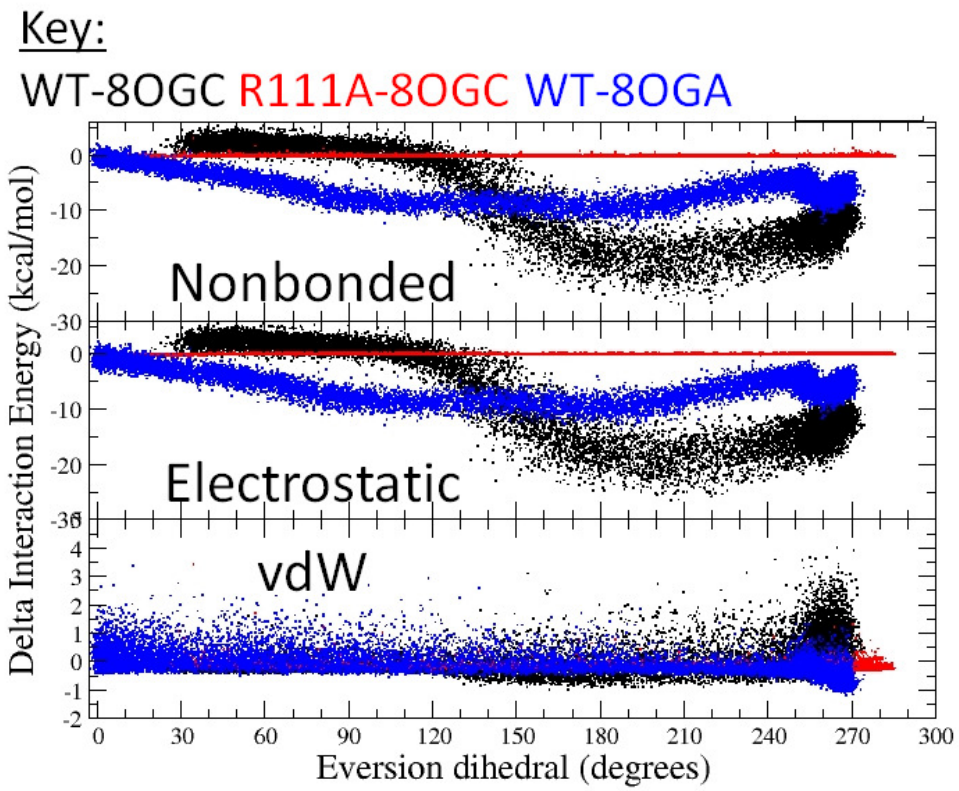

Figure 5-11 Interaction energy between the backbone of 8OG with the orphaned base (receptor) and residue 111 sidechain (ligand) and zeroed in the intrahelical site vs. eversion dihedral. The top graph shows the combined nonbonded interaction energy. The second graph shows the electrostatic component and the bottom graph shows the van der Waals component of the energies.

5.3.3 8OG:A base pair as a negative control for base eversion

Mutating the substrate base pair from 8OG:C to an 8OG:A bound in a Hoogsteen orientation allows us to investigate the mechanism of the Fpg base eversion pathway in the context of the least favorable substrate. The 8OG:A mispair results from insertion of an A opposite 8OG damage during replication. As part of the GO cycle, the 8OG:A base pair becomes a substrate for MutY, whose job is removing the misincorporated A to preserve the original DNA sequence. The 8OG:A base pair is the least preferred substrate for Fpg because removing the 8OG in this case would promote the mutation of the original G:C to a T:A base pair. This system can serve as a negative control for base eversion, since it is not the biological substrate and Fpg has the least enzymatic activity for the 8OG:A base pair.[16]

The PNEB pathway for the 8OG:A system proceeds through the major groove, although it maintains an important similarity to the minor groove path (Figure 5-12). The 8OG base is initially bound in a *syn* orientation rather than the canonical anti orientation around the glycosyl angle. In much the same way as eversion via a minor groove pathway proceeds, the 8OG rotates around its glycosyl angle during the initial eversion stages. The base reorients itself in the intrahelical site in order to optimize favorable electrostatic contacts along the major groove pathway, which is basic. The oxygen O8 interaction with these residues (Arg263, Asn173, amide backbone of Gly264) would be favorable, while the interaction of the Watson-Crick hydrogen bonding face of 8OG would be unfavorable. It can be predicted from the rotation barrier in the minor groove pathway that this reorientation of 8OG would produce unfavorable steric clashes and limit base eversion of a *syn*-oriented residue.

