Exploring DNA Damage Recognition Using Molecular Modeling

A Dissertation Presented

by

Haoquan Li

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Chemistry

Stony Brook University

August 2014

Stony Brook University

The Graduate School

Haoquan Li

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

Robert Rizzo – Committee Member
Associate Professor, Department of Applied Mathematics & Statistics and Chemistry

Miguel Garcia-Diaz – Outside Member
Associate Professor, Department of Pharmacological Sciences

This dissertation is accepted by the Graduate School

Charles Taber
Dean of the Graduate School

Abstract of the Dissertation

Exploring DNA Damage Recognition Using Molecular Modeling

by

Haoquan Li

Doctor of Philosophy

in

Chemistry

Stony Brook University

2014

8-oxo-7,8,-dihydroguanine (oxoG) is one of the most prevalent DNA lesions, and has been linked to cancer and age-related diseases due to its ability to promote mutations. The oxoG damage can be removed by the Formamidopyrimidine-DNA glycosylase (Fpg) in prokaryotes and by human 8-oxoguanine glycosylase 1 (hOGG1) in human cells. Fpg and hOGG1 are structurally unrelated, but they have the same function of excising oxoG by extruding the oxoG out from the DNA helix into the enzyme's active site. Previous studies have revealed the damage recognition interactions in the active sites of these enzymes; however, their extremely fast diffusion rates along DNA almost exclude the possibility that they interrogate every encountered base in the active site. Thus, recognition of oxoG by Fpg and hOGG1 is hypothesized to occur not only at the active site, but also at the early stages of the base eversion pathway. All-atom molecular dynamics simulations were used to explore the oxoG recognition process during base eversion. Our simulations indicate that both Fpg and hOGG1 have several checkpoints for oxoG along the eversion pathway, and the geometries of those checkpoints are remarkably similar between the two enzymes, suggesting they employ similar strategies to efficiently detect oxoG from a sea of undamaged bases.

Dedication Page

This dissertation is dedicated to my beloved wife Xianyin Chen, who stands by me and takes care of me every day.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	**Full Form of Abbreviation**
Å	Angstrom
B. st.	Bacillus stearothermophilus
BER	Base excision repair
bp	Base pair
C	Cytosine
COM	Center of mass
DNA	Deoxyribonucleic acid
E. coli	Escherichia coli
Fpg	Formamidopyrimidine DNA glycosylase
G	Guanine
hOGG1	Human 8-oxoguanine glycosylase
LJ	Lennard-Jones
MD	Molecular Dynamics
NER	Nucleotide Excision Repair
oxoG	8-oxoguanine
PDB	Protein Data Bank
PMF	Potential of Mean Force
PNEB	Partial Nudged Elastic Band
RMSD	Root mean square deviation
TIP3P	Transferable Intermolecular Potential 3 Points
UDG	Uracil DNA Glycosylase
VDW	van der Waals
WHAM	Weighted Histogram Analysis Method

Acknowledgments

First, I would like to thank my advisor, Professor Carlos Simmerling, for accepting me into his lab and giving me the DNA repair project. Carlos has always been a thoughtful mentor, and I am extremely grateful for his support and guidance throughout my PhD research. I would also like to thank my committee members Professor Orlando Schärer, Professor Robert Rizzo and Professor Miguel Garcia-Diaz, for their helpful suggestions and comments during my committee meetings. I would sincerely like to thank Professor Arthur Grollman, Professor Dimitry Zharkov and Professor Carlos de los Santos, who have given me great guidance over the years for the project.

The members of the Simmerling lab have been my second family over the years. I thank Dr. Fu Lin and Dr. Cheng-Tsung Lai for the many helpful discussions from which I have learnt a lot. I also thank Dr. Arthur Campbell and Dr. Christina Bergonzo for their help in the project. Kevin Hauser, James Maier, Hai Nguyen, Dr. Carmenza Martinez, Dr. Yi Shang, He Huang, Koushik Kasavajhala, Kenneth Lam, Colleen Kirkup and Dr. Sally Pias have all helped me to solve problems in both work and life. I am thankful for all of these wonderful friends who have made science so much fun.

I have been fortunate to meet my 09’ Chem friends in Stony Brook and have so many good times together. I am thankful for having you guys in my life.

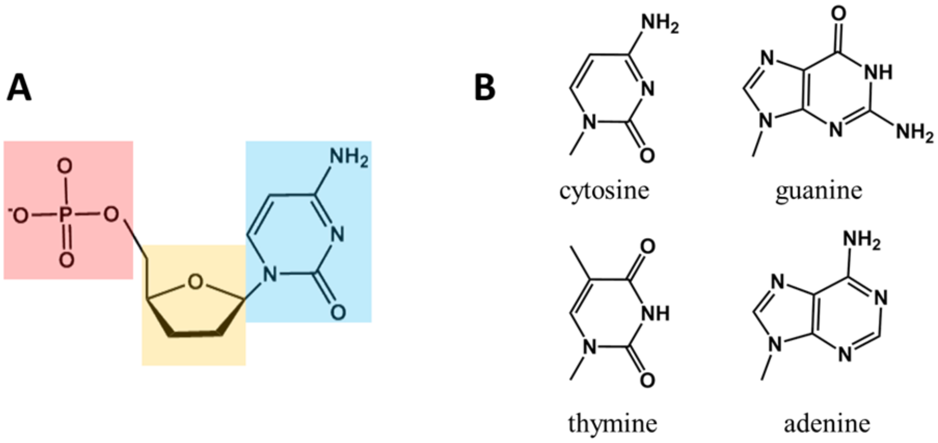
I would like to thank my family for being supportive over all these years. Finally I would like to express my deepest gratitude to my lovely wife, Xianyin Chen, whose love makes me feel complete.

1. Introduction

1.1. DNA Damage and Repair

1.1.1. Properties of DNA

Deoxyribonucleic acid (DNA) is the genetic material for almost all living organisms. A DNA molecule is a double helical polymer made of nucleotides.[1] Each nucleotide consists of three parts: a sugar ring (pentose 2-deoxyribose), a nucleobase group attached to the sugar, and a phosphate group connecting the C3' and the C5' atoms of the sugar rings (Figure 1-1A). The nucleobases are nitrogen-containing aromatic molecules, and there are four types of nucleobases in DNA: cytosine (C), thymine (T), guanine (G) and adenine (A). C and T are pyrimidines, which contains two nitrogens and four carbons in the aromatic ring; while G and A are purines, which are pyrimidines joined with an imidazole (Figure 1-1B). It is the sequence of these four nucleobases along the polymer that encodes the genetic information.



cytosine guanine thymine adenine

Figure 1-1. (A). A nucleotide comprises a phosphate group (red), a sugar ring (orange) and a nucleobase (blue). (B). Structures of the four nucleobases.

In duplex DNA, the two strands are typically held together in an antiparallel pattern, by hydrogen bonds between the nucleobases, forming a double helix that winds around a helix axis in a right-handed spiral. Each pyrimidine-type nucleobase pairs with a specific purine-type nucleobase: a C forms three hydrogen bonds to a G, whereas a T forms two hydrogen bonds to an A (Figure 1-2).[1] Such pairing pattern is often called Watson-Crick base pairing. Each base pair stacks with their two neighboring pairs via the favorable pi-pi interaction.

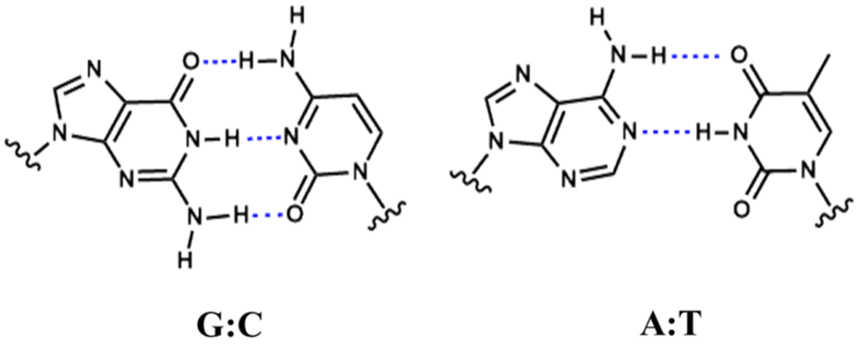


Figure 1-2. Watson-Crick DNA base pairing patterns. Hydrogen bonds are denoted by dotted lines.

There are many possible conformations for the double helix of DNA, and the most common form under the conditions found in cells is called the B-DNA form (Figure 1-3). In a B-DNA, two distinct grooves can be found between the strands: the major groove and the minor groove. These two grooves do not have a same size, since the strands are located unsymmetrically with respect to each other (Figure 1-3). The grooves are adjacent to the base pairs, and since the major groove is wider than the minor groove, the bases are more accessible from the major groove side.

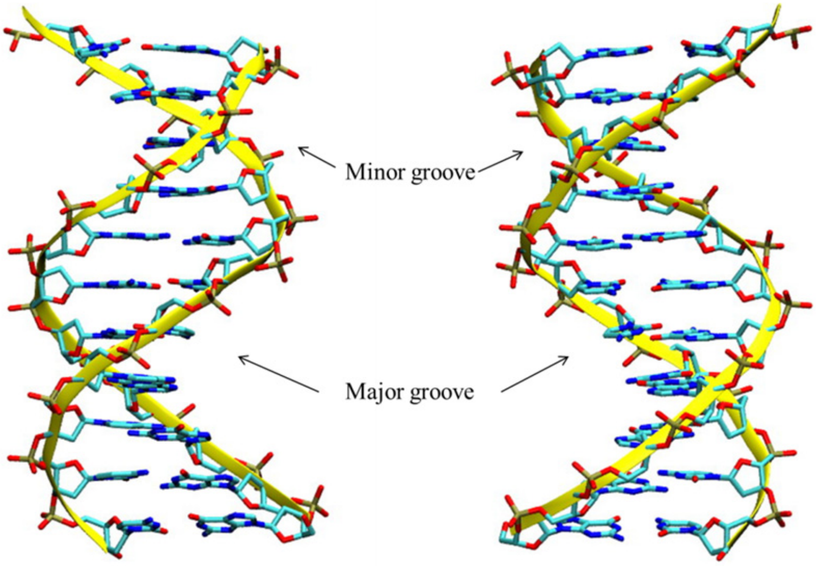


Figure 1-3. The minor and major grooves in a B-DNA, viewed from two different perspectives.

1.1.2. DNA repair pathways and DNA glycosylases

DNA integrity is continuously threatened by exogenous and endogenous agents such as UV, reactive oxygen species and alkylation agents.[2] The resulting DNA damage, if not repaired, may block replication, transcription or cause mutations that may eventually lead to aging and cancer.[3] Organisms have evolved a number of DNA repair pathways, such as direct reversal, recombination, mismatch repair, base excision repair (BER) and nucleotide excision repair (NER), to detect and repair various types of DNA damage.

While bulky helix-distorting lesions are repaired by the NER pathway, single damaged bases in DNA are mostly repaired by the BER pathway[4], which includes numerous variations, each specific for a different type of lesion. Commonly, the BER pathway is initiated by a DNA glycosylase, which removes a damaged base and creates an apurinic/apyrimidinic (AP) site. An AP endonuclease then nicks the DNA strand adjacent to the AP site by creating a 3'-OH terminus, which is then extended by a polymerase, accompanied by excision of the AP site. After the polymerase places this repair "patch" (single nucleotide or long patch), the nick in the strand is then sealed by a DNA ligase. In vivo, there are some additional scaffold proteins in BER, such as XRCC1 and Poly (ADP-ribose) Polymerases, which can stimulate and modulate the activities of the core enzymes discussed above.

The BER pathway is initiated by DNA glycosylases. DNA glycosylases share a common principle of action: they locate their substrate lesions by three-dimensional hopping combined with one-dimensional sliding along the DNA,[5, 6] and then extrude the damaged base out from the DNA duplex to the active pocket, where the catalytic reaction occurs. They hydrolyze the N-glycosidic bond between the substrate base and deoxyribose, leaving an AP site in DNA.[7] There are two main classes of glycosylases: monofunctional and bifunctional. In the case of monofunctional glycosylases, an activated water molecule displaces the damaged base by attacking C1' of the sugar moiety; in the case of bifunctional glycosylases (also known as DNA glycosylases/AP lyases), an amino group of the enzyme attacks the C1' to form a Schiff base intermediate which facilitates a series of electron rearrangements, resulting in strand cleavage, degradation of deoxyribose and regeneration of free enzyme.[8-10] Thus, a DNA strand break is yielded, with 3' α,β-unsaturated aldehyde and a 5' phosphate group. Some glycosylases further excise the aldehyde to yield a one-nucleotide gap with 3'- and 5'-terminal phosphate groups.[9] Although sharing a similar mechanism, the DNA glycosylases are structurally diverse and can be assigned to one of the three superfamilies discussed below.

The Uracil-DNA glycosylase (UDG) superfamily. UDG is probably the most extensively studied DNA glycosylase and its base excision mechanism (dissociative SN1-type) is essentially understood.[11] UDG can excise uracil from single or double stranded DNA, and does not produce breaks in DNA after base excision.[12]

The Fpg/Nei superfamily. This family is named after two prototypical homologous bifunctional glycosylases in *E. coli*: formamidopyrimidine-DNA glycosylase (Fpg) and endonuclease VIII (Nei). Most members in this superfamily use their N-terminal proline residue as the key catalytic nucleophile. In addition to the N-terminal domain, Fpg/Nei proteins have a helix-two-turn-helix (H2TH) domain and a zinc (or zinc-less) finger motif. Most Fpg/Nei proteins have broad substrate specificity.

The HhH-GPD superfamily. This family contains a range of structurally related proteins repairing diverse lesions, including alkylation, oxidative and hydrolytic damage. The superfamily is named after its hallmark Helix-hairpin-helix and Gly/Pro rich loop followed by an absolutely conserved Asp. This core HhH/GPD structure is versatile enough to cover diverse lesions, including oxidative, alkylation and hydrolytic damage. A representative member of this superfamily is human 8-oxoguanine glycosylase 1 (hOGG1), whose structure and catalytic mechanism are well characterized.[10, 13]

Although the BER have been extensively studied, there are still unresolved questions. For example, it is still unknown how DNA glycosylases detect damaged bases embedded in a million-fold excess of undamaged bases. A full understanding of mechanisms in the BER pathway can help us develop new tools for biochemistry studies (such as small-molecule modulators of glycosylase activity) and might lead to new therapeutics.

1.1.3. Properties of 8-oxoguanine and the GO repair pathway

8-oxoguanine (oxoG) is one of the most common oxidative DNA damage, with a frequency of ~1 oxoG per 10^6 guanines in eukaryotic cells.[14, 15] OxoG is generated from oxidation of guanine, keeping the overall shape unchanged; the only structural difference is located on the imidazole ring: guanine has a lone pair of electrons at the N7 atom and a hydrogen (H8) at the C8 atom, whereas oxoG has a hydrogen (H7) on N7 and an oxygen (O8) on C8 (Figure 1-4). OxoG is a particularly deleterious DNA damage because it can adopt a *syn* conformation to form a stable 8-oxoG:A base pair (Figure 1-4),[16, 17] and thus oxoG may promote G-C to T-A transversion mutation.[18] Increase in the frequency of the oxoG lesions has been linked to cancer and aging.[3, 19]

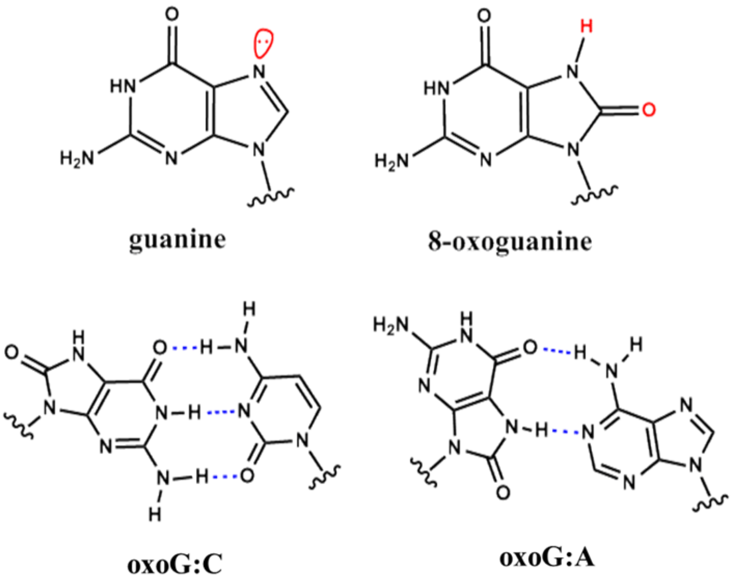


Figure 1-4. Top: comparison of guanine and oxoG. The lone pair of electrons on N7 of guanine and the H7/O8 atoms of oxoG are colored in red. Bottom: base paring of oxoG:C and oxoG:A.

The pathway that repairs the oxoG lesion is called the GO repair pathway. In bacteria, the GO repair pathway consists of MutT, Fpg (also known as MutM) and MutY.[20] Their function analogs in human are MTH1, hOGG1 and MUTYH, respectively.[21] MutT (or MTH1) hydrolyses the free 8-oxo-dGTP into 8-oxo-dGMP so as to prevent polymerases from incorporating oxoG into DNA. The bifunctional glycosylase Fpg (or hOGG1) excises oxoG from the 8-oxoG:C base pair, leaving an AP site or a nick on the DNA backbone, which is subsequently processed to restore the correct G:C base pair by other enzymes involved in the BER pathway. MutY (or MUTYH) recognizes the oxoG:A base pair and removes the inappropriate A, providing an opportunity to form an oxoG:C which is the substrate for Fpg or hOGG1. Interestingly, it has been shown that pol λ, the eukaryotic polymerase in the BER pathway, prefers to incorporate cytosine over adenine to pair opposite oxoG in the presence of auxiliary proteins, and thus prevents futile cycles of oxoG repair.[22] Figure 1-5 summarizes the BER pathway for oxoG.