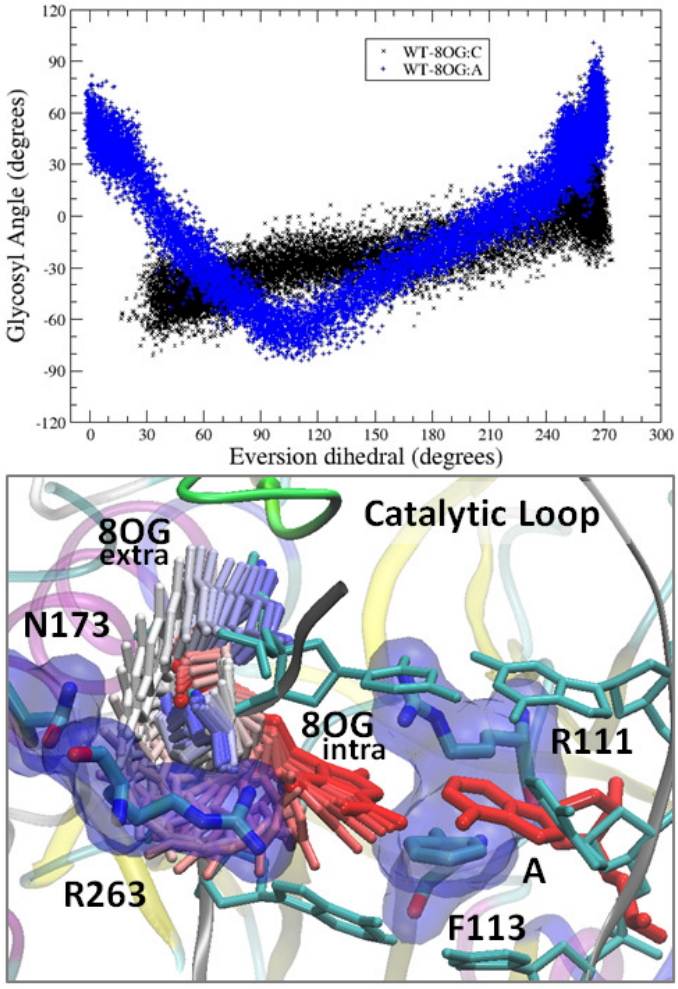

Figure 5-12 Top: Eversion paths through the major groove for WT (black) and 8OGA (blue) systems. Since the eversion dihedral is based on the position of the imidazole ring of 8OG, the values seen in the intrahelical base in the syn orientation are lower than those seen in the WT. This reflects the position of the 5-membered ring pointing towards the minor groove in the *syn* orientation (8OG:A), and towards the major groove when in the *anti* orientation (8OG:C). Bottom: Restart files from PNEB simulation of 8OG:A system show major groove pathway of eversion. The base is colored by timestep from red to blue. Note initial rotation of 8OG in the 8OG:A system occurs during the early stages of base eversion (from an eversion dihedral of 0^{o} to 90^{o}), which reorients the 8OG base below Arg263. After this interaction at an eversion dihedral value of 90^{o}, the pathway is similar to the WT system.

The initial *syn* to *anti* rotation in the 8OG:A system resolves unfavorable electrostatic clashes in the intrahelical location (defined as an eversion dihedral value below 60°). The normalized histogram of the interaction energy between the sidechain of Arg263 and the 8OG base, shown in Figure 5-13, shows a more favorable interaction when 8OG is in an *anti* orientation, as seen when base paired opposite C, and a less favorable interaction when 8OG is in a *syn* conformation, when mispaired with A. Poor electrostatic interactions between the $8OG_{syn}$ and Arg263 promote rotation around the glycosyl angle while the base is intrahelical.

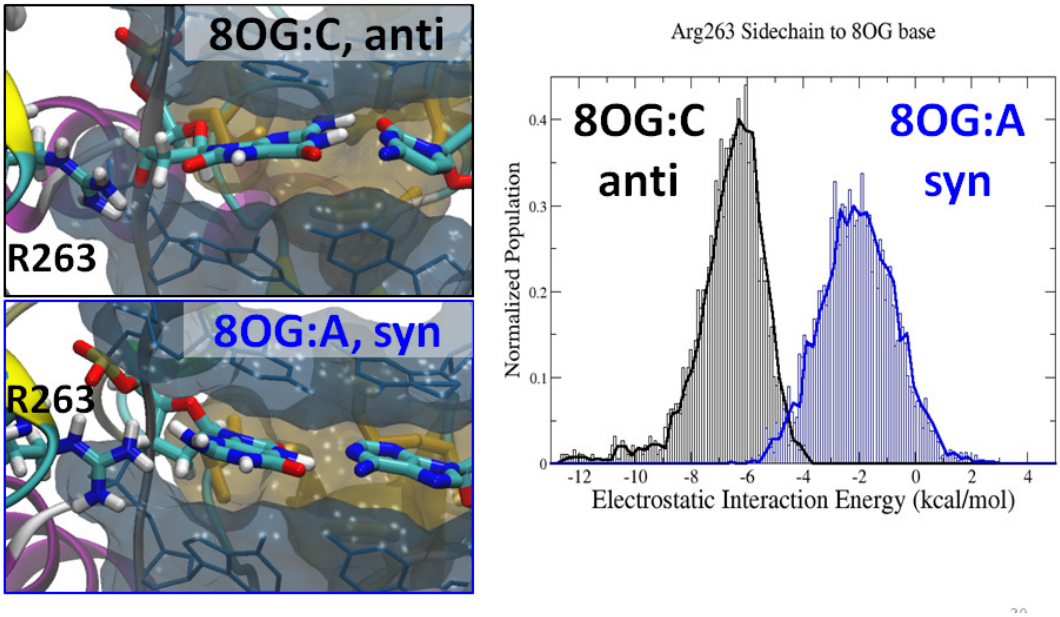

Figure 5-13 Left, top: Intrahelical position of 8OG:C base pair, shown with the O8 atom oriented towards Arg263 in an *anti* conformation. Left, bottom: Intrahelical position of 8OG:A base pair, with Watson-Crick hydrogen bonding face oriented towards Arg263 in a *syn* conformation. Right: Histogram of electrostatic interaction energy for intrahelical region of WT-8OG:C (black) and WT-8OG:A (blue) systems. The intrahelical region was defined as all structures with an eversion dihedral angle value of 60° and below.

Interaction energies of the everting 8OG base to its surrounding DNA quantify the unfavorable energy during initial base eversion (Figure 5-14). These energies are zeroed in the extrahelical site (120^{o}). During the initial rotation around the glycosyl angle (0^{o} to 70^{o} in the eversion dihedral, 60^{o} to -60^{o} in the glycosyl torsion) the vdW relative interaction energy barrier is much less favorable for the 8OG:A system than the WT simulation. Once rotation around the glycosidic torsion is accomplished, and the base is no longer in the intrahelical location, the energies are more similar. This can be expected since the O8 of the 8OG base in the 8OG:A system is now oriented similarly to its WT counterpart.

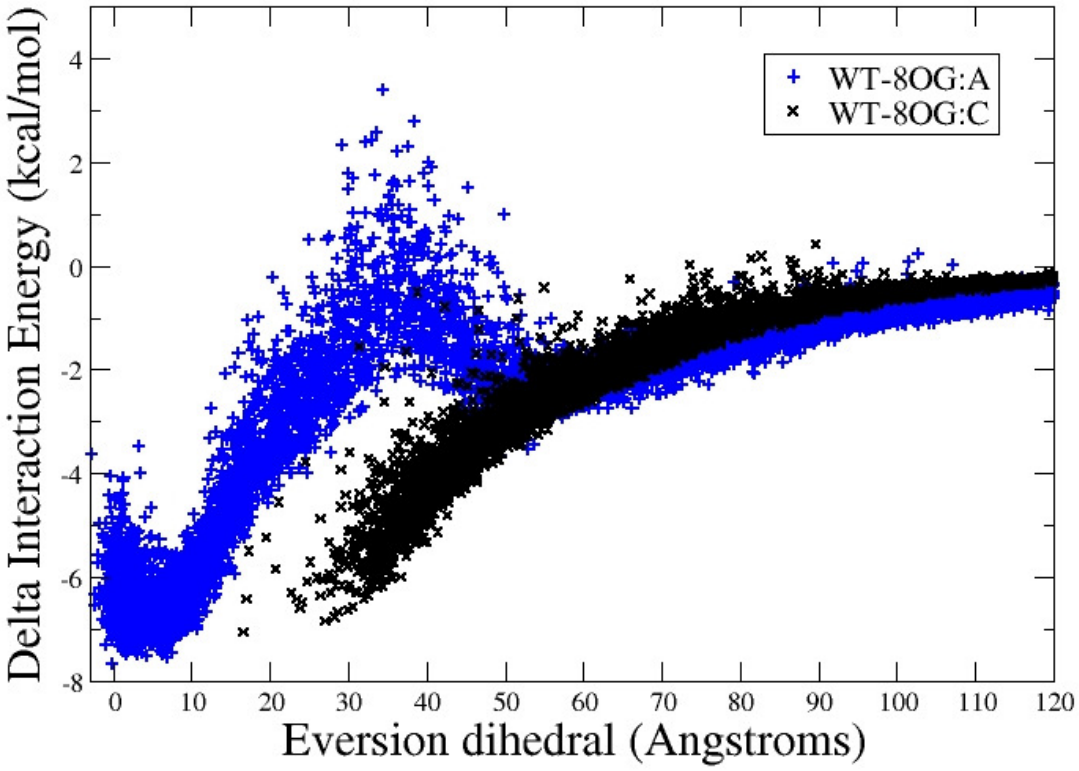

Figure 5-14 Relative van der Waals interaction energy vs. eversion dihedral between the 8OG (ligand) with the orphaned base and surrounding base pairs (receptor). The top graph shows the combined nonbonded interaction energy. The second graph shows the electrostatic component and the bottom graph shows the van der Waals component of the energies.

The simulations pose two distinct hypothesis about how Fpg prevents processing of 8OG:A base pairs. First, rotation of 8OG around the glycosyl angle while in an intrahelical orientation would never be a favorable process and thus is unlikely to occur. Second, the inability of Arg111 to intercalate into the helix during the latter part of eversion does not stabilize the transition state structure to the degree seen in the WT system. Future work on the WT-8OG:A mutant system will use 2D umbrella sampling to generate the PMF and compare the free energies of these processes. However, intrahelical rotation around the glycosyl angle is extremely unfavorable; we can tell from the minor groove PMF in the WT system that this action results in a high free energy barrier. This is most likely too high an energy penalty to pay for the 8OG to proceed along the eversion pathway, acting as discrimination against *syn* oriented 8OG. Destabilization of the transition state seems to be an additional check by the R111 residue of Fpg to assure 8OG opposite an A does not reach the active site and is not processed.

5.4 Conclusion

The goal of this work was to ascertain whether the intercalating wedge residues which played an active role in base eversion. To this end, we calculated the pathways of 8OG base eversion in the presence of WT Fpg, F113A Fpg, R111A Fpg, and WT Fpg bound to a non-substrate 8OG:A base pair. The Phe113 wedge actively alters the 8OG:C base pair, increasing the buckle and destabilizing it to a non-optimal geometry. Wedge insertion accompanies binding to Fpg and binding energy is shifted to destabilizing the base pair. The F113A system has optimal geometry in the intrahelical location, evidenced by the lower free energy of this region in the PMF. The better geometry allows hydrogen bonds to persist for longer values of the distance

reaction coordinate, while the WT system has strained bonds which more readily break, allowing eversion to occur.

The Arg111 intercalating residue has been hypothesized to recognize/stabilize 8OG base pair partners and discriminate pyrimidine from purine opposite 8OG. Here, we found that it plays an active role in stabilizing a transition state intermediate along the base eversion pathway. By removing the Arg111 side chain in the R111A mutant system, stabilization of the transition state structure was abolished. A favorable interaction between the R111, 8OG backbone, and orphaned C was diminished when C was replaced with A, showing discrimination against a poor substrate relative to the WT simulation at this late intermediate. This correlates well with the catalytic activity of Fpg.

6 SUMMARY

In the present work the process by which 8OG damage is repaired by Fpg was examined. This work links two important steps in this repair pathway – the recognition of intrahelical damage by Fpg, and the eversion profile of the damage from the DNA helix into the active site of the enzyme. The dynamics of these events are difficult to probe with experiments, so simulations have been used to gain atomic level insight into these high-energy low-populated events. This required the development of timescale-independent simulation methods to determine paths of conformational transitions.

Simulations show a distinct difference in the dynamics of the 8OG base when compared to its undamaged G counterpart. The 8OG backbone torsions prefer to populate a BII substate of DNA. Upon Fpg binding, the free energy penalty for shifting the 8OG to the bound conformation is half that of the G. This provides evidence towards the conformational capture of 8OG by the enzyme, as well as selection against undamaged bases due to the larger free energy barrier for binding G. Interestingly, conformational characteristics of the BII substate which 8OG prefers have been shown to lead to base eversion. Here, 8OG is displaced towards the major groove, which is the proposed direction of base eversion into the active site where catalysis occurs.

The eversion of 8OG from an intrahelical position into the active site of Fpg where catalysis occurs is a critical step in repair. The major and minor groove pathways of lesion processing were simulated and their free energies compared. The major groove free energy pathway is significantly lower than the minor groove pathway, indicating it is the preferred pathway for lesion eversion. Base eversion via the major groove has also been proposed for the human analog of Fpg, hOGG1.[76; 121] The simulations indicate conservation between the

prokaryotic and eukaryotic enzymes, suggesting functional similarity in the dynamics of 8OG processing.

Development of the PNEB method allowed isolation of the minimum energy pathways of conformational transitions in the Fpg-DNA complex. In general, this method can be targeted to a specific transition by using atom masks to designate parts of a larger system on which to perform the calculation, while the rest of the system still experiences a standard MD simulation environment. It ultimately allows calculation of the most likely pathway for conformational transitions without imposing restraints along a reaction coordinate which may or may be a good description of the path. Often a reasonable reaction coordinate can be characterized from analysis of the PNEB pathway.