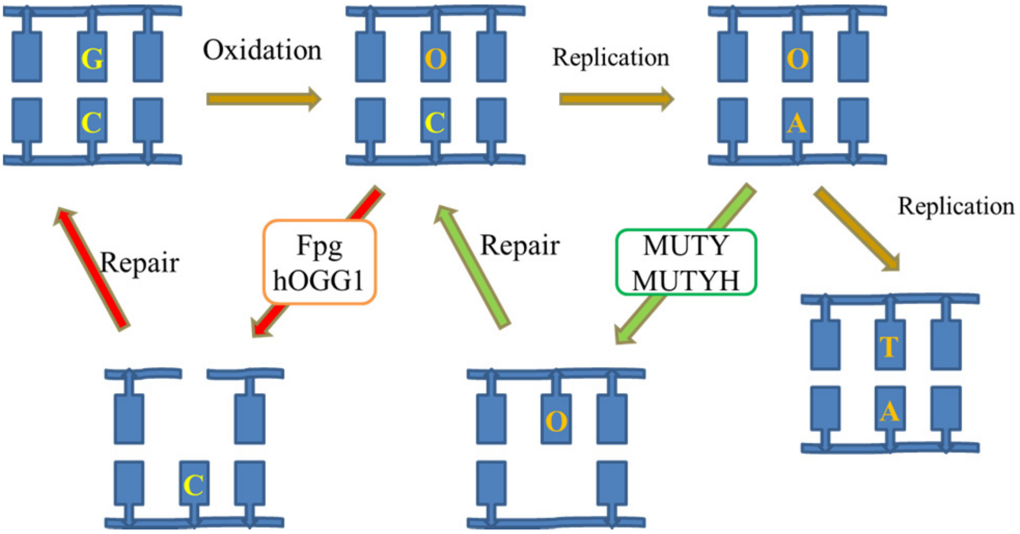


Figure 1-5. The BER pathway for oxoG. OxoG is denoted by the ‘O’ in orange. The arrows in tan indicate the transversion mutation process caused by oxoG. ‘Repair’ represents the

processing steps carried out by deoxyribophosphate lyase, AP endonuclease, and DNA polymerase and DNA ligase.

1.1.4. Excision and recognition of oxoG by Fpg and hOGG1

Fpg and hOGG1 initiate BER for oxoG by excising the oxoG paired to a cytosine. Catalytic reactions by Fpg and hOGG1 have been well characterized (Figure 1-6). The N-terminal proline of Fpg attacks the C1' atom of the sugar of oxoG, forming a Schiff base intermediate.[9, 23] Similarly, hOGG1 cleaves the glycosidic bond of oxoG by nucleophilic attack of the C1' atom with the ε-NH_2 group of Lys249.[10] After the oxoG base is cleaved, Fpg and hOGG1 act as an AP lyase and nick the DNA backbone. Fpg catalyzes successive β- and δ-elimination reactions, leaving a single nucleoside gap in DNA,[24, 25] while hOGG1 only carries out β-elimination at the 3'-side of the AP site.[10, 26]

While the oxoG excision mechanisms of Fpg and hOGG1 are well understood, their oxoG recognition mechanisms are less clear. It is still unknown how these enzymes can efficiently recognize oxoG from a vast excess of undamaged bases and without the need for energetic consumption. Stopped-flow kinetics studies have suggested that the pre-excision oxoG recognition process by Fpg and hOGG1 involve at least these common steps: (i) initial encounter and extrusion of oxoG, (ii) insertion of certain enzyme residues into DNA, and (iii) insertion of oxoG into the active site of enzyme which isomerizes for catalysis.[27, 28]

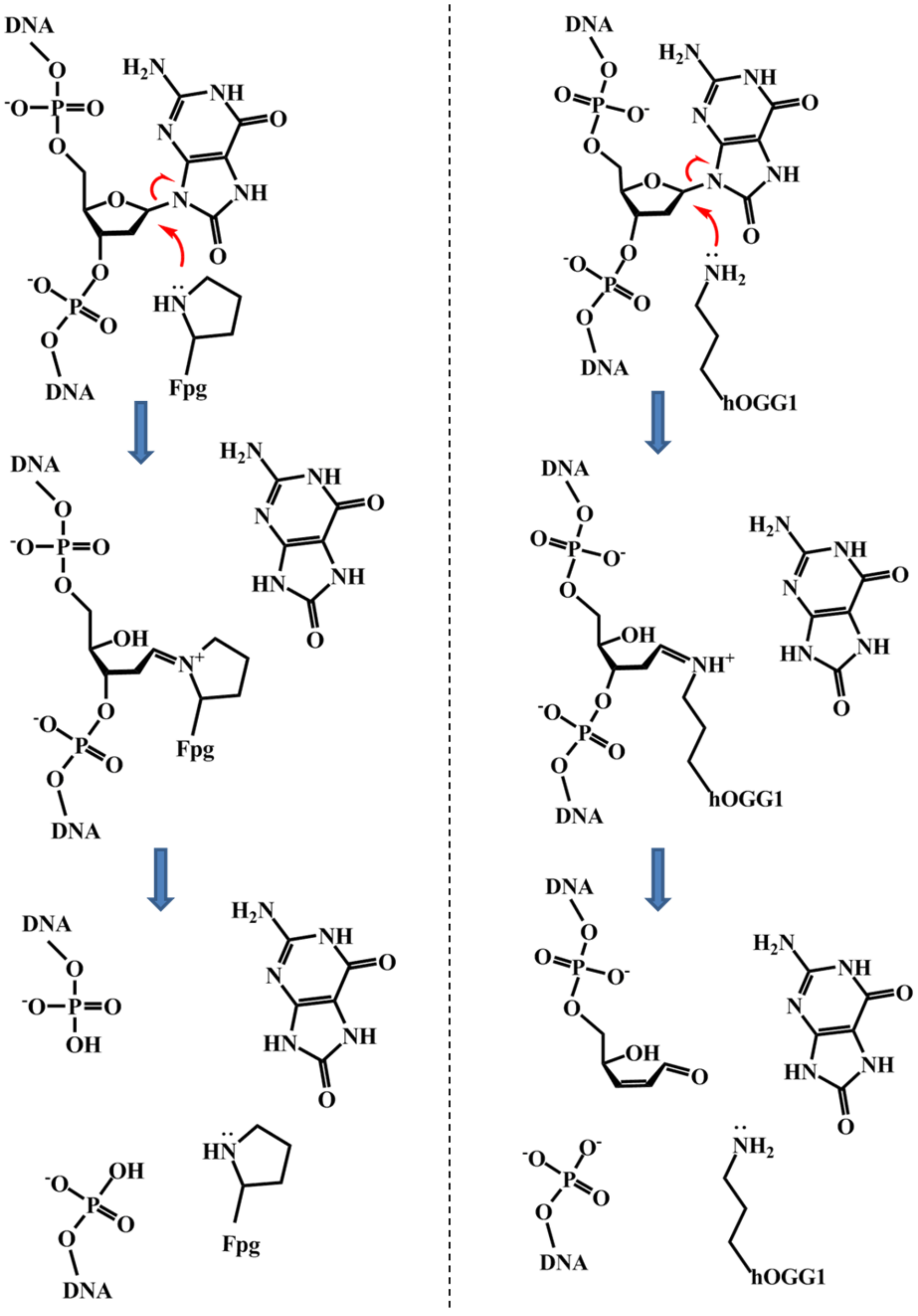
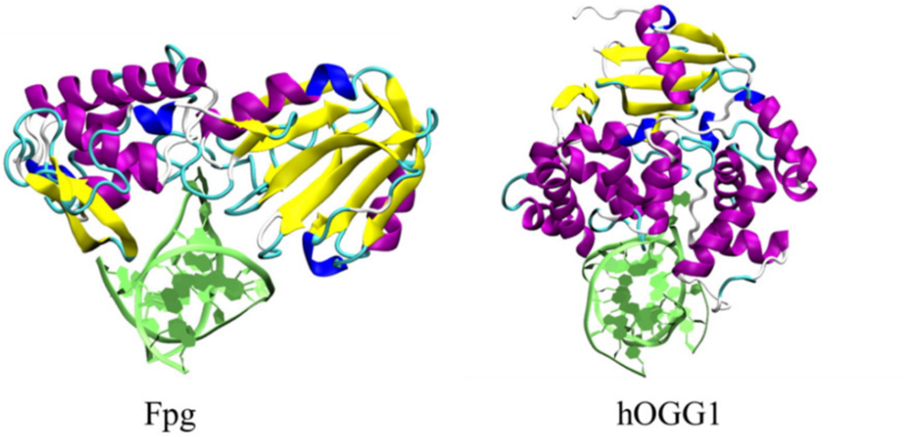


Figure 1-6. Overview of base excision repair catalyzed by Fpg (left) and hOGG1 (right). The first step shows the glycosylase activity and the second step shows the AP lyase activity.

Structures of Fpg and hOGG1 have been captured using crystallography. In general, Fpg contains around 270 residues and has two domains and a central hinge region where the catalytic site is located. The DNA binds to the central cleft, contacted by a conserved zinc finger and a helix-two turn-helix DNA binding motif. In contrast, hOGG1 contains 345 residues and has completely different fold: it contains mostly α helices contacting the DNA and some β sheets away from the DNA (Figure 1-7). Nevertheless, structures of catalytically inactive Fpg and hOGG1 bound with oxoG-containing DNA show certain similarity in handling the DNA (PDB ID: 1R2Y[29] and 1EBM[13]). Both Fpg and hOGG1 bend the DNA at the site of the target base pair, extrude oxoG into the active site and insert several residues into the void left by the flipped oxoG. The orphan cytosine is recognized by an arginine in both enzymes and the base pair 3' to the lesion is buckled by an intercalating aromatic residue (Phe113 in *Bst*-Fpg or Y203 in hOGG1). More importantly, in both enzymes the oxoG in the active site is recognized by a hydrogen bond made to its protonated N7 atom. Unlike oxoG, G does not process a protonated N7, therefore Fpg and hOGG1 can directly "read" the structural difference between oxoG and G by such hydrogen bond. The structure of Fpg bound with undamaged DNA has also been solved (PDB ID: 2F5O).[30] In this structure, a disulfide crosslink was engineered into the protein-DNA complex so as to capture this fleeting state. The undamaged G:C base pair being interrogated is not opened at all, but it is still buckled by the "wedge" residue Phe113 which intercalates into the duplex. The structure of hOGG1 interrogating undamaged DNA was at first captured using a proximal disulfide crosslink between hOGG1 and cytosine of the target pair (PDB ID: 1YQK[31]). The target G is thus forced by the crosslink to flip out from the DNA helix. The extrahelical G is not inserted into the active site like oxoG; instead, it stays at an exo-site adjacent to the active site (Figure 1-8). Another structure of hOGG1 bound with undamaged DNA (PDB ID: 3IH7) was captured using a disulfide crosslink at a site four base pairs away from the target pair, and thus the distal crosslink does not directly disrupt the target pair.[32] Surprisingly, the target G is not intrahelical; instead, the G is inserted into the active site of hOGG1, adopting a conformation nearly same as that of the extrahelical oxoG in 1EBM (Figure 1-8). Molecular dynamics simulation suggested that such insertion of G into the active site may result from suppression of structural fluctuations around the active site caused by the crosslink.[32] In this hOGG1-DNA complex, hOGG1 is fully functional in catalysis, but it fails to excise the G present in the active site. This finding suggests that hOGG1 has the ability to discriminate between oxoG and G at the level of catalysis. The mechanism of such catalytic checkpoint for oxoG is still unclear. Another structure of hOGG1 targeting an undamaged base pair is captured using proximal crosslink (PDB ID: 2I5W), in which the target G slightly flips out from the intrahelical position and forms a hydrogen bond to the N7 atom of the flanking oxoG.[33] This structure can be viewed as a tentative early intermediate in the base extrusion process, since the slight base pair opening is forced by the crosslink.

The crystal structures discussed above may have provided static pictures along the base eversion (or base flipping) pathway in Fpg and hOGG1. Recognition of oxoG may occur before, during or after base eversion. The direct oxoG recognition revealed so far occurs at the active site, where the protonated N7 of oxoG is contacted by the backbone carbonyl of S219 in Fpg or G42 in hOGG1.[13, 29] From the crystal structures, no direct interaction is seen made to recognize the O8 atom of oxoG, which is the other structural difference compared to G. It was suggested that a dipole-dipole interaction between K249/C253 and the O8/H7 atom pair of oxoG may play a role in oxoG recognition,[31] but such interactions are not necessary for oxoG eversion/excision, since the K249C/C253K double mutation does not substantially affect hOGG1's activity.[34]



Fpg

hOGG1

Figure 1-7. Comparison of the overall structures of Fpg (PDB ID: 1R2Y[29]) and hOGG1 (PDB ID: 1EBM[13]).

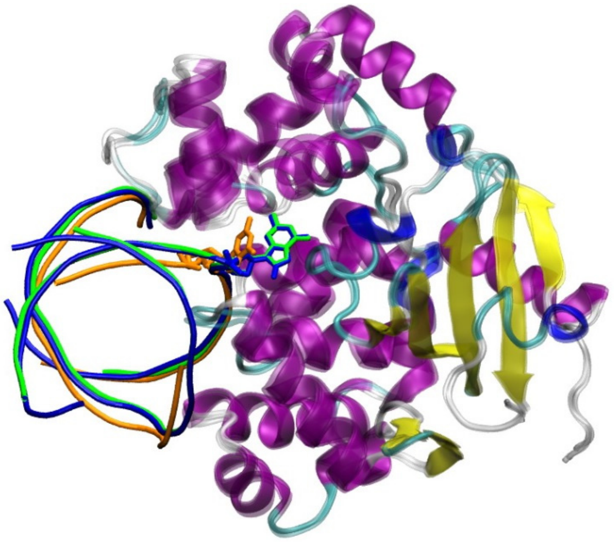
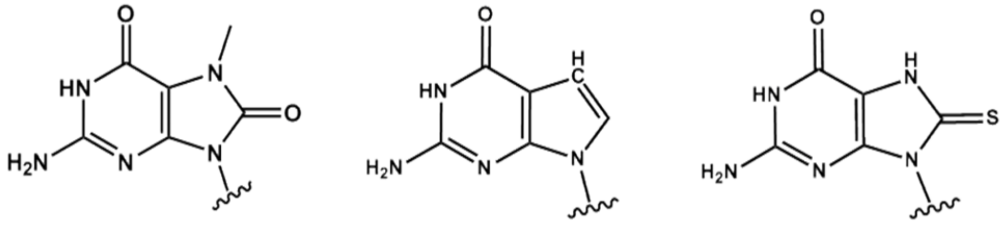


Figure 1-8. Comparison of the structures of 1EBM, 1YQK and 2I5W. The three structures were superimposed by the protein part. The DNA of 1EBM, 1YQK and 2I5W are colored in blue, green and orange, respectively, and the proteins are colored by secondary structure.

Hamm *et al* synthesized several oxoG analogs that lack either the protonated N7 or the O8 atom as compared to oxoG (Figure 1-9), and found that those oxoG analogs are all active substrates for Fpg and hOGG1, but with lower binding affinities as compared to oxoG.[35] Their findings suggest that Fpg and hOGG1 should contact both the protonated N7 and the O8 atoms, since substitution on either of these atoms result in lower binding affinity. The crystal structures of Fpg and hOGG1 have revealed direct contact to the protonated N7 of oxoG, but the structural details that show how Fpg and hOGG1 recognize the O8 atom are still missing. Those unrevealed contacts that contribute to oxoG recognition probably occur during early base eversion, which has high-energy and low-population states that are inaccessible by experimental approaches. In addition, high-speed imaging experiments showed that hOgg1 and Fpg slide along DNA in a rate that approaches the upper limit for one-dimensional diffusion.[36] Thus, it is unlikely that Fpg or hOGG1 extrudes every encountered base into the active site to examine DNA damage; instead, damage recognition should occur during early base eversion so that Fpg or hOGG1 does not waste time in fully opening the bases during rapid scanning.



7-methyl-8-oxo-2‘-guanine 7-deaza-2’-guanine 8-thio-2’-guanine

Figure 1-9. Synthesized oxoG analogs that are active substrates with Fpg and hOGG1.

Computer simulations have been used to explore the transient base eversion process. Qi *et al* found that eversion of oxoG through the minor groove is energetically more favorable than that of G.[37] However, our lab has calculated the full free energy landscape for oxoG eversion through the minor and major groove, which indicates that the major groove eversion encounters a significantly lower energy barrier.[38] Interestingly, major groove eversion has also been suggested for hOGG1, as the crystal structures have shown the intermediates along the tentative major groove pathway.[31, 33, 39, 40]

Overall, considerable effort has made to reveal the damage recognition mechanism of Fpg and hOGG1. There is a large amount of structural data that provides insights into certain states of the base eversion pathways, but the dynamic and energetic details of the pathways are still unclear, hindering our understanding of the damage recognition mechanism. In this work, molecular modeling is used to investigate the base eversion pathways in Fpg and hOGG1, so as to provide structural, dynamic and energetic basis of the damage recognition process.

1.2. Molecular Dynamics

Molecular dynamics (MD) simulations are widely used in structural biology.[41] MD has a unique advantage of connecting energy, structure and dynamics for conformational changes. By solving Newton's equations of motion for a system, MD produces trajectories of structures showing conformational changes over time. Such trajectories of atomic motions can then be viewed and analyzed using graphic software such as MOIL-view,[42] pymol[43] and VMD.[44] Currently, typical timescale for MD simulations is nanosecond or microsecond; but with advanced time-independent methods, MD can sample the dynamics of biological events that occur on much longer timescales. MD is particularly useful when obtaining structures of those low-populated intermediates which are usually not accessible by experiments. When used together complementarily, MD simulations and experimental methods, such as X-ray crystallography and nuclear magnetic resonance spectroscopy (NMR), can provide a complete picture of the structure and dynamics of biomolecules.

1.2.1. Basic Physics and Force Field

MD numerically solves the Newton's equations of motion to predict the motion of the atoms in a particular system.

$$\vec{F_i} = m_i \cdot \vec{a_i} \qquad (1\text{-}1)$$

$\vec{F_i}$ is the force exerted on atom i, mi is the mass of atom i, and $\vec{a_i}$ is the acceleration of i. $\vec{F_i}$ can also be expressed as the gradient of the potential energy:

$$\vec{F_i} = -\vec{\nabla_i} V \qquad (1\text{-}2)$$

Combining equations (1-1) and (1-2) and taking the second derivative of $\vec{a_i}$ yields:

$$-\vec{\nabla_i} V = m_i \cdot \frac{d^2 \vec{x_i}}{dt^2} \qquad (1\text{-}3)$$

Thus the derivative of the potential energy of the given system is related to the changes of atom position ($\vec{x_i}$) as a function of time (*t*).