As previously mentioned the PNEB method is being employed to find the pathway of translocation of Fpg along DNA. Kinetics experiments suggest a low barrier for sliding along DNA, and the Phe113 wedge is proposed to interrogate base pairs searching for the substrate 8OG:C. Current results from Alexandra Carvalho show PNEB pathways preferentially insert the Phe wedge to interrogate all in a stretch of three base pairs, given no structural information about the interrogation of the central base pair. This method is also being used to study hOGG1, the human analog of Fpg, for dynamic similarity in base eversion by Haoquan Li.

ENDNOTES AND BIBLIOGRAPHY

1. Chargaff, E., Zamenhof, S. & Green, C. (1950). Composition of human desoxypentose nucleic acid. *Nature* **165**, 756-7.

2. Watson, J. D. & Crick, F. H. C. (1953). Molecular Structure of Nucleic Acids - a Structure for Deoxyribose Nucleic Acid. *Nature* **171**, 737-738.

3. David, S. S., O'Shea, V. L. & Kundu, S. (2007). Base-excision repair of oxidative DNA damage. *Nature* **447**, 941-950.

4. Friedberg, E. C., Walker, G. C. & Siede, W. (1995). *DNA repair and mutagenesis*, ASM Press, Washington, D.C.

5. Lindahl, T. (1993). Instability and Decay of the Primary Structure of DNA. *Nature* **362**, 709-715.

6. Scharer, O. D. (2003). Chemistry and biology of DNA repair. *Angewandte Chemie-International Edition* **42**, 2946-2974.

7. Hoeijmakers, J. H. J. (2001). Genome maintenance mechanisms for preventing cancer. *Nature* **411**, 366-374.

8. Cheng, K. C., Cahil, D.S., Kasai, H., Nishimura, S., Loeb, L.A. (1992). 8-Hydroxyguanine, an Abundant Form of Oxidative DNA Damage, Causes G -> T and A ->C Substitutions. *Journal of Biological Chemistry* **267**, 166-172.

9. Michaels, M. L., Tchou, J., Grollman, A. P. & Miller, J. H. (1992). A repair system for 8-oxo-7,8-dihydrodeoxyguanine. *Biochemistry* **31**, 10964-8.

10. Fowler, R. G., White, S. J., Koyama, C., Moore, S. C., Dunn, R. L. & Schaaper, R. M. (2003). Interactions among the Escherichia coli mutT, mutM, and mutY damage prevention pathways. *DNA Repair (Amst)* **2**, 159-73.

11. Chetsanga, C. J. & Lindahl, T. (1979). Release of 7-Methylguanine Residues Whose Imidazole Rings Have Been Opened from Damaged DNA by a DNA Glycosylase from Escherichia-Coli. *Nucleic Acids Research* **6**, 3673-3684.

12. Cabrera, M., Nghiem, Y. & Miller, J. H. (1988). Mutm, a 2nd Mutator Locus in Escherichia-Coli That Generates G.C-]T.A Transversions. *Journal of Bacteriology* **170**, 5405-5407.

13. Tchou, J., Kasai, H., Shibutani, S., Chung, M. H., Laval, J., Grollman, A. P. & Nishimura, S. (1991). 8-Oxoguanine (8-Hydroxyguanine) DNA Glycosylase and Its

Substrate-Specificity. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 4690-4694.

14. Zharkov, D. O., Rieger, R. A., Iden, C. R. & Grollman, A. P. (1997). NH2-terminal proline acts as a nucleophile in the glycosylase/AP-lyase reaction catalyzed by Escherichia coli formamidopyrimidine-DNA glycosylase (Fpg) protein. *J Biol Chem* **272**, 5335-41.

15. Grollman, A. P. & Moriya, M. (1993). Mutagenesis by 8-Oxoguanine - an Enemy Within. *Trends in Genetics* **9**, 246-249.

16. Tchou, J., Bodepudi, V., Shibutani, S., Antoshechkin, I., Miller, J., Grollman, A. P. & Johnson, F. (1994). Substrate-Specificity of Fpg Protein - Recognition and Cleavage of Oxidatively Damaged DNA. *Journal of Biological Chemistry* **269**, 15318-15324.

17. Bhagwat, M. & Gerlt, J. A. (1996). 3'- and 5'-strand cleavage reactions catalyzed by the Fpg protein from Escherichia coli occur via successive beta- and delta-elimination mechanisms, respectively. *Biochemistry* **35**, 659-65.

18. Gilboa, R., Zharkov, D. O., Golan, G., Fernandes, A. S., Gerchman, S. E., Matz, E., Kycia, J. H., Grollman, A. P. & Shoham, G. (2002). Structure of formamidopyrimidine-DNA glycosylase covalently complexed to DNA. *J Biol Chem* **277**, 19811-6.

19. Fromme, J. C. & Verdine, G. L. (2003). DNA lesion recognition by the bacterial repair enzyme MutM. *J Biol Chem* **278**, 51543-8.

20. Sugahara, M., Mikawa, T., Kumasaka, T., Yamamoto, M., Kato, R., Fukuyama, K., Inoue, Y. & Kuramitsu, S. (2000). Crystal structure of a repair enzyme of oxidatively damaged DNA, MutM (Fpg), from an extreme thermophile, Thermus thermophilus HB8. *Embo Journal* **19**, 3857-3869.

21. Banerjee, A., Santos, W. L. & Verdine, G. L. (2006). Structure of a DNA glycosylase searching for lesions. *Science* **311**, 1153-7.

22. Qi, Y., Spong, M. C., Nam, K., Banerjee, A., Jiralerspong, S., Karplus, M. & Verdine, G. L. (2009). Encounter and extrusion of an intrahelical lesion by a DNA repair enzyme. *Nature* **462**, 762-U79.

23. Fromme, J. C. & Verdine, G. L. (2002). Structural insights into lesion recognition and repair by the bacterial 8-oxoguanine DNA glycosylase MutM. *Nat Struct Biol* **9**, 544-52.

24. Kuznetsov, N. A., Koval, V. V., Zharkov, D. O., Vorobjev, Y. N., Nevinsky, G. A., Douglas, K. T. & Fedorova, O. S. (2007). Pre-steady-state kinetic study of substrate specificity of Escherichia coli formamidopyrimidine--DNA glycosylase. *Biochemistry* **46**, 424-35.

25. Koval, V. V., Kuznetsov, N. A., Zharkov, D. O., Ishchenko, A. A., Douglas, K. T., Nevinsky, G. A. & Fedorova, O. S. (2004). Pre-steady-state kinetics shows differences in processing of various DNA lesions by Escherichia coli formamidopyrimidine-DNA glycosylase. *Nucleic Acids Res* **32**, 926-35.

26. Zharkov, D. O. & Grollman, A. P. (2005). The DNA trackwalkers: principles of lesion search and recognition by DNA glycosylases. *Mutat Res* **577**, 24-54.

27. Lindorff-Larsen, K., Piana, S., Dror, R. O. & Shaw, D. E. (2011). How fast-folding proteins fold. *Science* **334**, 517-20.

28. Karplus, M. & McCammon, J. A. (2002). Molecular dynamics simulations of biomolecules (vol 9, pg 646, 2002). *Nature Structural Biology* **9**, 788-788.

29. Alder, B. J., Wainwright, T.E. (1957). Phase Transition for a Hard Sphere System. *Journal of Chemical Physics* **27**, 1208-1209.

30. Stillinger, F. H., Rahman, A. (1974). Improved Simulations of Liquid Water by Molecular-Dynamics. *Journal of Chemical Physics* **60**, 1545-1557.

31. McCammon, J. A., Gelin, B. R., Karplus, M. . (1977). Dynamics of Folded Proteins. *Nature* **267**, 585-590.

32. Weiner, S. J., Kollman, P. A., Nguyen, D. T. & Case, D. A. (1986). An All Atom Force-Field for Simulations of Proteins and Nucleic-Acids. *Journal of Computational Chemistry* **7**, 230-252.

33. Cornell, W. D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W, Kollman, P.A. (1995). A Second Generation Force-Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J Am Chem Soc* **117**, 5179 - 5197.

34. Bayly, C. I., Cieplak, P., Cornell, W. D. & Kollman, P. A. (1993). A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the Resp Model. *Journal of Physical Chemistry* **97**, 10269-10280.

35. Cieplak, P., Cornell, W. D., Bayly, C. & Kollman, P. A. (1995). Application of the Multimolecule and Multiconformational Resp Methodology to Biopolymers - Charge Derivation for DNA, Rna, and Proteins. *Journal of Computational Chemistry* **16**, 1357-1377.

36. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A. & Simmerling, C. (2006). Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins-Structure Function and Bioinformatics* **65**, 712-725.

37. Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T., Caldwell, J., Wang, J. & Kollman, P. (2003). A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *Journal of Computational Chemistry* **24**, 1999-2012.

38. Best, R. B., Buchete, N. V. & Hummer, G. (2008). Are current molecular dynamics force fields too helical? *Biophys J* **95**, L07-9.

39. Wickstrom, L., Okur, A. & Simmerling, C. (2009). Evaluating the performance of the ff99SB force field based on NMR scalar coupling data. *Biophys J* **97**, 853-6.

40. Lange, O. F., van der Spoel, D. & de Groot, B. L. (2010). Scrutinizing molecular mechanics force fields on the submicrosecond timescale with NMR data. *Biophys J* **99**, 647-55.

41. Cheatham, T. E., 3rd, Cieplak, P. & Kollman, P. A. (1999). A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *J Biomol Struct Dyn* **16**, 845-62.

42. Perez, A., Marchan, I., Svozil, D., Sponer, J., Cheatham, T. E., Laughton, C. A. & Orozco, M. (2007). Refinenement of the AMBER force field for nucleic acids: Improving the description of alpha/gamma conformers. *Biophysical Journal* **92**, 3817-3829.

43. Tsui, V. & Case, D. A. (2000). Theory and applications of the generalized Born solvation model in macromolecular Simulations. *Biopolymers* **56**, 275-291.

44. Joung, I. S. & Cheatham, T. E., 3rd. (2009). Molecular dynamics simulations of the dynamic and energetic properties of alkali and halide ions using water-model-specific ion parameters. *Journal of Physical Chemistry B* **113**, 13279-90.

45. Perez, A., Luque, F. J. & Orozco, M. (2007). Dynamics of B-DNA on the microsecond time scale. *Journal of the American Chemical Society* **129**, 14739-14745.

46. Lavery, R., Zakrzewska, K., Beveridge, D., Bishop, T. C., Case, D. A., Cheatham, T., 3rd, Dixit, S., Jayaram, B., Lankas, F., Laughton, C., Maddocks, J. H., Michon, A., Osman, R., Orozco, M., Perez, A., Singh, T., Spackova, N. & Sponer, J. (2010). A systematic molecular dynamics study of nearest-neighbor effects on base pair and base pair step conformations and fluctuations in B-DNA. *Nucleic Acids Res* **38**, 299-313.

47. Henzler-Wildman, K. & Kern, D. (2007). Dynamic personalities of proteins. *Nature* **450**, 964-972.

48. Kuznetsov, N. A., Koval, V. V., Zharkov, D. O., Vorobjev, Y. N., Nevinsky, G. A., Douglas, K. T. & Fedorova, O. S. (2007). Pre-steady-state kinetic study of substrate specificity of Escherichia coli formamidopyrimidine-DNA glycosylase. *Biochemistry* **46**, 424-435.

49. Nguyen, D. T. & Case, D. A. (1985). On finding stationary states on large-molecule potential energy surfaces. *The Journal of Physical Chemistry* **89**, 4020-4026.

50. Bolhuis, P. G., Chandler, D., Dellago, C. & Geissler, P. L. (2002). Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Annual Review of Physical Chemistry* **53**, 291-318.

51. Dellago, C., Bolhuis, P. G., Csajka, F. S. & Chandler, D. (1998). Transition path sampling and the calculation of rate constants. *Journal of Chemical Physics* **108**, 1964-1977.