$\vec{\nabla_i} V$ is determined by the force field, a potential energy function that describes the physical interactions of particles being simulated. The Amber force field is one of the most commonly used force fields and is of particular interest in this work. In the Amber force field, the potential energy $V(r^N)$ consists of bonded (bond, angles and dihedral) and non-bonded (electrostatic and van der Waals) terms, as shown in equation (1-4) below[45].

$$V(r^N) = \sum_{bonds} \frac{1}{2} K_b (b - b_0)^2 + \sum_{angles} \frac{1}{2} K_\theta (\theta - \theta_0)^2 + \sum_{torsions} \frac{1}{2} V_n [1 + \cos(n\omega - \gamma)] + \sum_{i<j}^{N} \varepsilon_{i,j} \left[\left(\frac{R_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{ij}}{r_{ij}} \right)^{6} \right] + \sum_{i<j}^{N} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \quad (1\text{-}4)$$

The first and second terms are harmonic potentials which correlate with deviation of the actual bond length (b) and actual bond angle (θ) to the reference length (b_0) and reference angle (θ_0), respectively. The third term is the torsional potential describing how the energy changes as the bond rotates. ω is the torsion angle and γ is the phase factor. The fourth term is the Lennard-Jones potential that represents the van der Waals interactions. R_{ij} is the distance at which the potential reaches its minimum, $\varepsilon_{i,j}$ is the well depth, and r_{ij} is the distance between the atoms. The fifth term is the Coulomb potential that represents the non-bonded electrostatic interactions between a pair of charges (q_i and q_j).

The protein force field used in this work is ff99SB[46], which was developed in the Simmerling lab. This force field has been shown to be highly accurate in reproducing experimental data.[47, 48] The most recent improvements to the Amber force field for nucleic acids, the parmbsc0 corrections for the alpha and gamma backbone dihedrals,[49] are also employed in this work.

1.3. Advanced Simulation Methods

One of the key challenges of MD is the timescale problem. The typical accessible timescale for conventional MD, ranging from femtosecond to microsecond, cannot cover many of the biologically interesting processes occurring on a larger timescale. For example, eversion of oxoG in Fpg and hOGG1 occurs on a millisecond timescale.[27, 28] Enhanced conformational sampling techniques, which can capture large-scale and long-time conformational changes, are thus needed to investigate those biological problems. In this work, we used the partial nudged elastic band method to find the minimum energy pathways for eversion of oxoG and G, and used umbrella sampling to calculate the free energy along the resulting pathways. This methodology has been successfully applied to investigate the energetic preference of oxoG eversion through the major and minor grooves in Fpg.[38]

1.3.1. Partial Nudged Elastic Band

Partial nudged elastic band[50] (PNEB) is method that finds the minimum energy pathway between two given endpoint structures that define the start state and the end state. Intermediate states between the two endpoints can be used to seed the path. Multiple copies (known as "beads") of the endpoints and seeded intermediates are created, and all these beads are then linked to form a continuous path. In PNEB, the two endpoint beads are fixed, and the other beads are pulled into an initial interpolating path which is subjected to optimization. Spring forces were applied to the beads and force each bead to maintain at an average separation between its neighboring beads, and the forces from the molecular potential are projected out in the perpendicular direction to the tangent of the path.

The PNEB implementation in the AMBER package[51] can allow users to defined subset of the system to which the nudged elastic band forces are applied. Thus, this method is particularly useful for investigating local conformational changes in large, complicated systems, or in explicitly solvated systems.

One of the limitations of PNEB is that the final path depends on the initial guess, which is usually the most direct path. Sometimes the resulting path developed from the most direct path is not the dominant path, so multiple PNEB runs with different initial guesses, created by adding one or more seeding intermediates, may be needed to find the desired path.

1.3.2. Umbrella Sampling

Umbrella sampling is a technique which can improve sampling of a system where ergodicity is hindered by high energy barriers on the free energy surface.[52] The method uses MD simulations to determine the probability distribution for the system along a reaction pathway, and then the potential of mean force (PMF), or the free energy landscape can be determined by inverting the Boltzmann distribution.[53-55] In umbrella sampling simulations, the transition pathway of interest is described by a reaction coordinate, and is divided into many small regions (also known as "windows") along the reaction coordinate. A harmonic biasing potential is applied to each window and it confines the distribution of the system to a desired region, which overlaps with the distribution of the neighboring windows. Thus, the transition pathway is fully sampled along the reaction coordinate. The Weighted Histogram Analysis Method (WHAM)[54, 55] is used to remove the bias of each window and calculate the free energy based on the resulting unbiased probability distribution.

$$\boldsymbol{A(x) = -k_B T \ln P'(x) - U'(x)} \qquad (1\text{-}5)$$

Equation (1-5) describes the calculation of unbiased free energy (*A(x)*) using the biased population (*P'(x)*) and the biasing potential (*U'(x)*). k_B is the Boltzmann constant, and T is the system temperature.

To successfully perform umbrella sampling simulations, several parameters need careful consideration. One of them is the reaction coordinate. The reaction coordinate should well represent the transition pathway of interest, since the free energy of the system is the projected on the reaction coordinate. Sometimes sole reaction coordinate may not be sufficient to describe the complicated conformational transitions; in such cases, multiple dimensions may be used in umbrella sampling.[37, 38, 56] The other parameters that should be carefully chosen are the window size (or the difference between the reference values of the neighboring windows) and the force constant of the biased potential. Too small a windows size means too many windows, which will consume unnecessary computational time. Choosing a windows size that is too large may cause error in PMF calculation because the distribution from neighboring windows may not overlap. Discontinuity between windows may also be the result of biasing potentials that are too strong, which confine the windows in a small region along the reaction coordinate. Also, the biasing potentials should be strong enough to force the simulation to sample those high-energy regions.

One of the limitations of umbrella sampling is that the accuracy of the PMF strongly depends on the chosen reaction coordinate(s). For pathways of complicated conformational changes, the reaction coordinate may not well represent the desire path or the system may drift towards the orthogonal degrees of freedom, creating discontinuity on the pathway. Using more reaction coordinates to restrain the system on the desired path may be a solution to these problems, but in such cases the required computation time is significantly increased.

1.4. Overview of Research Projects

1.4.1. Structural and energetic basis for 8-oxoguanine recognition by Fpg

8-oxoguanine (oxoG) is one of the most prevalent DNA lesions and it is generated from oxidation of guanine (G). Formamidopyrimidine glycosylase (Fpg) scans oxoG from a sea of undamaged bases by fast sliding along the DNA, and extrudes oxoG into the extrahelical active site where the oxoG is excised. Since it is impractical for Fpg to fully extrude every encountered base into the active site for damage recognition, it is hypothesized that damage recognition occurs not only at the active site, but also at the early stages of the base eversion pathway, which are difficult to probe using experiments. We thus used molecular dynamics simulations to explore the base eversion process of oxoG and G in Fpg and focused on the structural and energetic details for the mechanism of early oxoG recognition. We found that Fpg selectively facilitates eversion of oxoG by stabilizing the several intermediate states along eversion pathway, in which the eversion of G is hindered by unfavorable interactions. Our findings suggest that recognition of oxoG indeed occurs during early base eversion, and we also identified important residues that are responsible for recognition of oxoG, supported by biochemical analysis of site-directed Fpg mutants disturbing the interactions in the intermediates.

1.4.2. How do Fpg and hOGG1 achieve same function with complete different structures?

Human 8-oxoguanine DNA N-glycosylase 1 (hOGG1) is a functional homolog of Fpg but they are not structurally related. These two enzymes have similar strategies to process the oxoG damage: they search for oxoG by rapid diffusion along DNA, kink the DNA at the target base and extrude oxoG to the active site where the catalysis occurs. However, structural and energetic details for oxoG recognition during base flipping are still unknown. Here we used MD simulations to reveal the mechanism of oxoG recognition by hOGG1 and compared it to that by Fpg, and found that hOGG1, like Fpg, has several checkpoints for oxoG along the eversion pathway, and the geometries of those checkpoints are remarkably similar between the two enzymes, suggesting they employ similar strategies to efficiently detect oxoG from a sea of undamaged bases.

1.4.3. Mutant studies reveal the role of Fpg's aromatic wedge in damage recognition

Inserting a bulky wedge residue into the DNA to interrupt the base stacking around the target base and promotes base eversion may be a common feature of DNA glycosylases.[57] Previously we have used MD simulations and stopped-flow kinetics experiments to investigate the oxoG recognition by wild type Fpg and its mutants with substitution for the wedge residue F113, and found that Fpg actively uses Phe113 to destabilize the intrahelical states of oxoG and G eversion (Kuznetsov *et al*, in preparation). Here we used a new reaction coordinate that allows us to explore the free energy surface of the full base eversion pathway, and found that the F113A mutation affects the intrahelical state and a late extrahelical state. Our data suggest that the wedge residue Phe113 may play a role in oxoG recognition by destabilizing the intrahelical state to promote oxoG eversion and by stabilizing the extrahelical state to facilitate further eversion of oxoG.

2. Structural and energetic basis for 8-oxoguanine recognition by Fpg

Acknowledgments

The material presented in this chapter contains direct excerpts from a manuscript that is in preparation by Haoquan Li, Christina Bergonzo, Arthur J. Campbell, Anton V. Endutkin , Fu Lin, Carlos de los Santos, Dmitry O. Zharkov, Arthur Grollman and Carlos Simmerling. The introduction and conclusion sections were written by Arthur J. Campbell, and revised by Christina Bergonzo and Haoquan Li, with suggestions from Carlos Simmerling and Dmitry O. Zharkov. The method section and the results and discussion section were written by Haoquan Li and Dmitry O. Zharkov with suggestions from Carlos Simmerling. Haoquan Li performed the simulations and data analysis, which are based on the work flow developed by Arthur J. Campbell. Anton V. Endutkin performed the biochemical experiments.

2.1. Introduction

Oxidative DNA damage is one of the most common types of damage known to affect the genome.[58, 59] The main source of oxidative DNA damage is from reactive oxygen species (ROS), which arise due to endogenous and exogenous factors, such as aerobic metabolism and exposure to ionizing and photosensitized UV radiation, respectively.[60] One common product of ROS of particular interest is oxoG, which is formed by oxidation of a normal guanine base at the C8 position, as shown in Figure 2-1A.[61] The oxoG lesion remains structurally similar to a normal guanine (G) nucleobase, with a two atom difference on the major groove face in B-form duplex

DNA: oxoG has an O8 and an H7 on the face, G has an H8 and a lone pair on N7, altering the electrostatics of oxoG compared to G. If the oxoG lesion remains uncorrected, it tends to adopt a Hoogsteen orientation and guide incorporation of adenine as a base pairing partner during replication (Figure 2-1B), promoting a G:C to T:A transversion mutation which has been linked with cancers and various age related diseases.[58, 62,63]

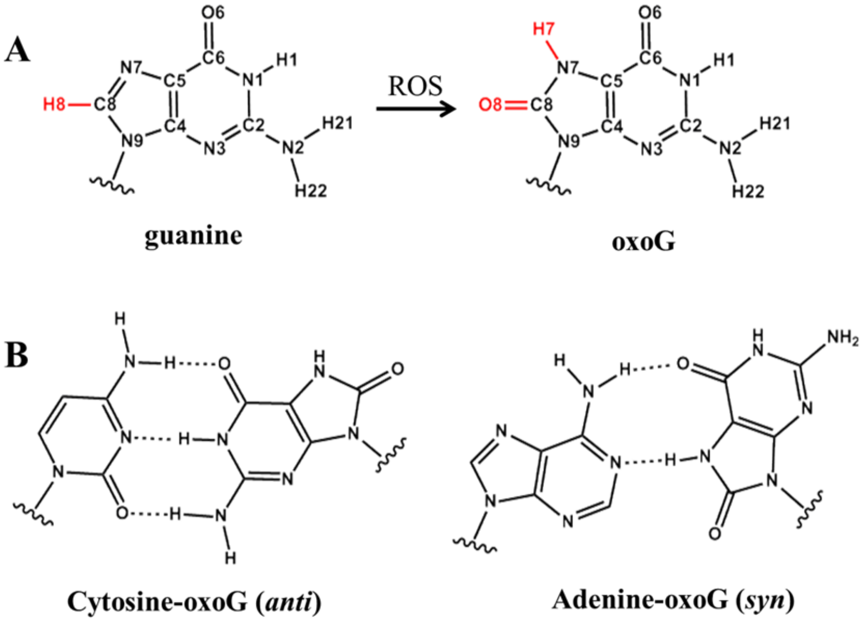


Figure 2-1. A, chemical structure of guanine and oxoG. The structural difference between the two bases is highlighted in red. B, Watson-Crick paring of oxoG:C and Hoogsteen pairing of oxoG:A.

Base excision repair (BER) is a cellular mechanism that repairs various types of DNA damage.[64] In human cells, fewer than a dozen BER glycosylases usually process 100 to 10,000 lesions per cell per day.[3, 65, 66] These glycosylases share similar structural features with their bacterial counterparts, including an intercalating aromatic/aliphatic wedge that enters through the minor groove, an overall bending/kinking of the DNA upon binding, and a base eversion process that repositions the damaged nucleobase into an extrahelical protein active site for excision.[67]

The BER DNA glycosylases that target oxoG as one of their substrates in human and bacterial cells are hOGG1 and Fpg respectively. These glycosylases are functional analogs that excise the oxoG lesion opposite a cytosine in duplex DNA. The catalytic mechanisms of oxoG excision by hOGG1 and Fpg have been well characterized.[9, 10, 13, 23, 68, 69] Both enzymes carry out nucleophilic attack on the C1' atom of the deoxyribose sugar of oxoG, initiating base excision, and then create a nick on the DNA backbone.[13, 23] Although hOGG1 and Fpg have different tertiary structures, they contain analogous residues proposed to perform similar functions during recognition and base eversion. Both enzymes have aromatic intercalating wedges (Tyr203 and Phe113 in hOGG1 and *B. st* Fpg, respectively) that enter through the minor groove at the site of damage.[13, 30] Additionally, arginine sidechains (204 for hOGG1 and 111 for *B. st* Fpg) intercalate through the minor groove once damage is bound in the active site, recognizing the proximal hydrogen bond acceptor atoms N3 and O2 of the cytosine (C) base paired to the oxoG.[13, 70] The set of hydrogen bonds that form between the arginine residue(s) of each enzyme and the orphan base is specific for a cytosine, and is disrupted by any other nucleobase in this position, thus contributing to a biologically relevant substrate specificity of hOGG1 and Fpg.

Previous crystallographic work has been able to provide putative snapshots along the eversion process.[29, 30, 37, 71-73] Of particular interest are two crystallographic structures of *Bacillus stearothermophilus* (*B. st*) Fpg bound to DNA that define two states: one in which an undamaged G nucleotide is intrahelical (PDB 2F5O),[30] and another where a damaged oxoG is trapped in an extrahelical conformation in the active site of a catalytically impaired mutant enzyme (PDB 1R2Y).[74] Although these two endpoint snapshots are useful for understanding two biologically relevant states, they give an incomplete picture of how Fpg discriminates against G in favor of oxoG during a dynamic process of eversion, thus hindering our understanding of the damage recognition mechanism. Using fluorescence stopped-flow experiments, we have earlier shown that G is not rejected immediately while remaining in an intrahelical conformation but at least partially everted (Kuznetsov *et al*, in preparation). We hypothesize that base eversion is needed for oxoG recognition, and at least one or more recognition steps occur at the early stages of base eversion, since the extremely rapid sliding of Fpg along DNA strongly suggests that Fpg does not fully evert every encountered base to the active site.[36] To test this hypothesis, we investigated the eversion pathway for oxoG and G so as to understand how recognition occurs. Molecular dynamics simulations were used to model the low-populated higher-energy states of base eversion whose structures and dynamics are inaccessible by experimental methods. The free energy profiles of eversion from an intrahelical, unopened conformation to an extrahelical conformation, where the target base is bound by the active site loop in its pre-excision complex, are calculated. Fpg's mechanism for discrimination along the eversion pathways can be understood by structural analysis, energy decomposition, as well as biochemical mutation analysis of residues critical for oxoG recognition.

2.2. Methods

The Amber 11 and 12 suites of programs were used for all calculations in this work.[51, 75] The ff99SB force field[46] was used with the parmbsc0 DNA backbone parameters for all systems.[49] The parameters for oxoG were based on the parameters developed by Miller et al.[76]

2.2.1. System preparation and equilibration

The endpoint structures of the base eversion pathway were generated from crystal structure 2F5O[30] (intrahelical endpoint) and 1R2Y[29] (extrahelical endpoint) in the same way as in our previous work.[38] Previously we set the catalytic proline (Pro1) neutral because we were interested in investigating the scenario directly preceding the catalytic reaction, in which the Pro1 needs to be deprotonated to act as a nucleophile.[77] In this work, we focus on the early stages of base eversion process which are probably prior to the proton rearrangement events at the active site. We used the H++ program, which predicts pK_a based on the standard continuum solvent methodology and produces results comparable to the experimentally determined values,[78] to estimate the pK_a of Pro1 in the equilibrated intrahelical structure, and found that Pro1 has a pK_a of 9.3. Thus, here we set Pro1 to be positively changed. The protonation state of other residues, the sequences and initial coordinates of the endpoint structures are same as in our previous work.[38]

The endpoint structures were solvated in a truncated octahedron with ~12 Å buffer. Each of the solvated structures was minimized and equilibrated in five steps: (i) a 10000-step minimization with positional restraint on the heavy atoms; (ii) a 100ps MD simulation in which the system was heated linearly to the target temperature of 330K, which reflects the biological temperature of *Bacillus stearothermophilus*, while the heavy atoms of the complex were restrained; (iii) a 100ps and a 250ps MD simulations with the heavy atoms of the complex fixed by 100 and 10 kcal/mol/Å^2 restraints, respectively; (iv) a 100ps, a 200ps and a 250ps MD simulations with the heavy atoms of the protein and DNA backbones fixed by 10, 1 and 0.1 kcal/mol/Å^2 restraints, respectively; (v) a final 2ns unrestrained simulation. During minimization and equilibration SHAKE was employed to constrain bonds involving hydrogen atoms[79], and a 1 fs time step was used. The particle mesh Ewald method[80, 81] was used to calculate Columbic interactions with a nonbonded cutoff of 8Å. Through step (iii) to (v) constant temperature of 330K and constant pressure of 1 atm were maintained by the weak-coupling algorithm.[82]

2.2.2. Base eversion pathway generation using PNEB

The protocol for running PNEB simulations was adopted from the previous work sampling oxoG eversion in Fpg through the major and minor grooves.[38] The initial temperature was set to 330K and the spring force constant was 2 kcal/mol/Å^2 for the first 100ps path optimization, then the spring constant was increased to 20 kcal/mol/Å^2 for the following 500ps. The systems were gradually heated to 380K over the next 100ps, and the temperature was kept over the next 200ps, and then was decreased back to 330K over the next 100ps. The final production runs were then performed over 500ps. During the annealing and the production steps the spring forces were set to 50 kcal/mol/Å^2.