52. Elber, R. & Karplus, M. (1987). A Method for Determining Reaction Paths in Large Molecules - Application to Myoglobin. *Chemical Physics Letters* **139**, 375-380.

53. Jonsson, H., Mills, G., Jacobsen, K.W. (1998). Nudged elastic band method for finding minimum energy paths of transitions. In *Classical and Quantum Dynamics in Condensed Phase Simulations, Edition B* (Berne, J., Ciccotti, G., Cooker, D.F., ed.), pp. 385-404. World Scientific.

54. Jónsson, H. (1994). Quantum and thermal effects in H2 dissociative adsorption: Evaluation of free energy barriers in multidimensional quantum systems. *Physical Review Letters* **72**, 1124 -1127.

55. Mills, G., Jónsson, H. & Schenter, G. K. (1995). Reversible work transition state theory: application to dissociative adsorption of hydrogen. *Surf. Sci.* **324**.

56. Henkelman, G. & Jonsson, H. (2000). Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *Journal of Chemical Physics* **113**, 9978-9985.

57. Sheppard, D., Terrell, R. & Henkelman, G. (2008). Optimization methods for finding minimum energy paths. *Journal of Chemical Physics* **128**, -.

58. Chu, J. W., Trout, B. L. & Brooks, B. R. (2003). A super-linear minimization scheme for the nudged elastic band method. *Journal of Chemical Physics* **119**, 12708-12717.

59. Henkelman, G., Uberuaga, B. P. & Jonsson, H. (2000). A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *Journal of Chemical Physics* **113**, 9901-9904.

60. Maragakis, P., Andreev, S. A., Brumer, Y., Reichman, D. R. & Kaxiras, E. (2002). Adaptive nudged elastic band approach for transition state calculation. *Journal of Chemical Physics* **117**, 4651-4658.

61. Trygubenko, S. A. & Wales, D. J. (2004). A doubly nudged elastic band method for finding transition states. *Journal of Chemical Physics* **120**, 2082-2094.

62. Carr, J. M., Trygubenko, S. A. & Wales, D. J. (2005). Finding pathways between distant local minima. *Journal of Chemical Physics* **122**, -.

63. Hestenes, M. R. & Stiefel, E. (1952). Methods of Conjugate Gradients for Solving Linear Systems. *Journal of Research of the National Bureau of Standards* **49**, 409-436.

64. Nocedal, J. (1980). Updating Quasi-Newton Matrices with Limited Storage. *Mathematics of Computation* **35**, 773-782.

65. Mathews, D. H. & Case, D. A. (2006). Nudged elastic band calculation of minimal energy paths for the conformational change of a GG non-canonical pair. *J Mol Biol* **357**, 1683-93.

66. Cisneros, G. A., Liu, H. Y., Lu, Z. Y. & Yang, W. T. (2005). Reaction path determination for quantum mechanical/molecular mechanical modeling of enzyme reactions by combining first order and second order "chain-of-replicas" methods. *Journal of Chemical Physics* **122**, -.

67. Torrie, G. M. & Valleau, J. P. (1977). Non-Physical Sampling Distributions in Monte-Carlo Free-Energy Estimation - Umbrella Sampling. *Journal of Computational Physics* **23**, 187-199.

68. Kumar, S., Bouzida, D., Swendsen, R. H., Kollman, P. A. & Rosenberg, J. M. (1992). The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules .1. The Method. *Journal of Computational Chemistry* **13**, 1011-1021.

69. Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. (1995). Multidimensional Free-Energy Calculations Using the Weighted Histogram Analysis Method. *Journal of Computational Chemistry* **16**, 1339-1350.

70. Roux, B. (1995). The Calculation of the Potential of Mean Force Using Computer-Simulations. *Computer Physics Communications* **91**, 275-282.

71. Geney, R., Layten, M., Gomperts, R., Hornak, V. & Simmerling, C. (2006). Investigation of salt bridge stability in a generalized born solvent model. *Journal of Chemical Theory and Computation* **2**, 115-127.

72. Huang, N., Banavali, N. K. & MacKerell, A. D., Jr. (2003). Protein-facilitated base flipping in DNA by cytosine-5-methyltransferase. *Proc Natl Acad Sci U S A* **100**, 68-73.

73. Song, K., Campbell, A. J., Bergonzo, C., de los Santos, C., Grollman, A. P. & Simmerling, C. (2009). An Improved Reaction Coordinate for Nucleic Acid Base Flipping Studies. *Journal of Chemical Theory and Computation* **5**, 3105-3113.

74. Blainey, P. C., van Oijent, A. M., Banerjee, A., Verdine, G. L. & Xie, X. S. (2006). A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 5752-5757.

75. Zharkov, D. O. (2007). [Structure and conformational dynamics of base excision repair DNA glycosylases]. Trans. Dimitry Zharkov. *Mol Biol (Mosk)* **41**, 772-86.

76. Banerjee, A., Yang, W., Karplus, M. & Verdine, G. L. (2005). Structure of a repair enzyme interrogating undamaged DNA elucidates recognition of damaged DNA. *Nature* **434**, 612-618.

77. Oda, Y., Uesugi, S., Ikehara, M., Nishimura, S., Kawase, Y., Ishikawa, H., Inoue, H. & Ohtsuka, E. (1991). Nmr-Studies of a DNA Containing 8-Hydroxydeoxyguanosine. *Nucleic Acids Research* **19**, 1407-1412.

78. Lipscomb, L. A., Peek, M. E., Morningstar, M. L., Verghis, S. M., Miller, E. M., Rich, A., Essigmann, J. M. & Williams, L. D. (1995). X-Ray Structure of a DNA Decamer Containing 7,8-Dihydro-8-Oxoguanine. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 719-723.

79. Broyde, S., Wang, L. H., Zhang, L., Rechkoblit, O., Geacintov, N. E. & Patel, D. J. (2008). DNA adduct structure-function relationships: Comparing solution with polymerase structures. *Chemical Research in Toxicology* **21**, 45-52.

80. Svozil, D., Sponer, J. E., Marchan, I., Perez, A., Cheatham, T. E., Forti, F., Luque, F. J., Orozco, M. & Sponer, J. (2008). Geometrical and electronic structure variability of the sugar-phosphate backbone in nucleic acids. *Journal of Physical Chemistry B* **112**, 8188-8197.

81. Trieb, M., Rauch, C., Wellenzohn, B., Wibowo, F., Loerting, T. & Liedl, K. R. (2004). Dynamics of DNA: B-I and B-II phosphate backbone transitions. *Journal of Physical Chemistry B* **108**, 2470-2476.

82. Wibowo, F. R., Rauch, C., Trieb, M. & Liedl, K. R. (2005). M.TaqI facilitates the base flipping via an unusual DNA backbone conformation. *Biopolymers* **79**, 128-138.

83. Bertrand, H. O., Ha-Duong, T., Fermandjian, S. & Hartmann, B. (1998). Flexibility of the B-DNA backbone: effects of local and neighbouring sequences on pyrimidine-purine steps. *Nucleic Acids Research* **26**, 1261-1267.

84. Yang, W. (2006). Poor base stacking at DNA lesions may initiate recognition by many repair proteins. *DNA Repair* **5**, 654-666.

85. Cheng, X. L., Kelso, C., Hornak, V., de los Santos, C., Grollman, A. P. & Simmerling, C. (2005). Dynamic behavior of DNA base pairs containing 8-oxoguanine. *Journal of the American Chemical Society* **127**, 13906-13918.

86. Banerjee, A., Santos, W.L., Verdine, G.L. (2006). Structure of a DNA Glycosylase Searching for Lesions. *Science* **311**, 1153-1157.

87. D.A. Case, T. A. D., T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, K.M. Merz, D.A. Pearlman, M. Crowley, R.C. Walker, W. Zhang, B. Wang, S. Hayik, A. Roitberg, G. Seabra, K.F. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, P. Beroza, D.H. Mathews, C. Schafmeister, W.S. Ross, and P.A. Kollman (2006). AMBER9. University of California, San Francisco.

88. Miller, J. H., Fan-Chiang, C. C. P., Straatsma, T. P. & Kennedy, M. A. (2003). 8-Oxoguanine enhances bending of DNA that favors binding to glycosylases. *Journal of the American Chemical Society* **125**, 6331-6336.

89. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A. & Simmerling, C. (2006). Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **65**, 712-25.

90. Jorgnesen, W. L., Chandrasekhar, J., Madura, J.D., Impey, R.W., Klein, M.L. (1983). Comparison of Simple Potential Functions for Simulating Liquid Water. *Journal of Chemical Physics* **79**, 926-935.

91. Berendsen, H. J. C., Postma, J. P. M., Vangunsteren, W. F., Dinola, A. & Haak, J. R. (1984). Molecular-Dynamics with Coupling to an External Bath. *Journal of Chemical Physics* **81**, 3684-3690.

92. Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. (1977). Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *Journal of Computational Physics* **23**, 327-341.

93. Shao, J. Y., Tanner, S. W., Thompson, N. & Cheatham, T. E. (2007). Clustering molecular dynamics trajectories: 1. Characterizing the performance of different clustering algorithms. *Journal of Chemical Theory and Computation* **3**, 2312-2334.

94. Bergonzo, C., Campbell, A. J., de los Santos, C., Grollman, A. P. & Simmerling, C. (2011). Energetic Preference of 8-oxoG Eversion Pathways in a DNA Glycosylase. *Journal of the American Chemical Society* **133**, 14504-14506.

95. Cheatham, T. E., Miller, J. L., Fox, T., Darden, T. A. & Kollman, P. A. (1995). Molecular-Dynamics Simulations on Solvated Biomolecular Systems - the Particle Mesh Ewald Method Leads to Stable Trajectories of DNA, Rna, and Proteins. *Journal of the American Chemical Society* **117**, 4193-4194.

96. Darden, T., York, D. & Pedersen, L. (1993). Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *Journal of Chemical Physics* **98**, 10089-10092.

97. Simmerling, C., Miller, J. L. & Kollman, P. A. (1998). Combined locally enhanced sampling and Particle Mesh Ewald as a strategy to locate the experimental structure of a nonhelical nucleic acid. *Journal of the American Chemical Society* **120**, 7149-7155.

98. Hawkins, G. D., Cramer, C. J. & Truhlar, D. G. (1995). Pairwise Solute Descreening of Solute Charges from a Dielectric Medium. *Chemical Physics Letters* **246**, 122-129.

99. Hawkins, G. D., Cramer, C. J. & Truhlar, D. G. (1996). Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *Journal of Physical Chemistry* **100**, 19824-19839.

100. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. (1983). Comparison of Simple Potential Functions for Simulating Liquid Water. *Journal of Chemical Physics* **79**, 926-935.

101. Loncharich, R. J., Brooks, B. R. & Pastor, R. W. (1992). Langevin Dynamics of Peptides - the Frictional Dependence of Isomerization Rates of N-Acetylalanyl-N'-Methylamide. *Biopolymers* **32**, 523-535.

102. O'Neil, L. L., Grossfield, A. & Wiest, O. (2007). Base flipping of the thymine dimer in duplex DNA. *Journal of Physical Chemistry B* **111**, 11843-11849.

103. Roe, D. R., Okur, A., Wickstrom, L., Hornak, V. & Simmerling, C. (2007). Secondary structure bias in generalized born solvent models: Comparison of conformational ensembles and free energy of solvent polarization from explicit and implicit solvation. *Journal of Physical Chemistry B* **111**, 1846-1857.

104. Feig, M. (2008). Is alanine dipeptide a good model for representing the torsional preferences of protein backbones? *Journal of Chemical Theory and Computation* **4**, 1555-1564.

105. Pryor, W. A. (1986). Oxy-Radicals and Related Species: Their Formation, Lifetimes, and Reactions. *Ann. Rev. Physiol.* **48**, 657-667.

106. Tchou, J. & Grollman, A. P. (1993). Repair of DNA containing the oxidatively-damaged base, 8-oxoguanine. *Mutat Res* **299**, 277-87.

107. Bjoras, M., Seeberg, E., Luna, L., Pearl, L. H. & Barrett, T. E. (2002). Reciprocal "flipping" underlies substrate recognition and catalytic activation by the human 8-oxo-guanine DNA glycosylase. *Journal of Molecular Biology* **317**, 171-177.