2.2.3. Reaction coordinate

Previously our lab has developed a reaction coordinate, the modified COM pseudodihedral angle (CPDb), to describe base eversion.[83] CPDb was then used in PMF calculation for oxoG eversion in Fpg.[38] In that work, since rotation around the glycosidic bond was important for base eversion through the minor groove as well as for base entering the active site through the major groove, and CPDb was not sensitive to such rotation, a second reaction coordinate, the glycosidic torsion angle, was needed in the PMF calculation.[38] In this work, we focus on eversion through the major groove only, and we were able to develop a modified CPDb (Figure 2-2, hereafter referred to as the eversion angle) which is sensitive to the glycosidic rotation of the everted base at the active site. Figure 2-3 shows the comparison of CPDb and the newly developed eversion angle in describing the PNEB path of oxoG eversion. In the region near the extrahelical endpoint, CPDb remains ~270° while the glycosidic angle changes from ~0° to ~70°, indicating that CPDb cannot describe the final 70° rotation of the glycosidic bond. On the other hand, the glycosidic rotation from ~0° to ~30° near the endpoint, representing the process of the base entering the active site, correlates with the change of the eversion angle from ~250° to ~270°, which is an improvement to CPDb. The eversion angle cannot separate the glycosidic rotation from ~30° to ~70° either, but such rotation is probably just normal fluctuation of the base in the active site, since our lab has calculated the PMF for 360° rotation of the glycosidic bond of oxoG in Fpg's active site, and found that the global minimum is at ~55° and the energy is roughly the same from ~30° through ~70°.[84] Although the eversion angle is not sensitive to the base fluctuation in the active site, it can still describe the base eversion path up to the base enters the active site, and thus was used as the reaction coordinate for our PMF calculations.

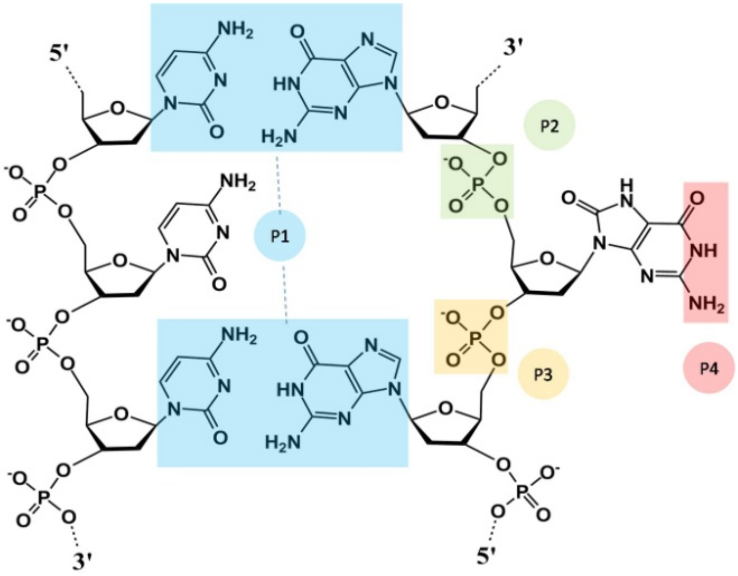


Figure 2-2. Definition of the eversion angle. The four points that define the eversion angle are boxed in different colors: P1, the center of mass (COM) of the four bases flanking the target base pair; P2, the COM of the 5' phosphate group; P3, the COM of the 3' phosphate group; P4, the COM of the Watson-Crick edge.

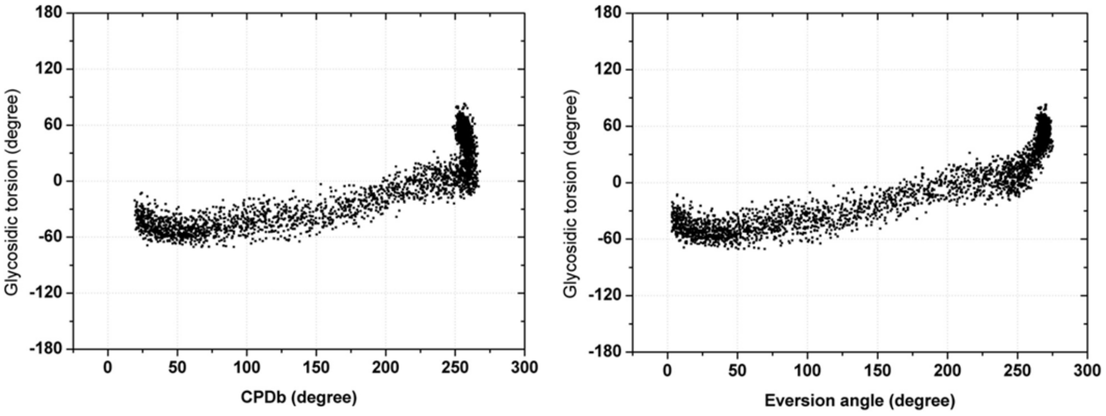


Figure 2-3. Glycosidic torsion angle as a function of CPDb (left) and eversion angle (right), showing these two reaction coordinates' sensitivity to the rotation of the glycosidic bond. Data points are calculated from the NEB production trajectories of the oxoG system.

2.2.4. Umbrella sampling and PMF calculation

Umbrella sampling (US) was used to obtain the PMF as a function of the eversion angle. The US protocol was similar to that in the previous studies sampling base eversion.[38, 83] For each WT system, two completely different sets of initial structures were taken from the PNEB production trajectories and US was performed independently for 2.5ns for each set. In each US run, 64 windows were evenly spaced along the eversion angle with ~4° intervals, and were restrained by 0.183 kcal/mol/deg^2 umbrella potential along the eversion distance in an NVT ensemble. The time step was 2 fs and the temperature was maintained at 330K using a Langevin thermostat[85] with a 75.0 ps^{-1} collision frequency. The eversion angle values were recorded every time step and were then analyzed using the weighted histogram analysis method (WHAM).[55] The error bars were then calculated from the difference of the two independent runs. Since the choice of zero point in free energy is arbitrary, the intrahelical endpoint is set to 0 kcal/mol in the PMF to highlight the difference along the path.

One of the limitations of PMF calculation using umbrella sampling is that the accuracy of the free energy strongly depends on the choice of reaction coordinate. Due to the complexity of the base eversion process, it is unlikely to find a reaction coordinate which can perfectly describe all the possible conformational changes along the pathway. The system has multiple degrees of freedom other than the one along the reaction coordinate, and thus if the system drifts in certain windows, it may cause discontinuity along the desired pathway. To test whether we have this problem, we calculated the RMSD of the target base during umbrella sampling, referenced to the initial structure of each window. As shown in Figure 2-4, we found that the majority of the windows have the RMSD below 2.5Å, a value comparable to that of the normal fluctuation of the intrahelical oxoG during equilibration. Several windows have higher RMSD (~4 Å), which indicates possible drifting of structure and may cause inaccuracy in PMF calculation. In future work, using multiple reaction coordinates may help to reduce such structural drifting.

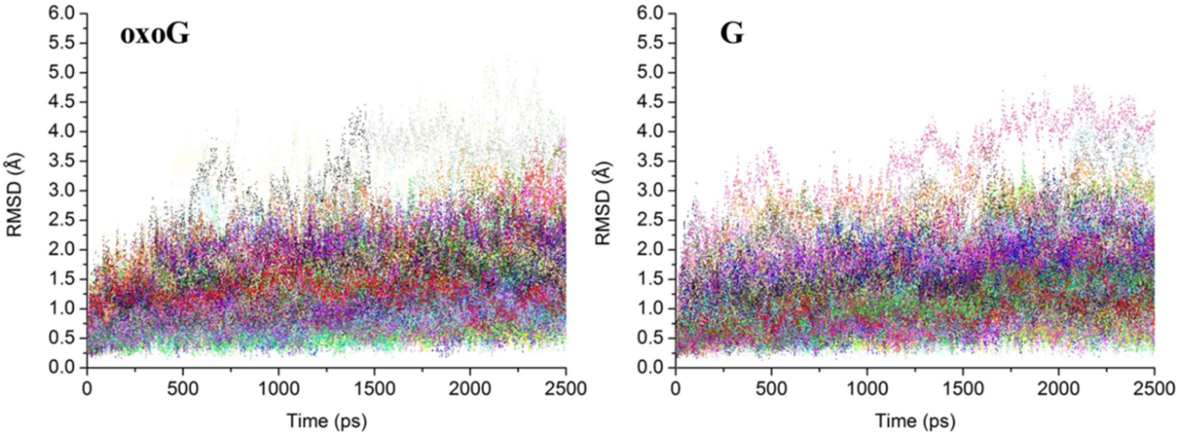


Figure 2-4. RMSD of the oxoG/G base in WT Fpg during umbrella sampling, referenced to the starting structure of each window. Structures were fit to the protein backbone. All of the 128 windows of each system were included.

2.2.5. Analysis of computational data and sequence conservation

To explore the structures of the oxoG and the G systems along the base eversion pathways, the umbrella sampling trajectories of these two systems were visualized using VMD.[44] Important residue to residue interactions along the base eversion pathways were then identified. To investigate the interactions, distances and nonbonded energies that can describe the interactions were calculated from the combined umbrella sampling trajectories. The electrostatic and van der Waals interaction terms were independently calculated and combined to give the total nonbonded interaction energy. For each system, the eversion angle and variables (distances and nonbonded energies) were analyzed for each picosecond from their respective umbrella sampling simulations. Each average variable data point was calculated for every 5° bin of eversion angle, and the error bars were calculated from the difference of the two independent umbrella sampling runs. The structures shown here were picked from the US trajectories and are representative structures exhibiting the important interactions. The sequence conservation chart was generated using Weblogo.[86, 87]

2.2.6. Biochemical assays of the Fpg mutants

The variants of *E. coli* Fpg harboring mutations of the residues Arg108, Asn168, and Arg258 (corresponding to Arg111, Asn173, and Arg263 of *B. st* Fpg) were produced in pET-

28a(+) plasmid using QuikChange Lightning site-directed mutagenesis kit (Agilent Technologies, Santa Clara, CA). Mutant and wild-type Fpg were expressed and purified as described.[88] Activity assays were performed on the 5′-^{32}P-labeled duplex oligonucleotide substrate with the damaged strand of the sequence 5′-CTCTCCCTTCXCTCCTTTCCTCT-3′ (X = oxoG or AP site) and the opposite strand fully complementary and bearing C opposite to the lesion. To obtain the AP substrate, 10 pmol of the duplex containing uracil in the required position was treated with 1 U *E. coli* uracil–DNA glycosylase (New England Biolabs, Ipswich, MA) immediately before reaction with Fpg. The fraction of the active enzyme in the preparations of wild-type Fpg and all mutants was determined by $NaBH_4$ trapping with the AP substrate as described[27] and was between 10% and 85% depending on the preparation. In the kinetic experiments, the reaction mixture contained 50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 1 mM ethylenediaminetetraacetate, 1 mM dithiothreitol, 2–100 nM oxoG or AP substrate and 1 nM (for oxoG substrate) or 0.2 nM (for AP substrate) wild-type or mutant Fpg (active form). The reaction was allowed to proceed for 10 min at 30°C and terminated by adding an equal volume of 95% formamide/20 mM EDTA dye and heating at 95°C for 3 min (oxoG substrate) or 1 min (AP substrate). The products were separated by electrophoresis in 20% polyacrylamide gel containing 8 M urea and quantified using Molecular Imager FX system (Bio-Rad, Hercules, CA). The K_M and k_{cat} values were determined from 3–5 independent experiments by non-linear regression using SigmaPlot v9.0 (SPSS, Chicago, IL).

2.3. Results and discussion

2.3.1. Simulation strategy

As discussed in introduction, damage recognition may occur along the base eversion path. To understand how Fpg discriminates against G in favor of oxoG, we computationally compared the eversion process of G and oxoG in duplex DNA bound by Fpg. Since base eversion in Fpg occurs in a millisecond timescale and is unlikely to be sampled by conventional MD simulations,[27] in this work, PNEB and umbrella sampling, which are time-independent methods, were used to structurally and energetically characterize the base eversion process. The same methods has been successfully applied to calculating the energetic preference of base eversion through the minor or the major groove in Fpg.[38] Here the base eversion pathways from the unopened position (the intrahelical endpoint) to the active site of Fpg (the extrahelical endpoint) are generated using PNEB, followed by PMF calculation using umbrella sampling and WHAM.[55] By investigating the structural and energetic differences between eversion of oxoG and G, we can then understand how oxoG is distinguished from G by Fpg.

2.3.2. Energy profiles of oxoG and G eversion show differences in free energy

As shown in the free energy profiles (Figure 2-5), the overall energetic difference between the endpoints of both pathways (Figure 2-5) is consistent with the two crystallographic structures 2F5O[30] and 1R2Y[74], in which G favors the intrahelical state (2F5O) and oxoG, the extrahelical state (1R2Y). The free energy profiles also indicate that eversion of oxoG is kinetically more favorable than that of G, because the energy barrier to eversion for G (~22 kcal/mol) is much higher than for oxoG (~7 kcal/mol). The energy barrier for oxoG calculated in this work is also lower than that of the minor groove path for oxoG in previous calculations[38] and thus the major groove path would have faster rate. Specifically, G favors the intrahelical position by ~20 kcal/mol with respect to its extrahelical state, whereas for oxoG, the extrahelical position is 7 kcal/mol lower in energy than its intrahelical position (Figure 2-5). The free energy profiles of the two pathways show significant differences in multiple stages, indicating several potential oxoG-recognizing steps along the base eversion pathway, through which we may understand the mechanism of damage recognition. The details of the potential oxoG-recognizing steps are discussed below, and for convenience they are referred to as Step I~IV (labeled in Figure 2-5).

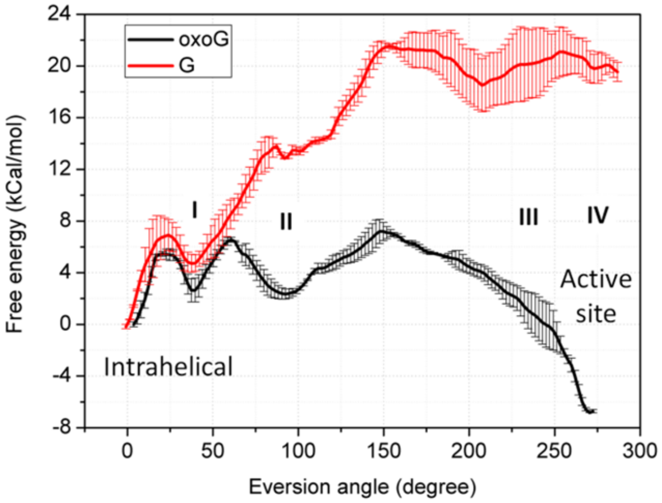


Figure 2-5. Comparison of free energy profiles for eversion of oxoG (black) and G (red). Four potential oxoG recognition steps are labeled I, II, III and IV. The error bars reflect the difference between two independent US runs.

2.3.3. Early discrimination between oxoG and G: Steps I and II

After both nucleobases leave the intrahelical space and enter the major groove eversion pathway, they reach a metastable state (Step I), indicated by the energy minimum at the eversion angle of ~40° of the PMF (Figure 2-5, with structures shown in Figure 2-6). When G reaches the metastable state it forms two key interactions with the residues on the zinc finger beta hairpin: first, a cation-pi stacking interaction with Arg263, a strictly conserved residue in all Fpg (Figure 2-7); second, a hydrogen bond between the N7 atom on the imidazole ring of G and the backbone amide H of Gly264 (Figure 2-6A), which is not a conserved residue (Figure 2-7), suggesting this interaction is not imperative for Fpg's function. For oxoG, the N7 atom is protonated, precluding the formation of a hydrogen bond to Gly264 as seen in the G system. To show the oxoG/G-Gly264 interaction along the base eversion pathway instead of in a single structural snapshot, we measured the averaged distances between the backbone N atom of Gly264 and the N7 of oxoG/G during umbrella sampling, and found that this distance is indeed significantly larger for oxoG than for G (Figure 2-8A). The cation-pi stacking between Arg263 and G does not occur for oxoG either, as indicated by the lower nonbonded energy of Arg263-oxoG compared to that of Arg263-G (Figure 2-8B). This is probably because oxoG moves away from Arg263 to form a hydrogen bond between the protonated N7 of oxoG and the second 5′ phosphate (hereafter referred to as p^1, Figure 2-6B). Since the N7 of G is deprotonated, the N7-p^1 hydrogen bond is not seen in the G system. Figure 2-9 clearly indicates that the N7-p^1 hydrogen bond in oxoG is formed at the eversion angle of ~40°, whereas in the G system the N7-p^1 distance is much larger. It is possible that the repulsive N7-p^1 interaction in the G system pushes G towards Arg263 and Gly264.

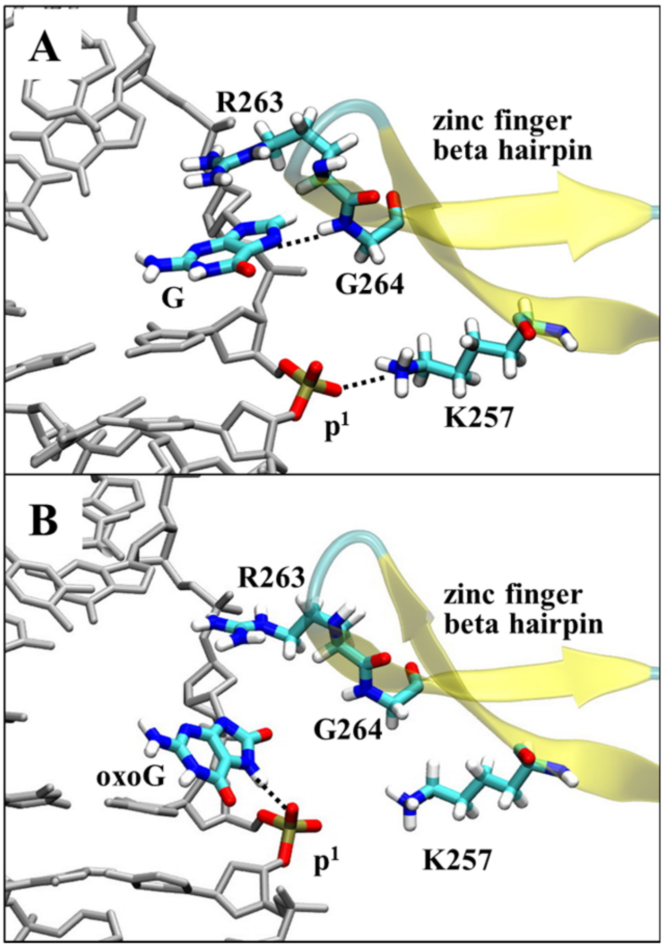


Figure 2-6. Comparison of the G system (A) and the oxoG system (B) structures at the eversion angle of ~40° (Step I). Hydrogen bonds are depicted as dotted lines.

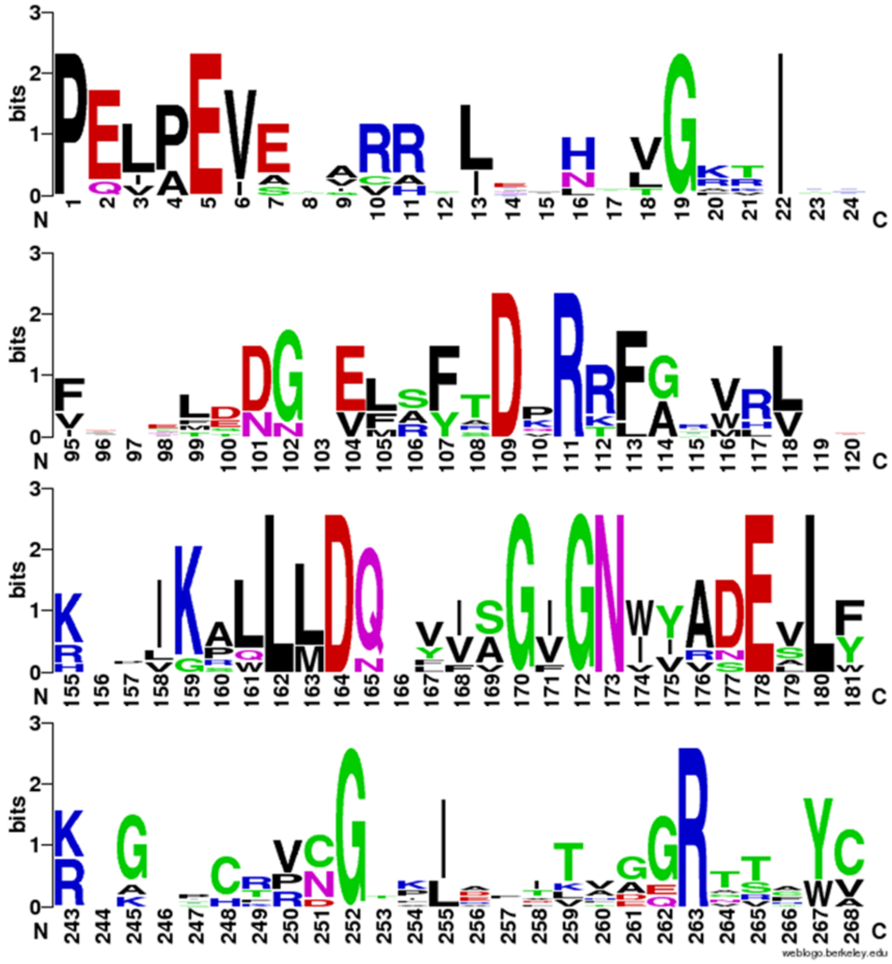


Figure 2-7. Sequence conservation around Asn173 among *B. st* Fpg (gi|38492995), *E. coli* Fpg (gi|15804176), *Mycobacterium tuberculosis* Fpg (gi|148662769), *Arabidopsis thalania* Fpg (gi|18404050), *Oryza sativa* Fpg (gi|115469160), *Candida albicans* Fpg (gi|3850130), and *Neurospora crassa* Fpg (gi|157072070).

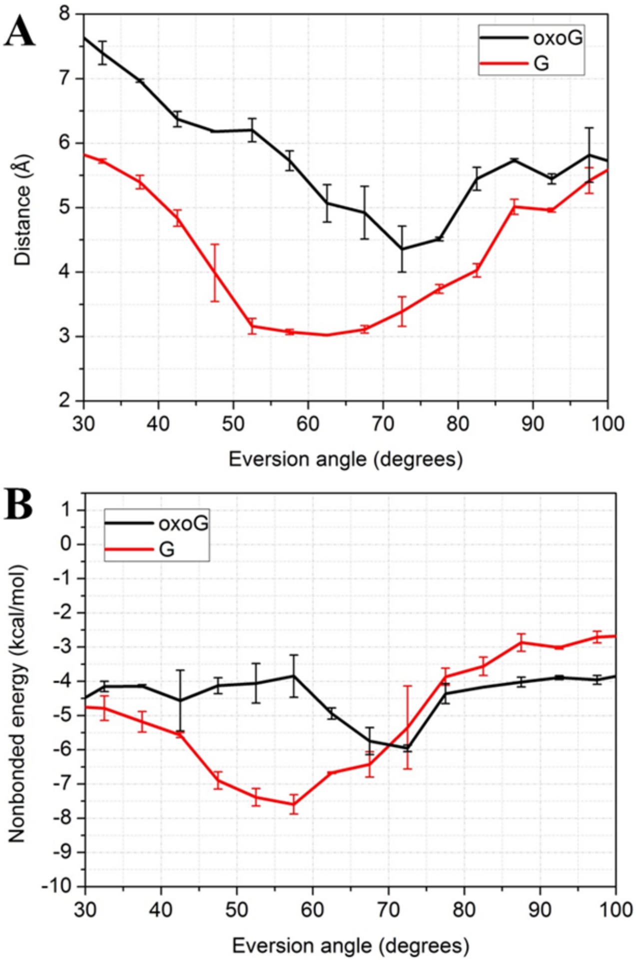


Figure 2-8. A, comparison of distances between the backbone N atom of Gly264 and the N7 atom of oxoG/G (black/red). The G-Gly264 hydrogen bond was formed at the eversion angle of 50°–75°, while oxoG is further away from Gly264. B, comparison of pairwise nonbonded energies between the base moiety of oxoG/G (black/red) and Arg263. The error bars reflect the difference between two independent US runs.

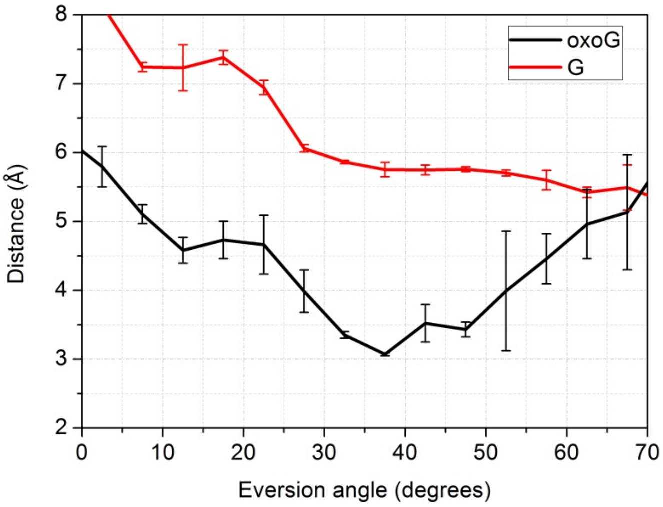


Figure 2-9. Comparison of distances between the O2P atom of p[1] and the N7 atom of oxoG (black) and G (red). The oxoG- p[1] hydrogen bond was formed at the eversion angle of ~40° (Step I). The error bars reflect the difference between two independent US runs.

After Step I, oxoG or G needs to go through the gap between the DNA backbone and the zinc finger hairpin (Figure 2-10), and this process corresponds to the eversion angle of 60~90°. The PMF of G at this transition region shows significant difference from that of oxoG (Figure 2-5). To go through, G has to overcome a 10 kcal/mol energy barrier (compared to ~4 kcal/mol for oxoG), suggesting that the transition is extremely unfavorable for G. Two main factors may contribute to this high barrier, as discussed below.

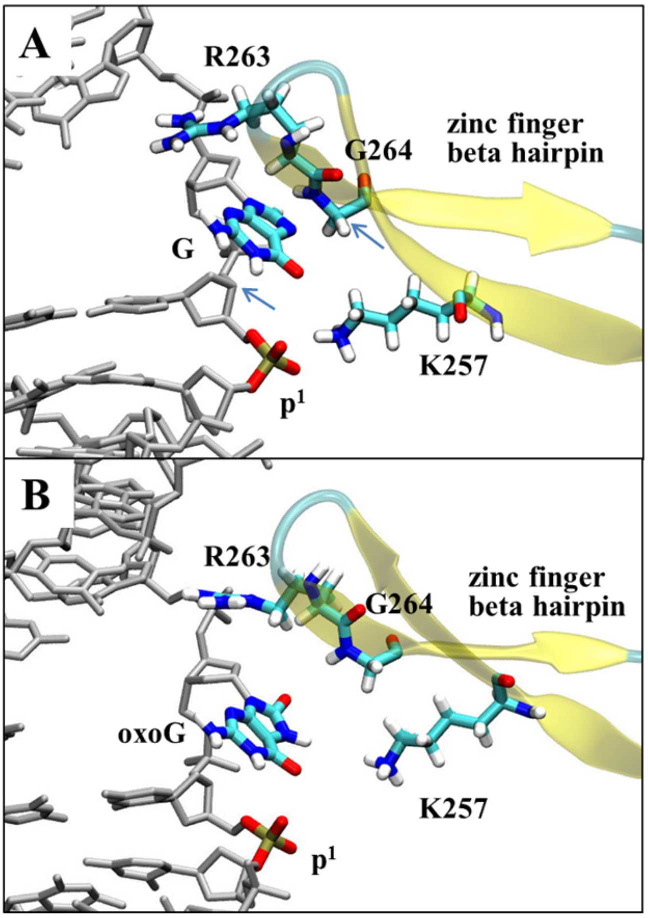


Figure 2-10. Comparison of the G system (A) and the oxoG system (B) structures at the eversion of ~75°. The distance between the C3′ of the 5′ nucleotide of the oxoG/G and the Cα of Gly264 (denoted by two blue arrows) are used to estimate the gap between the zinc finger hairpin and the DNA (discussed below).

First, there is a repulsive interaction between p^1 and the O6/N7 atoms of G, as suggested by the unfavorable pair-wise nonbonded energy shown in Figure 2-11A. OxoG has the O6 atom but not a lone pair of electrons on N7, so it does not make a strong repulsive interaction with p^1 as does G. Moreover, in the G system, the position of p^1 is fixed by the close contact of Lys257 (Figure 2-10A); whereas in the oxoG system, p^1 is disengaged from Lys257 (Figure 2-10B) and thus is free to move away from the O6 atom to avoid unfavorable interaction, so the distance between O6 of oxoG/G and Lys257 is larger in the oxoG system than in the G system (Figure 2-11B). Thus, the fixed p^1 in the G system selectively hinders eversion of G. Second, the gap between DNA and the zinc finger hairpin is wider in the oxoG system than in the G system. In the G system, the hairpin is oriented close to the DNA backbone, whereas in the oxoG system, Arg263 and Gly264 do not closely interact with oxoG, and the hairpin is further away from DNA than in the G system (Figure 2-10A,B). To roughly estimate how close the zinc finger hairpin is to the DNA backbone, we calculated the distance between C3' of the 5' nucleotide of the oxoG/G and the Cα of Gly264 (denoted by arrows in Figure 2-10A), which indicates that the gap in the oxoG system is significantly larger than in the G system throughout the intrahelical endpoint to Step II (Figure 2-12).

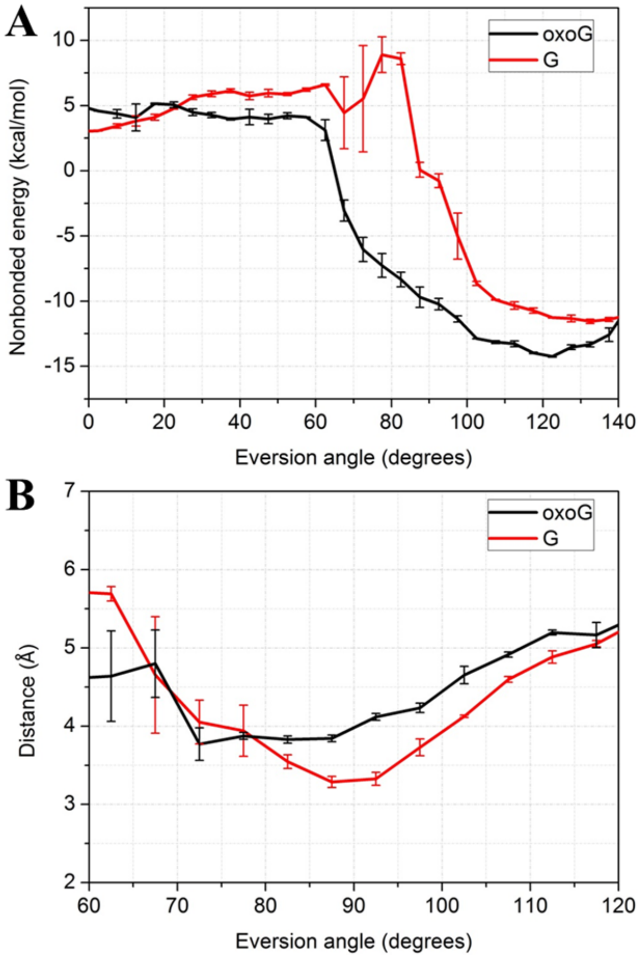


Figure 2-11. A, comparison of pairwise nonbonded energies between the base moiety of oxoG/G (black/red) and p^1. B, comparison of distances between the O6 atom of oxoG/G (black/red) and the O2P atom of p^1. The error bars reflect the difference between two independent US runs.

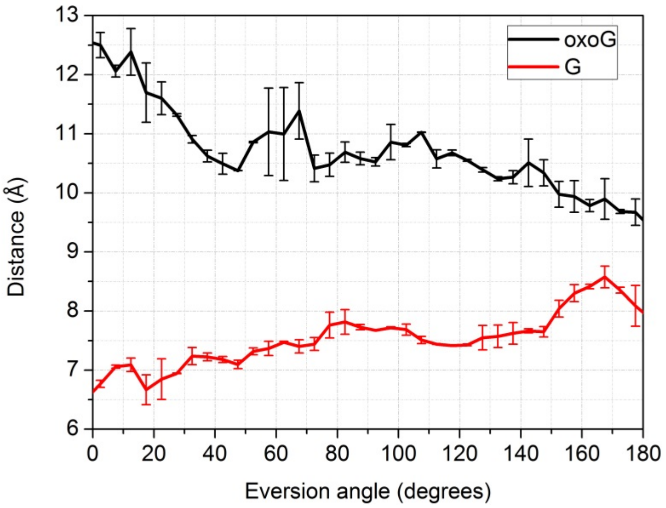


Figure 2-12. Comparison of distances between the Cα of Gly264 and the C3’ of the 5’ nucleotide of the oxoG (black) and G (red), showing that the gap between the zinc finger hairpin and DNA is wider in for oxoG than for G. The error bars reflect the difference between two independent US runs.

We suggest two reasons for the oxoG system having a larger gap. First, in the intrahelical endpoint, the 5′-side of oxoG is untwisted more than that of G. As shown in Figure 2-13, the twist angle between the target and the 5’ base is significantly smaller in the oxoG system than in the G system. Thus, the DNA strand containing oxoG rotates away from the zinc finger hairpin, making a wider gap than in the G system. This finding is consistent with a recent computational study indicating that oxoG in duplex DNA induces untwisting to its 5’-side base step because of the base−sugar (O8 of oxoG and O4’ of the sugar ring) repulsion.[89] Second, as discussed above, in Step I, G closely interacts with Arg263 and Gly264, which may draw the zinc finger hairpin nearer to DNA. Therefore, the gap for oxoG to flip through is wider than that for G; the hairpin thus sterically hinders G from further eversion (Figure 2-10A).

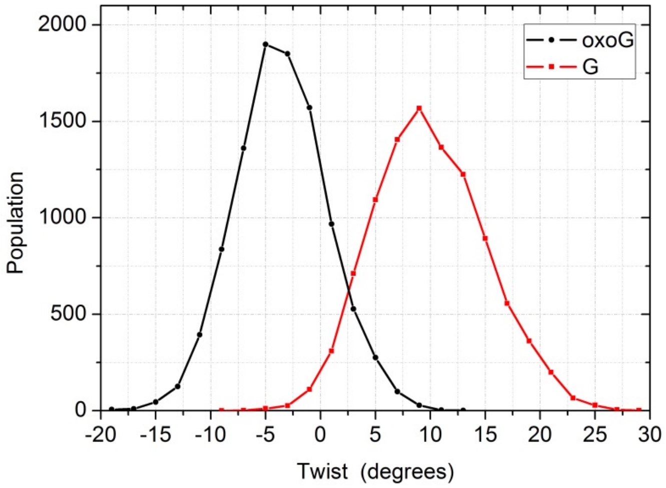


Figure 2-13. Comparison of twist angles between the target and the 5' base step in the oxoG system (black) and the G system (red). Only the intrahelical windows (the first two windows in each umbrella sampling run) were included in the calculation.

PMF of oxoG eversion shows an energy minimum at the eversion angle of ~90° (Step II), probably because oxoG is stabilized by p^1 and Asn173: the Watson-Crick edge of oxoG forms two hydrogen bonds to p^1 and the O8 atom of oxoG hydrogen bonds to the amide of Asn173 (Figure 2-14B). As G continues to flip out and reaches the eversion angle of ~90° of eversion angle, it forms two hydrogen bonds to p^1 as does oxoG; but unlike oxoG, G does not form a hydrogen bond to Asn173 since G does not have the O8 atom. Figure 2-15 indicates that Asn173 has favorable interaction with oxoG during early stages of eversion before Step II, probably due to the long distance O8-Asn173 interaction, and this interaction becomes strongest when the O8-Asn173 hydrogen bond is formed at Step II, at the eversion angle of ~90°. On the other hand, G has significantly less favorable interaction with Asn173 since it does not hydrogen-bond to Asn173. Thus, Asn173 contributes to the free energy difference between oxoG and G at step II and is probably part of the oxoG/G discrimination mechanism.