108. Cao, C. Y., Jiang, Y. L., Stivers, J. T. & Song, F. H. (2004). Dynamic opening of DNA during the enzymatic search for a damaged base. *Nature Structural & Molecular Biology* **11**, 1230-1236.

109. D.A. Case, T. A. D., T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo,, M. Crowley, R. C. W., W. Zhang, K.M. Merz, B.Wang, S. Hayik, A. Roitberg, G. Seabra, I., Kolossváry, K. F. W., F. Paesani, J. Vanicek, X.Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke,, L. Yang, C. T., J. Mongan, V. Hornak, G. Cui, D.H. Mathews, M.G. Seetin, C. Sagui, V. Babin, & Kollman, a. P. A. (2008). AMBER 10. University of California, San Francisco.

110. Perlow-Poehnelt, R. A., Zharkov, D. O., Grollman, A. P. & Broyde, S. (2004). Substrate discrimination by formamidopyrimidine-DNA glycosylase: Distinguishing interactions within the active site. *Biochemistry* **43**, 16092-16105.

111. Stote, R. H. & Karplus, M. (1995). Zinc-Binding in Proteins and Solution - a Simple but Accurate Nonbonded Representation. *Proteins-Structure Function and Genetics* **23**, 12-31.

112. Onufriev, A., Bashford, D. & Case, D. A. (2004). Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins-Structure Function and Bioinformatics* **55**, 383-394.

113. Bergonzo, C., Campbell, A. J., Walker, R. C. & Simmerling, C. (2009). A Partial Nudged Elastic Band Implementation for Use With Large or Explicitly Solvated Systems. *International Journal of Quantum Chemistry* **109**, 3781-3790.

114. Tchou, J. & Grollman, A. P. (1995). The catalytic mechanism of Fpg protein. Evidence for a Schiff base intermediate and amino terminus localization of the catalytic site. *J Biol Chem* **270**, 11671-7.

115. Norman, D. P. G., Chung, S. J. & Verdine, G. L. (2003). Structural and biochemical exploration of a critical amino acid in human 8-oxoguanine glycosylase. *Biochemistry* **42**, 1564-1572.

116. Radom, C. T., Banerjee, A. & Verdine, G. L. (2007). Structural characterization of human 8-oxoguanine DNA glycosylase variants bearing active site mutations. *Journal of Biological Chemistry* **282**, 9182-9194.

117. Slupphaug, G., Mol, C. D., Kavli, B., Arvai, A. S., Krokan, H. E. & Tainer, J. A. (1996). A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. *Nature* **384**, 87-92.

118. Song, K., Hornak, V., de Los Santos, C., Grollman, A. P. & Simmerling, C. (2006). Computational analysis of the mode of binding of 8-oxoguanine to formamidopyrimidine-DNA glycosylase. *Biochemistry* **45**, 10886-94.

119. Song, K., Kelso, C., de los Santos, C., Grollman, A. P. & Simmerling, C. (2007). Molecular Simulations reveal a common binding mode for glycosylase binding of oxidatively damaged DNA lesions. *Journal of the American Chemical Society* **129**, 14536-+.

120. Campbell, A. J. (2011). Exploring DNA Damage and Repair through Molecular Dynamics Simulations. 3460574, State University of New York at Stony Brook.

121. Banerjee, A. & Verdline, G. L. (2006). A nucleobase lesion remodels the interaction of its normal neighbor in a DNA glycosylase complex. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 15020-15025.

122. Serre, L., de Jesus, K. P., Boiteux, S., Zelwer, C. & Castaing, B. (2002). Crystal structure of the Lactococcus lactis formamidopyrimidine-DNA glycosylase bound to an abasic site analogue-containing DNA. *Embo Journal* **21**, 2854-2865.

123. Scharer, O. D. & Campbell, A. J. (2009). Wedging out DNA damage. *Nat Struct Mol Biol* **16**, 102-4.

124. Bruner, S. D., Norman, D. P. G. & Verdine, G. L. (2000). Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature* **403**, 859-866.

125. Fromme, J. C., Banerjee, A., Huang, S. J. & Verdine, G. L. (2004). Structural basis for removal of adenine mispaired with 8-oxoguanine by MutY adenine DNA glycosylase. *Nature* **427**, 652-656.

126. Parikh, S. S., Mol, C. D., Slupphaug, G., Bharati, S., Krokan, H. E. & Tainer, J. A. (1998). Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA. *Embo Journal* **17**, 5214-5226.

127. Zaika, E. I., Perlow, R. A., Matz, E., Broyde, S., Gilboa, R., Grollman, A. P. & Zharkov, D. O. (2004). Substrate discrimination by formamidopyrimidine-DNA glycosylase - A mutational analysis. *Journal of Biological Chemistry* **279**, 4849-4861.

128. Plum, G. E., Grollman, A. P., Johnson, F. & Breslauer, K. J. (1995). Influence of the oxidatively damaged adduct 8-oxodeoxyguanosine on the conformation, energetics, and thermodynamic stability of a DNA duplex. *Biochemistry* **34**, 16148-16160.

129. D.A. Case, T. A. D., T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R., Luo, R. C. W., W. Zhang, K.M. Merz, B. Roberts, B. Wang, S. Hayik, A. Roitberg,, G. Seabra, I. K., K.F. Wong, F. Paesani, J. Vanicek, J. Liu, X. Wu, S.R. Brozell,, T. Steinbrecher, H. G., Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H., Mathews, M. G. S., C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and & Kollman, P. A. (2010). AMBER 11. University of California, San Francisco.

130. Morales, J. L. & Nocedal, J. (2000). Automatic preconditioning by limited memory quasi-Newton updating. *Siam Journal on Optimization* **10**, 1079-1096.

131. Tsui, V. & Case, D. A. (2000). Molecular Dynamics Simulations of Nucleic Acids with a Generalized Born Solvation Model. *Journal of the American Chemical Society* **122**, 2489-2498.

132. Nikita A. Kuznetsova, A. J. C., Christina Bergonzo, Grigory V. Mechetina, Vladimir V. Kovala, Carlos de los Santos, Arthur P. Grollman, Olga S. Fedorova, Dmitry O. Zharkova, and Carlos Simmerling (2012). Active Destabilization of Damaged Base Pairs by a DNA Glycosylase Wedge.