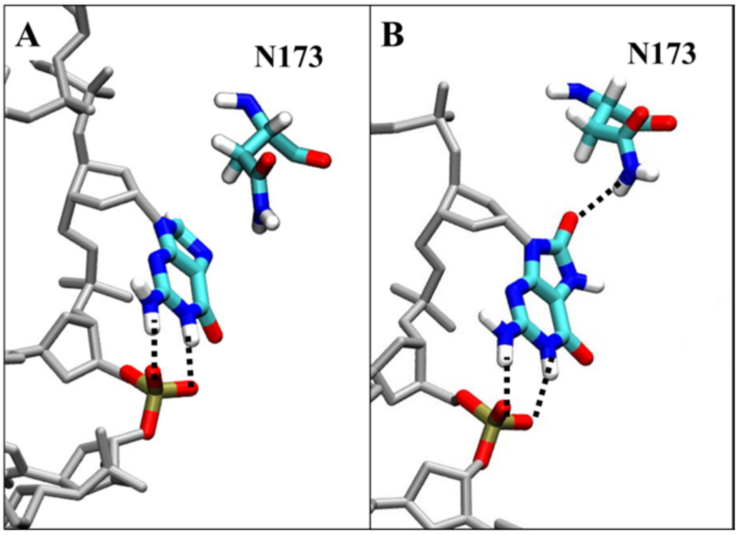


Figure 2-14. Comparison of the G system (A) and the oxoG system (B) structures at the eversion of ~90° (Step II).

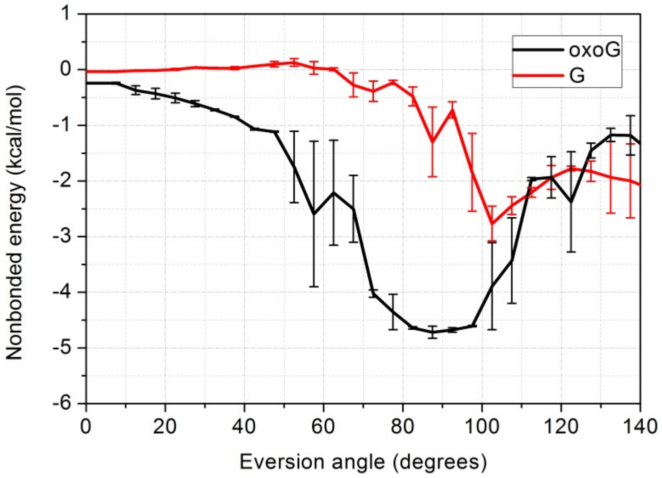


Figure 2-15. Comparison of pairwise nonbonded energies between Asn173 and the base moiety of oxoG (black) and G (red). The error bars reflect the difference between two independent US runs.

A recent article suggested a minor groove eversion pathway in Fpg, identifying a partially everted G in the minor groove pathway as a putative intermediate.[37, 90] The enzyme in this structure was covalently bound to the DNA using disulfide crosslinking technology where a N173C mutation was made and a crosslink was formed between residue 173 and the DNA backbone.[90] This work assumed that Asn173 does not play a significant role in base eversion, nor did it take into account that an N173C crosslink may sterically block the major groove pathway. Since we identify Asn173 as a key residue in recognition via a major groove path, the minor groove intermediate obtained by mutation of this residue may not reflect a biologically relevant structure.

2.3.4. Recognition of oxoG in Step III and IV

To escape Step II and flip further toward the active site, G has to overcome another barrier of ~9 kcal/mol, after which the free energy remains roughly stable (Figure 2-5); whereas oxoG only needs to overcome an energy barrier of 5–6 kcal/mol, and then the free energy oxoG drops by ~14 kcal/mol from the eversion angle of ~150° to ~280° (from Step III to Step IV, Figure 2-5). The energy drop in the oxoG system is likely because the O8 of oxoG is stabilized by the positively charged Pro1 when transiting to the active site (Figure 2-16A). Calculation of pairwise nonbonded energy between oxoG and Pro1 indicates that the favorable oxoG/Pro1 interaction emerges early in base eversion and becomes stronger as the base flips (Figure 2-17A); the nonbonded energy is strongest when O8 of oxoG hydrogen bonds to Pro1 at the eversion angles of 220°–280° (Figure 2-17A). Pro1 also makes a hydrogen bond to the N7 of G (Figure 2-16B), contributing to the energy minimum at the eversion angle of 200° in the PMF (Figure 2-5). However, compared to oxoG, the interaction of Pro1 with G is less favorable, especially after the eversion angle of 100°, as indicated by the nonbonded energy (Figure 2-17A). This is probably because O8 is an exocyclic atom whereas N7 is a ring atom, and thus Pro1 can be closer to O8 than to N7 during base eversion. Indeed, in most areas of the eversion path, the O8-Pro1 interaction in the oxoG system has closer distance than in the N7-Pro1 in the G system (Figure 2-17B). The difference in the strength of the interaction with Pro1 probably contributes to the divergence of the free energy of oxoG and G in Step III.

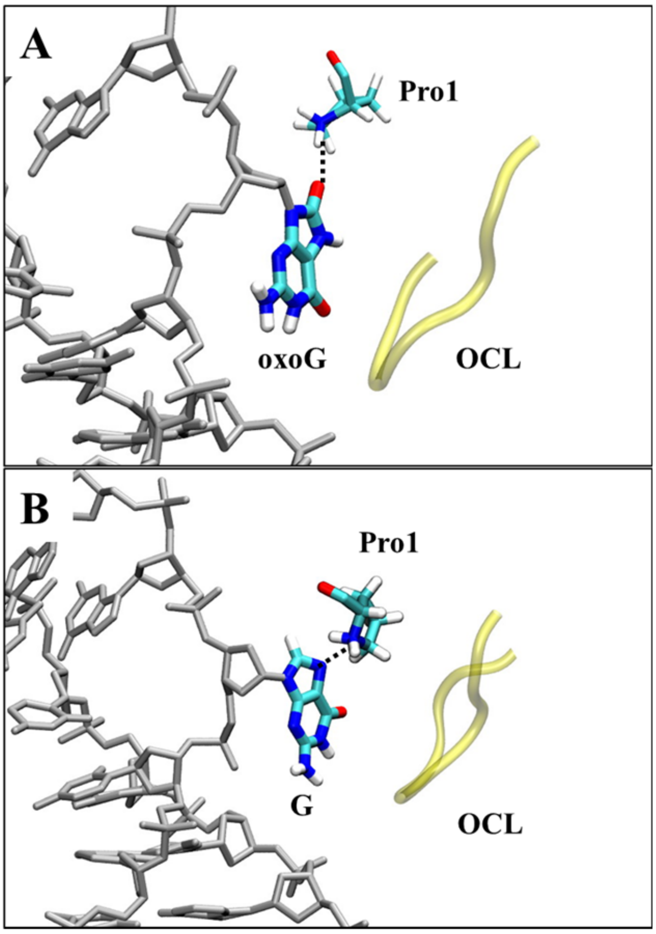


Figure 2-16. Comparison of the oxoG system structure at the eversion angle of ~220° (A) and the G system structure at the eversion angle of ~200° (B).

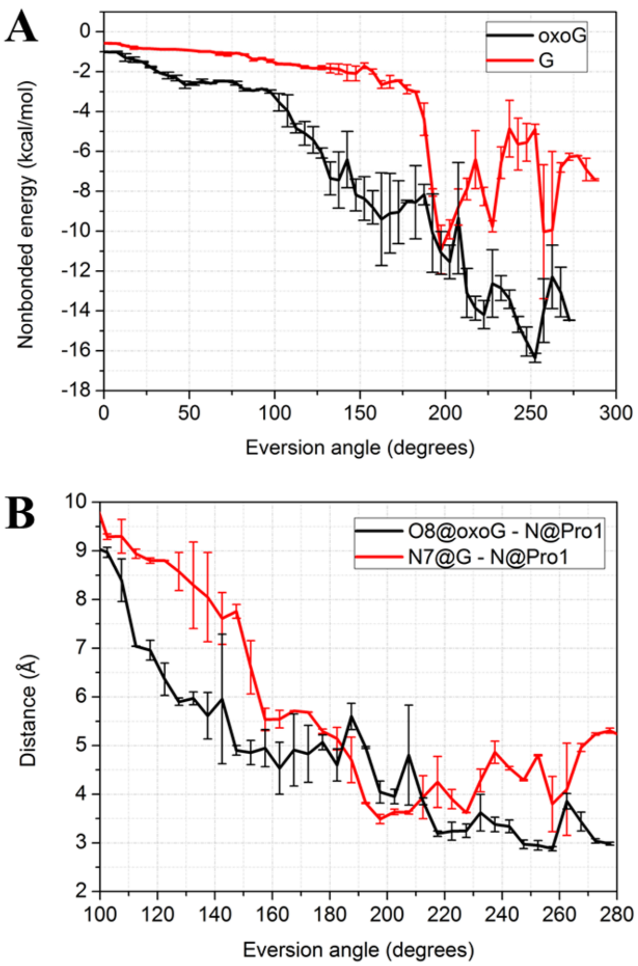


Figure 2-17. A, Comparison of pairwise nonbonded energies between the base moiety of oxoG/G (black/red) and Pro1. B, Comparison of distances from the amine of Pro1 to the O8 of oxoG (black) and to the N7 of G (red). The error bars reflect the difference between two independent US runs.

The base eversion pathway ends in the enzyme’s active site (at the eversion angle of ~260°–280°, Step IV). As indicated by the free energy profile, Step IV is a stable state for oxoG. In our simulations, the N7 of oxoG hydrogen bonds to the backbone amide of Ser219, and the O^6 of oxoG is contacted by the backbone amides of the oxoG-capping loop (OCL). These oxoG-stabilizing interactions were also seen in the crystal structure 1R2Y[29], in which an E2Q Fpg mutant is interrogating an oxoG in the active site, and thus our results suggest that the E2Q Fpg-oxoG complex is a good model of the WT system. In addition, the backbone amide and the side chain of Asn173 hydrogen bond to the 3′- and 5′ phosphate group (p^{-1} and p^{0}), respectively; and Arg263, stacking with Asn173, also contacts p^{-1} and p^{0} (Figure 2-18A). Thus Asn173 and Arg263 help to stabilize the everted nucleotide. Surprisingly, when G enters the active site, the OCL does not form any hydrogen bonds to G (Figure 2-18B), probably because the deprotonated N7 of G is rejected by the carbonyl of ser219 (Figure 2-19), and G rotates to make contacts with Arg79 and Glu77 with its Watson-Crick wedge (Figure 2-18B). Nevertheless, it would be a very rare occasion that a G could overcome the ~20 kcal/mol energy barrier to reach the active site, therefore the contacts made to G in this state are not as important for damage recognition as the ones occurs during the earlier states of base eversion.

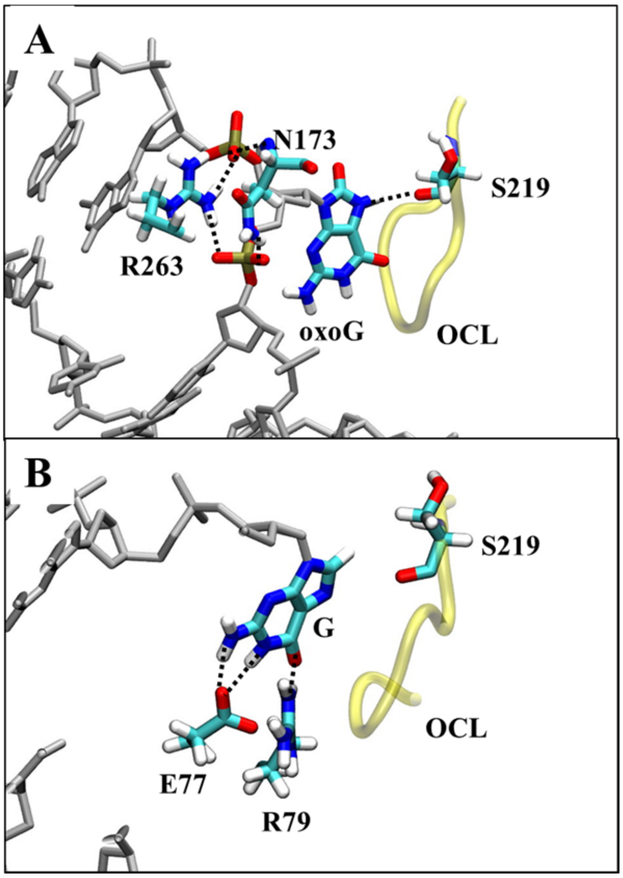


Figure 2-18. Comparison of the oxoG system structure at the eversion angle of ~275° (A) and the G system structure at the eversion angle of ~280° (B). For clarity, part of the DNA is not shown.

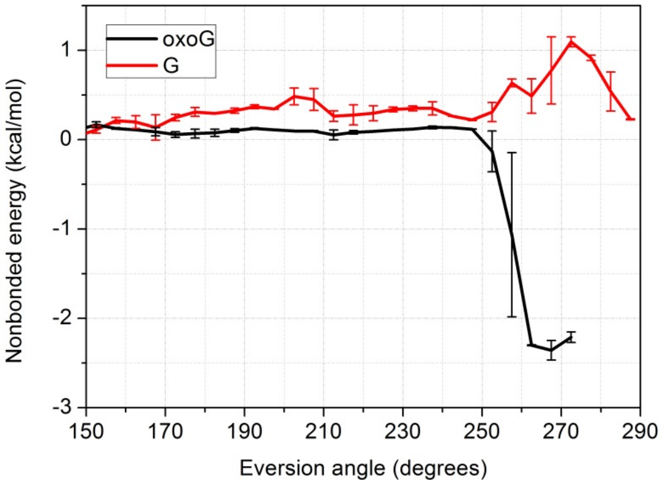


Figure 2-19. Comparison of pairwise nonbonded energies between the base moiety of oxoG/G (black/red) and Ser219 in OCL. The error bars reflect the difference between two independent US runs.

2.3.5. Arg111 recognizes the cytosine opposite oxoG and may promote oxoG eversion

Recognition of the C opposite oxoG is also important for damage repair by Fpg, since removal of an oxoG opposite a base other than C (such as A) can cause mutation. Crystallographic studies have indicated that Arg111 recognizes the C opposite oxoG by bidentate hydrogen bonds with the Watson-Crick edge of the C.[29] Previous simulations sampling the oxoG eversion through the minor groove showed that Arg111 first contacts the O2 atom of the C, competing with the intrahelical oxoG, and then invades into the DNA and forms bidentate hydrogen bonds, concomitant with the base pair opening.[37] Nevertheless, the role of Arg111 in the major groove eversion path, which is energetically more favorable than the minor groove one,[38] was unclear. Here our simulations indicate that Arg111 in the major groove path plays a similar role as in the minor groove path – it recognizes the opposite C in early stages of base eversion and may promote oxoG eversion by competing with oxoG for hydrogen bond interactions with the C. Figure 2-20 indicates the invasion process of Arg111. First, Arg111 is occluded from the DNA when the oxoG:C was intact (Figure 2-20A). When oxoG slightly shifts away from the original position, Arg111 moves close to O2 of the C and starts competing with N2 of the oxoG (Figure 2-20B). As the oxoG flips further, Arg111 comes closer to C and forms two hydrogen bonds to O2 (Figure 2-20C). The bidentate hydrogen bonds between Arg111 and the C are then developed as the oxoG is disengaged from the opposite C (Figure 2-20D). Thus, it appears that Arg111 competes with oxoG for hydrogen bonding to the opposite C and probably disrupts the base pair, promoting oxoG eversion.

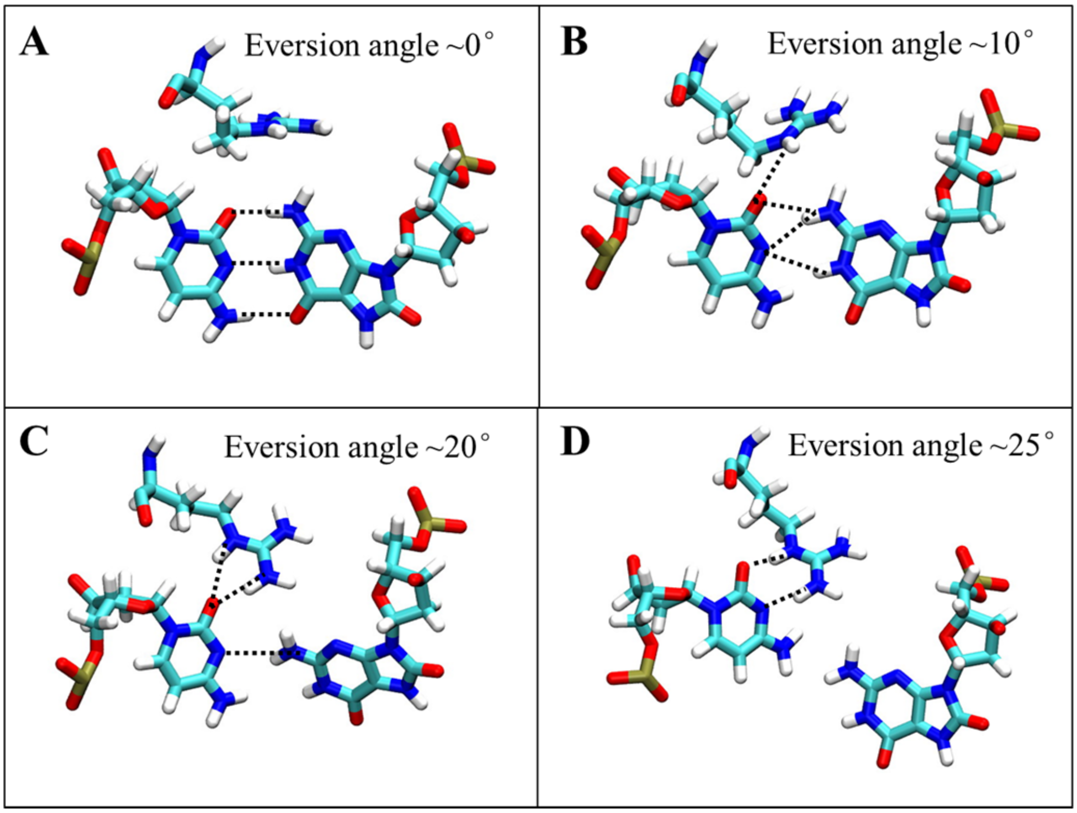


Figure 2-20. Structures of Arg111 interacting with the target oxoG:C pair in early stages of oxoG eversion.

2.3.6. The proposed oxoG/G discrimination mechanism of Fpg

Based on the findings discussed above, here we propose the oxoG/G discrimination mechanism by Fpg. In the intrahelical state, oxoG induces untwisting to the 5' base step, pushing the DNA backbone further away from the zinc finger hairpin as compared to that of the G system. Arg111 approaches the C opposite the oxoG and helps to disrupt the base pair. As oxoG opens from the pair, it quickly enters a metastable state in which it is stabilized by the DNA backbone phosphate p^1, whereas G is also in a metastable state with stabilizing interactions from Arg263 and Gly264, which draw the zinc finger hairpin nearer to DNA as compared to that of the oxoG system, and thus trapping G by sterically hindering G from further eversion (Step I). The wider hairpin-DNA gap in the oxoG system allows oxoG to pass through with significantly lower energy barrier as compared to G, and then oxoG enters the second metastable intermediate state where it is stabilized by p^1 as well as a hydrogen bond between O8 and Asn173, which is not observed in the G system (Step II). Eversion of both bases is then facilitated by the interaction from Pro1, which is stronger in the oxoG system than in the G system (Step IV). In the end, oxoG enters the active site and is contacted by Ser219 and other interactions from the oxoG-capping loop, whereas G is rejected by the OCL, probably due to the unfavorable interaction from Ser219, and instead stays in an adjacent position where it is contacted by Arg79 and Glu77 (Step IV).

2.3.7. Mutation studies

Among the key residues in the proposed damage discrimination steps, Arg111, Arg263, Asn173 and Pro1 are strictly conserved among Fpg (Figure 2-7), indicating their important roles in Fpg's function. In order to experimentally verify the critical residues predicted by molecular dynamics, our collaborators produced a set of corresponding mutations in Fpg from *E. coli*, an enzyme extensively characterized biochemically. Arg108 (Arg111 in *B. st* Fpg) was mutated into Lys (to conserve bulk, polarity, and charge), Gln (to conserve bulk and polarity, but not charge), or Leu (to conserve bulk only). Asn168 (Asn173 in *B. st* Fpg) was mutated into Asp to eliminate hydrogen-bonding donor capacity and introduce charge, or Gln to increase bulk moderately.

Finally, Arg258 (Arg263 in *B. st* Fpg) was mutated into Lys, Gln, or totally eliminated by conversion into Ala. Mutation studies of Pro1 was not performed, because Pro1 is needed for catalysis. The proteins were purified and their steady-state kinetics was determined.

All mutants were clearly compromised in their ability to cleave oxoG-containing DNA (Figure 2-21A, C). Characteristically, the activity on the AP substrates was affected to a much lesser degree, if at all (Figure 2-21B, D). This conservation strongly suggests that the mutations do not disturb the general enzyme structure, its ability to bind DNA or position the catalytic residues in a proper conformation, and that their detrimental effect is likely due to the disruption of the eversion process. AP sites are known to exist in a spontaneous dynamic equilibrium between intra- and extrahelical states,[91] and thus the requirements for active enzyme-induced eversion are likely relaxed. The mutants could be also cross-linked to the AP substrate by $NaBH_4$ treatment (not shown), providing a way to quantify the fraction of the active enzyme.

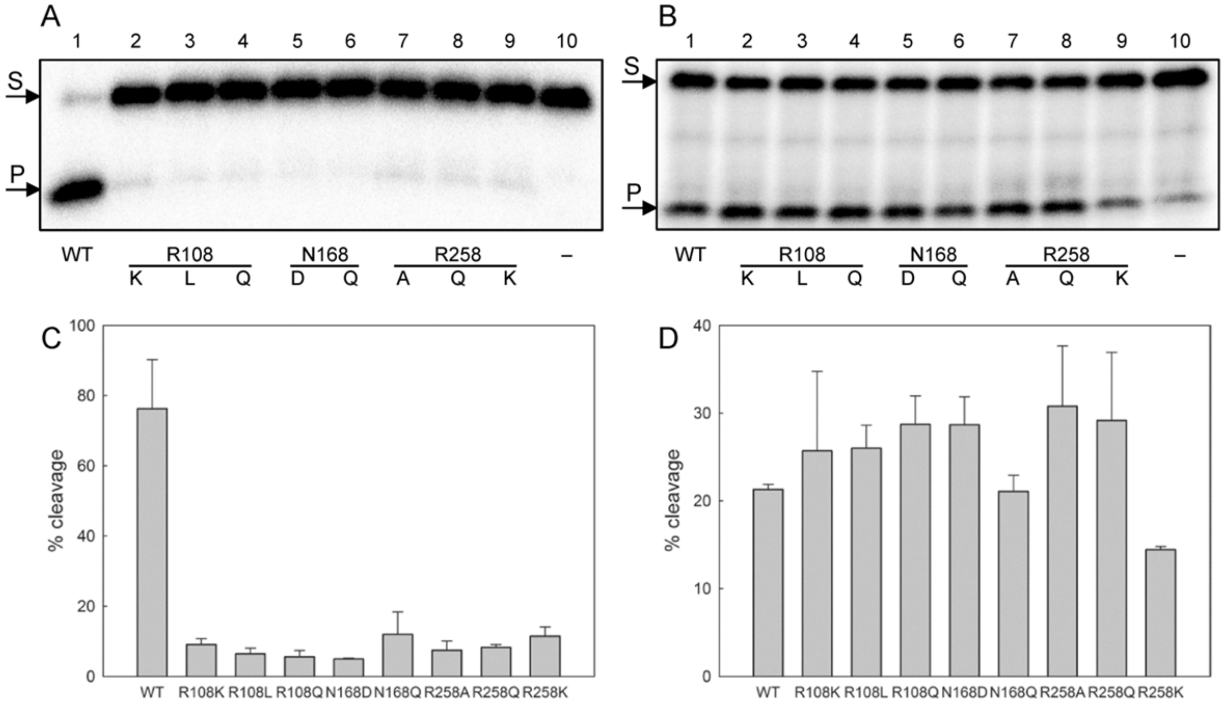


Figure 2-21. Cleavage of oxoG substrate (A, C) and AP substrate (B, D) by wild-type and mutant Fpg proteins. Representative gels (A, B) and mean ± s.d. (*n* = 3) are shown. The concentration of the substrate was 50 nM in all experiments. Fpg was taken at 2 nM (oxoG substrate) or 0.2 nM (AP substrate). S, substrate, P, cleaved product.

The kinetic constants obtained for oxoG substrate after normalization for the active enzyme concentration support the role of Arg108, Asn168, and Arg258 in the eversion process, as shown in Table 2-1. The mutations hardly affected K_M, in the cases when it could be measured independently, again indicating that the overall affinity for the substrate and stability of the enzyme–substrate complex did not change much. On the contrary, the catalytic constant was down at least 10-fold, suggesting the disruption of the events following substrate binding. The least affected was the N168Q mutant, which still has the ability to form all hydrogen bonds but presumably does it in a less favorable conformation due to the extra methylene group. On the other hand, the N168D mutation, abolishing the transient hydrogen bonds with O8 and DNA phosphates and introducing an unfavorable electrostatic repulsion of oxoG, was the most detrimental, matching the predictions of the simulations. Of all Arg258 mutants, the R258Q retained most activity, suggesting importance of stacking and hydrogen bonding interactions with this residue. Electrostatic interactions alone do not seem to play a significant role in the oxoG guidance through the eversion pathway, since the substitution of Lys for Arg258 had the same effect as the abolition of Arg258 altogether. Since the simulations indicate that Arg263 plays roles in two steps of oxoG recognition (Steps I and IV), and mutation experiments cannot separate these two steps, it is still unclear how the mutations of Arg263 effect Fpg's function. To solve this problem, i*n silico* mutagenesis may be need in future work. Finally, Arg108 mutants retained minor residual activity only if the charge was conserved. As discussed above, our simulations showed that the equivalent arginine, Arg111, first contact O2 of the C opposite to the oxoG and acts as a lever in penetrating the base stack and then competes with oxoG for the Watson–Crick bonding edge of the opposite C. Obviously, Lys could form a bond to O2 and penetrate the helix, partially destabilizing the intrahelical oxoG:C base pair, but fails to form the C-specific bonds later.

Table 2-1. Kinetic parameters of wild-type and mutant *E. coli* Fpg

Enzyme	**K_M, nM**	**k_{cat}, s^{-1}**	**k_{sp}, $nM^{-1}\cdot s^{-1}$**
wild-type	8.1 ± 2.3	1.7 ± 0.1	0.21 ± 0.06[a]
R108K	18.0 ± 4.0	0.070 ± 0.006	0.0039 ± 0.0009
R108L	n/d[b]	n/d	$(6.4 \pm 0.4)\times10^{-4}$
R108Q	n/d	n/d	$(6.4 \pm 0.3)\times10^{-4}$
N168D	n/d	n/d	$(4.7 \pm 0.2)\times10^{-4}$
N168Q	8.1 ± 1.9	0.16 ± 0.01	0.020 ± 0.005
R258A	n/d	n/d	$(7.6 \pm 0.3)\times10^{-4}$
R258Q	6.3 ± 3.0	0.08 ± 0.01	0.013 ± 0.006
R258K	n/d	n/d	$(6.0 \pm 0.4)\times10^{-4}$

[a] $k_{sp} = k_{cat}/K_M$

[b] Cleavage too low to separately calculate K_M and k_{cat}; k_{sp} calculated from the slope of the linear part of the v_0 *vs* [S] dependence.

2.4. Conclusion

To understand how Fpg efficiently recognizes oxoG from a vast excess of undamaged guanines, we modeled the free energy pathway of eversion for both the oxoG damage and an undamaged G using MD simulations. In addition to free energy profiles, structural analysis of the simulation data revealed specific interactions recognizing oxoG. Arg111 recognizes the C opposite the oxoG and probably promotes eversion of oxoG. Early interaction with the second 5' phosphate specifically recognizes the protonated N7 of oxoG; Asn173 and Pro1 play a role in recognizing O8 of oxoG and thus facilitating further eversion of oxoG. We then used biochemical mutation analysis of residues to verify the critical roles of Arg111, Asn173 and R263 in oxoG eversion. On the other hand, eversion of the undamaged G is hindered by the unfavorable early interactions from the DNA backbone, and the narrow gap between DNA and the zinc finger beta hairpin also prevents G from proceeding along the eversion pathway. Therefore, Fpg can discriminate against G in favor of oxoG in early stages of base eversion, and this early damage discrimination process is much more efficient than the one occurring in the active site, allowing Fpg to quickly detect oxoG during fast sliding.

3. Molecular Simulations reveal a common strategy of 8-oxoguanine recognition by DNA glycosylases

Acknowledgments

The material presented in this chapter contains direct excerpts from a manuscript that is in preparation by Haoquan Li, Dmitry O. Zharkov, Christina Bergonzo, Lin Fu, Arthur J. Campbell, Carlos de los Santos, Arthur Grollman, and Carlos Simmerling. The manuscript was written by Haoquan Li with suggestions from Carlos Simmerling and Dmitry O. Zharkov. Haoquan Li performed the simulations and data analysis. Christina Bergonzo, Lin Fu and Arthur J. Campbell developed the basic work flow for NEB and umbrella sampling.

3.1. Introduction

8-Oxoguanine (oxoG) is a mutagenic DNA lesion generated under oxidative stress and it differs from its precursor, guanine, by only two positions (O^8 and H^7). OxoG can be detected and repaired via the base excision repair (BER) pathway which is initiated by DNA *N*-glycosylases that can remove their cognate lesions and create apurinic/apyrimidinic (AP) sites. The human 8-oxoguanine-DNA glycosylase (hOGG1) is one of the major glycosylases involved in the human BER pathway, and is responsible for removal of oxoG from oxoG:C base pairs in duplex DNA. Formamidopyrimidine-DNA glycosylase (Fpg, also known as MutM) is a functional analog of hOGG1 in prokaryotes, but these two enzymes have completely different sequences and folds. Both hOGG1 and Fpg search for oxoG by sliding along DNA in a one-dimensional Brownian manner and the DNA is kinked at the interrogated site.[30, 36, 92] When an oxoG:C base pair is located, the oxoG is everted from the DNA helix and inserted deeply into the catalytic pocket before its *N*-glycosidic bond is hydrolyzed.

Base eversion (or base flipping) is a common feature of diverse glycosylase–DNA complexes. The target nucleotide must rotate out from the double helix so that the glycosylase can accommodate the damaged base at the extrahelical catalytic pocket and hydrolyze the glycosidic bond. This process involves a series of conformational changes described as “pinch–push–plug–pull”.[93] For both Fpg and hOGG1, three pre-excision steps of processing oxoG have been suggested using stopped-flow kinetics: (i) initial encounter and extrusion of oxoG (“pinch-push”), (ii) insertion of certain hOGG1 residues into DNA (“plug”), and (iii) insertion of oxoG into the active site of hOGG1 which isomerizes for catalysis (“pull”).[27, 28]

Crystallography studies have been providing structural information of the base eversion process in hOGG1. A putative early intermediate (PDB ID: 2I5W) along the base flipping pathway showed a opened G:C pair of which the G is slightly flipped-out towards the major groove, and is stabilized by an hydrogen bond to the adjacent oxoG.[33] In a putative later intermediate (PDB ID 1YQK), the target G is extrahelical but rejected from the active pocket, and instead stays at an adjacent “exo-site”.[31] In both of these intermediates, an engineered disulfide crosslink was used to force hOGG1 to interrogate a non-cognate G:C pair. This crosslink, between Cys149 (mutated from Asn149 in WT) and the cytosine of the target pair, occupies the space where the otherwise unopened G should be at and thus forces the G out of the helical stack. On the other hand, oxoG has also been seen in the exo-site when the active site is sterically obstructed by a Gln319Phe mutation (PDB ID 2NOF),[40] but compared to the G in the exo-site in 1YQK, the oxoG in 2NOF is slightly closer to the active site. Crystal structure of extrahelical oxoG in the active pocket of a catalytically inactive hOGG1 was captured without using a crosslink (PDB ID: 1EBM).[13]

Although the crystal structures with Cys149-Cyt crosslink (2I5W, 1YQK, 2NOF) do show the snapshots of base eversion along a putative major groove pathway, it should be mentioned that Cys149 inserts into DNA from the minor groove and forms a disulfide bond with the cytosine, and such an artificial crosslink prevents the target base from presenting in the minor groove direction. Nevertheless, the major groove eversion mechanism has been suggested for other glycosylases. For example, the crystal structure of uracil DNA glycosylase (UDG) suggested a major groove eversion mechanism, as UDG, like hOGG1, binds DNA from the minor groove and the minor groove path is sterically hindered[94]. Also, computational studies on the Fpg-DNA complex indicate that the eversion of oxoG through the major groove is energetically more favorable than through the minor groove[38]. Thus, it is plausible that hOGG1 also extrudes oxoG through a major groove pathway.

To date, the mechanism by which hOGG1 so specifically distinguishes the rare oxoG among a great excess of undamaged G is still not clear. Crystallographic studies have shown that hOGG1 can distinguish oxoG from G at the active site, by forming a discriminatory hydrogen bond between the carbonyl of Gly42 and the pyrrole N^7 atom of oxoG; if a guanine were present in the active site, its non-protonated N^7 would electrostatically clash with Gly42.[13, 31] In addition, the $Lys249^+/Cys253^-$ dipole has been proposed to help recognize the N^7/O^8 dipole of oxoG[31] but this dipole-dipole interaction is not imperative, since the double mutant Lys249Cys/Cys253Lys is able to excise oxoG.[34] Currently the Gly42/H^7 hydrogen bond is the only direct interaction observed to distinguish oxoG from G. However, biochemical studies have revealed several hOGG1 substrates besides oxoG possess O^8 but not H^7, such as 2,6-diamino-4-oxo-5*N*-

methylformamidopyrimidine (Me-FaPyG) and 7-methyl-8-oxo-2'-deoxyguanine.[35, 95] These findings suggest that both H^7 and O^8 may be used for lesion recognition, and thus there are probably some unrevealed interactions between hOGG1 and the H^7/O^8 atoms other than the single G42/H^7 hydrogen bond formed in the active pocket. In addition, the extremely fast diffusion rate of hOGG1 sliding along DNA almost excludes the possibility that hOGG1 interrogates every encountered DNA base in its active site, which is at the very end of the base eversion pathway.[36] Therefore, we hypothesize that the substrate recognition by hOGG1 occurs not only at the active site, but also at the early stages of the base eversion pathway. Unfortunately, potential early stage intermediates for the wild type functional system have relatively high energy and low population, and they are still inaccessible directly by experimental methods, thus hindering our understanding of the damage recognition mechanism by hOGG1.

To understand how hOGG1 recognizes oxoG against G, here we energetically and structurally characterize the base eversion pathway in hOGG1 through the use of molecular dynamics (MD), an approach that has a unique advantage of connecting energy, structure and dynamics for conformational changes. Following the protocol of our previous study on Fpg system,[38] here we compare the free energy and structure of oxoG and G eversion to understand the damage recognition mechanism of hOGG1. We found that in addition to the active site discrimination against oxoG and G, the system also provides several earlier levels of damage recognition, and this mechanism is similar to that of Fpg.

3.2. Methods

3.2.1. Force field and parameters

The SANDER module in the Amber 10 and Amber 11 suites of programs[51, 96] was used for molecular dynamics (MD) simulations. For all systems the ff99SB force field[46] was used with the parmbsc0 correction of the DNA alpha and gamma backbone parameters.[49] The parameters for oxoG were obtained from Miller et al.[76] The TIP3P[97] explicit solvent water model was used to solvate the protein-DNA complexes.

3.2.2. Initial structure selection for base flipping pathway calculation

Partial nudged elastic band[50] (PNEB) is a simulation method that finds a minimum energy pathway between two given structures (also called "endpoints") that define the start state and the end state, respectively. Intermediate states between the start and the end states can also be assigned beforehand and then all these structures will be linked to form a complete minimum potential energy path. In this case, the start state is hOGG1 interrogating an intrahelical unopened oxoG:C or G:C base pair (referred to as the "intrahelical state"), whereas the end state is a state that the target base is inserted into the active pocket of hOGG1 (referred to as the "in-pocket state"); an intermediate state is any state that the oxoG or G is neither paired with the opposite cytosine nor present at the active site.

To generate a base eversion pathway using PNEB, we need an intrahelical endpoint, an extrahelical endpoint, and one or more intermediates to seed the major groove path. The 1EBM[13], 2I5W[33] and 1YQK[31] crystal structures are reasonable candidates for initial structures of base flipping path generation. They represent the end state and two putative intermediate states of base flipping, respectively: in the 1EBM structure, the oxoG is everted into the active site of hOGG1; in the 2I5W structure, the target G is slightly everted towards the major groove and it forms a stabilizing hydrogen bond with the neighboring oxoG; 1YQK is a putative later intermediate in which the target G stays at an extrahelical exo-site along the major groove path. To fit these crystal structures into our oxoG eversion model, we reverted the hOGG1 back to wild type and made the DNA sequence consistent among all structures that oxoG is at the target flipping site and other bases are undamaged.

While the structures of the extrahelical endpoint (1EBM) and the intermediates (2I5W/1YQK) are available from experiments, currently there is no structure available for the intrahelical endpoint, therefore this endpoint was obtained by computationally modifying an existing hOGG1 structure. The 2I5W is the best candidate for precursor of the intrahelical state since the target base in the 2I5W is only slightly everted, thus it will be relatively easy to force the everted base to be fully intrahelical. The modification was performed after initial equilibration of the oxoG-complex model of 2I5W (see below).

For the G eversion model, the 1YQK structure was reserved to validate our methodology, because crystallographic studies have suggested that 1YQK is a relatively stable intermediate state for an extrahelical G and thus we expected our simulation can independently sample this structure. Therefore, we did not use 1YQK in the G-complex PNEB simulation as a seeding intermediate, but instead used another intermediate structure modified from 2I5W (see "Calculations of the base flipping paths using PNEB" in Methods). In our G eversion model, the hOGG1 is also wild type and the DNA sequence is same as that in the oxoG model except that the everted oxoG was changed to G.

3.2.3. Generation and equilibration of structures from crystal coordinates

All protein mutations present in the 1EBM, 1YQK and 2I5W structures were reverted back to wild-type residues using Swiss-PdbViewer[98]. Artificial disulfide crosslinks in the crystallization of 1YQK and 2I5W were not included in the simulations. The missing residues (Gly80, Asp81 and Lys82 in 1EBM and 1YQK; Gly80, Asp81, Lys82 and Ser83 in 2I5W) at the edge of protein globule far from the DNA binding site (about 40 Å away) were built using Modeller[99]. The DNA sequence for all structures in this work is 5'-GGGGTAGACCTGGAC-3', 5'-GTCCAXGTCTACCCC-3' ("X" is oxoG in the oxoG-complex or G in the G-complex), which is adopted from the sequence in 1EBM. When building the initial structures, the nucleotides that are in the crystal structure but not in the above sequence were deleted, and the ones in the above sequence but not in crystal structures were built in with UCSF Chimera[100] using the coordinates of a standard B-form duplex generated by 3DNA[101]. For base substitutions where G was changed to oxoG or vice versa, the atoms different between oxoG and G (the H^8 or O^8/H^7 atoms) were deleted and the desired substitution was then built using the tLEaP module in Amber.[102] Protonation states for all residues in 1EBM were determined using MolProbity[103] and were used as the standard for other structures. The structures were then solvated in ~14,500 TIP3P[97] explicit water molecules in a truncated octahedron box whose boundary is at least 10Å from the solute.

Each of the solvated structures was minimized and equilibrated in five steps: (i) a 10000-step minimization with positional restraint on the heavy atoms of the hOGG1-DNA complex; (ii) a 100ps MD simulation in which the system was heated linearly to the target temperature of 310K while the heavy atoms of the complex were restrained; (iii) a 100ps and a 250ps MD simulations with the heavy atoms of the complex fixed by 100 and 10 $\text{kcal}\cdot\text{mol}^{-1}\cdot\text{Å}^{2}$ restraints, respectively; (iv) a 100ps, a 200ps and a 250ps MD simulations with the heavy atoms of the protein and DNA backbones fixed by 10, 1 and 0.1 kcal·mol^{-1}·Å^{2} restraints, respectively; (v) a final 2ns unrestrained simulation. During minimization and equilibration SHAKE was employed to constrain bonds involving hydrogen atoms[79], and a 1 fs time step was used. The particle mesh Ewald method was used to approximate long-range Columbic interactions[80, 81] and the nonbonded cutoff was set to 8Å. Through step (iii) to (v) constant temperature of 310K (reflecting the human body temperature) and constant pressure of 1 atm were maintained by the weak-coupling algorithm.[82]

3.2.4. Generation and equilibration of the intrahelical structures

For the oxoG-complex, the initial structure for generating the intrahelical structure was taken from after the 100ps MD of step (iv) of the 2I5W equilibration (see above), then two steps were performed. (a) The oxoG (which was changed from the everted G of the original PDB structure before equilibration was performed) and the opposite C were forced back to pair together by using distance restraints on the Watson-Crick hydrogen bonds for 1ns, while the rest of DNA was fixed with 1 kcal·mol^{-1}·Å^{2} positional restraints. Meanwhile 1 kcal·mol^{-1}·Å^{2} targeted restraints, referencing to the coordinates of the 1KO9 structure, were applied on the protein backbone and also on the side chains of Asn149 and Arg204, so as to move them away from the "plugging" position to allow the oxoG to return intrahelical. (b) The structure was then equilibrated in a 250ps simulation with 0.1 kcal·mol^{-1}·Å^{2} restraints on the backbone heavy atoms, followed by 2ns unrestrained MD.

For the generation of the intrahelical G-complex, the initial structure was taken from after step (a) of the intrahelical oxoG generation and the oxoG was changed to G. Then a 10000-step minimization was performed with the heavy atoms of the protein-DNA complex restrained, followed by an equilibration step same as step (b) of the intrahelical oxoG-complex equilibration.

3.2.5. Calculations of the base flipping paths using PNEB

The PNEB protocol was adapted from our previous study on Fpg.[38] Because one intermediate is not sufficient to seed a major groove path in hOGG1, here we used two seeding intermediates. For both the oxoG-complex and the G-complex, the base flipping path was generated in PNEB simulations by linking 32 replicas in total along the path: 10 replicas of the equilibrated intrahelical structure, 10 replicas of the equilibrated in-pocket structure, and 6 replicas of each of the two guiding intermediates, which were needed to seed the eversion path through the major groove.[38] For the oxoG-complex, the intermediates generated from the coordinates of 2I5W and 1YQK (see above) served as guiding intermediates. For the G-complex, one of the guiding intermediates was generated from the coordinates of 2I5W, while the other guiding intermediate, unlike the oxoG-complex, was not generated from 1YQK, because 1YQK was reserved for later validation and not used in calculation of the eversion path. Thus, another intermediate structure was needed to substitute 1YQK to guide the path. This intermediate structure is generated by modify the 2I5W intermediate of G-complex: a 50ps MD simulation was performed on the equilibrated 2I5W intermediate, forcing the target G to flip further to 9.5 Å of the eversion distance (see definition at Figure 3-1) with 10 $\text{kcal}\cdot\text{mol}^{-1}\cdot\text{Å}^2$ restraints on the reaction coordinate.

In PNEB, the initial temperature was 310K and the spring forces were set to 1 $\text{kcal}\cdot\text{mol}^{-1}\cdot\text{Å}^2$ for the first 100ps path optimization, followed by a 500ps equilibration with 20 kcal·mol

$^{1}\cdot$Å^{2} spring forces. Over the next 100ps the structures were gradually heated to 380K and this temperature was kept for 200 ps, and then was decreased back to 310K over the next 100 ps. During this annealing process the spring forces were set to 50 kcal·mol$^{-1}\cdot$Å^{2}. The final 500ps production runs were then performed, with 30 kcal·mol$^{-1}\cdot$Å^{2} spring forces.

3.2.6. Choice of reaction coordinate for base flipping

The choice of a reaction coordinate is important for the success of umbrella sampling. Our lab has improved the reaction coordinate for base flipping, a modified COM pseudodihedral angle (CPDb) which describes how far the base is everted from the duplex[83]. Combination of CPDb and the glycosidic angle has been applied to measuring the progress of base flipping through both directions in Fpg[38]. However, CPDb may not accurately measure the base flipping paths in hOGG1, since it cannot distinguish the in-exo-site and the in-pocket states, both of which have CPDb around 210–220° (Figure 3-2). Thus, we decided to use a new reaction coordinate "eversion distance" (Figure 3-1) which can well separate the exo-site and the active site positions. As shown in Figure 3-3, the exo-site and the active site positions are at ~15 Å and ~18 Å of the eversion distance, respectively, while both positions have an eversion angle around 220°. The glycosidic angle measurement used for the Fpg base flipping[38] are not used in this work, because the eversion distance used here is also sensitive to the glycosidic bond rotation required for eversion (Figure 3-4); especially during the base entering the active site from the exo-site, where the glycosidic angle drops from ~0° to ~-120°, the eversion distance accordingly increases from ~15 Å to ~18 Å. Therefore, the eversion distance was used as the sole reaction coordinate in umbrella sampling.

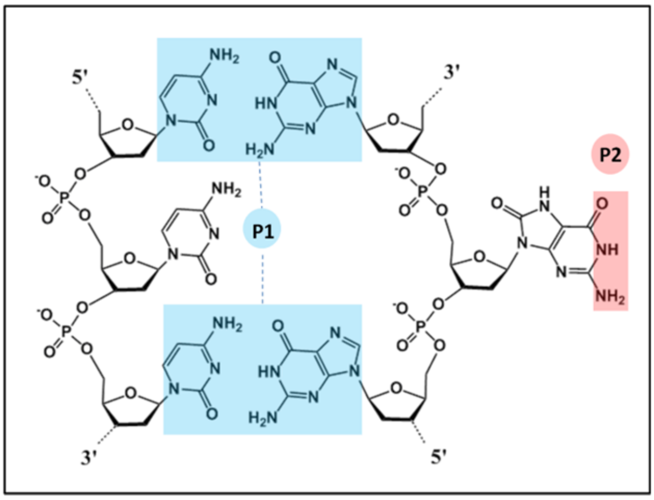


Figure 3-1. Eversion distance, the distance between P1 (boxed in blue) and P2 (boxed in red), used as reaction coordinate for base eversion.

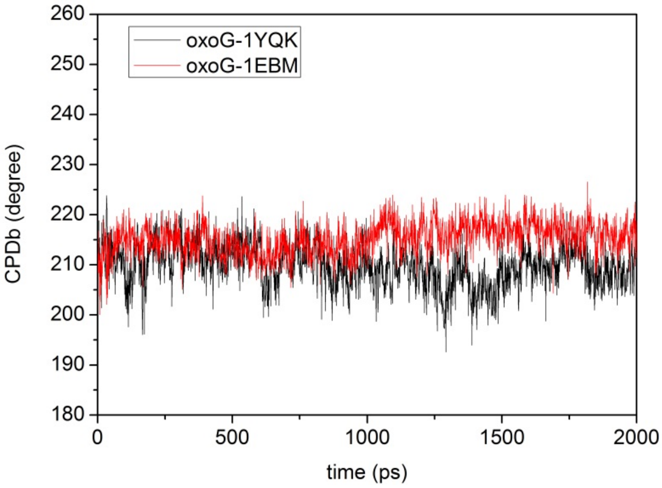


Figure 3-2. Measurement of CPDb during the 2ns MD of the oxoG-complex starting from the 1YQK and 1EBM structures.

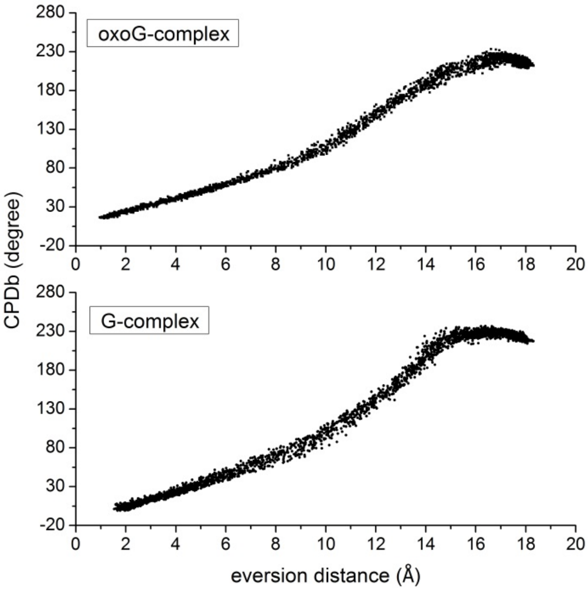


Figure 3-3. PNEB trajectories of eversion of oxoG and G plotted two-dimensionally in CPDb vs. eversion distance. The eversion angle keeps roughly the same value after the eversion distance reaches 16 Å.

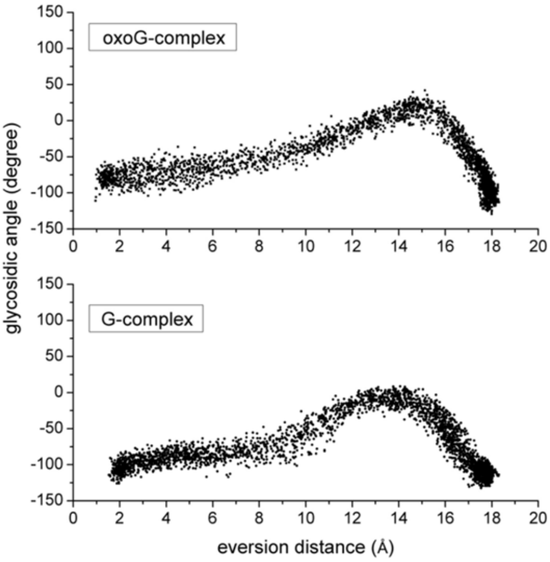


Figure 3-4. PNEB trajectories of eversion of oxoG and G plotted two-dimensionally in glycosidic angle vs. eversion distance. Along the whole pathways, the change of glycosidic angle

correlates with the change of eversion distance.

3.2.7. Calculations of the potential of mean force (PMF) for base flipping

Umbrella sampling (US) was used to obtain the PMF as a function of the eversion distance. In order to further enhance conformational sampling, Hamiltonian replica exchange was employed.[104] The US protocol was similar to those in our previous studies using conventional umbrella sampling.[83, 105] For both the oxoG-complex and the G-complex, two completely different sets of initial structures were taken from the PNEB production trajectories and the US was performed independently for 2.5ns in an NVT ensemble for each set. The US windows were evenly spaced along the eversion distance with 0.3 Å intervals, and was restrained by 10 kcal·mol^{-1}·Å^{2} umbrella potential along the eversion distance; the spacing of the windows and the umbrella restraint constant were determined by the criteria that the windows can overlap with their neighbors along the reaction coordinate, as shown in Figure 3-5. The time step was 2 fs and the replicas attempted to exchange with their neighbors in every 10 ps of the simulations. The temperature was maintained at 310K using a Langevin thermostat[85] with a 75.0 ps^{-1} collision frequency. The eversion distance values were recorded every time step and were then analyzed using the weighted histogram analysis method (WHAM).[55] The error bars were then calculated from the difference of the two independent runs. The intrahelical endpoint of the PMF is set to 0 kcal/mol for both systems.

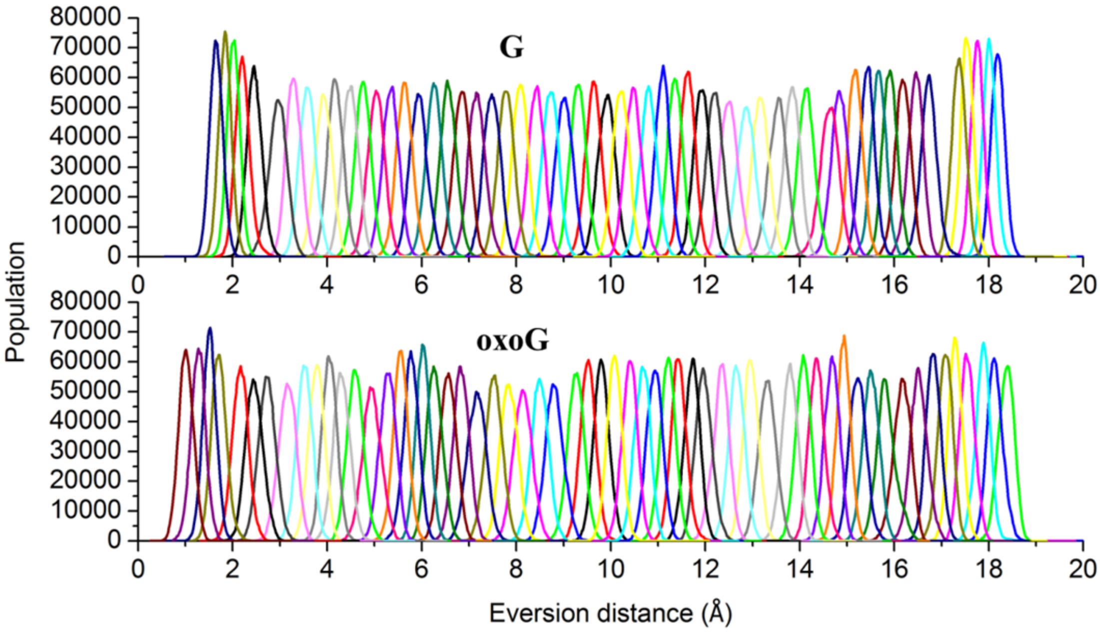


Figure 3-5. Overlap of the window distributions along the eversion distance in umbrella sampling. Each curve represents the population distribution of a window.

To check the structural shifting and possible discontinuity of the path, we introduce the glycosidic angle which describes the rotation of the glycosidic bond, one of the major motions of the flipping base, as an additional reaction coordinate for testing. Figure 3-6 indicates the G has higher fluctuation along the glycosidic angle than does oxoG, and both of them has highest fluctuation around the eversion distance of 8 Å, suggesting possible inaccuracy of PMF in that area. Nevertheless, the desired eversion pathways are sampled continuously by umbrella sampling.

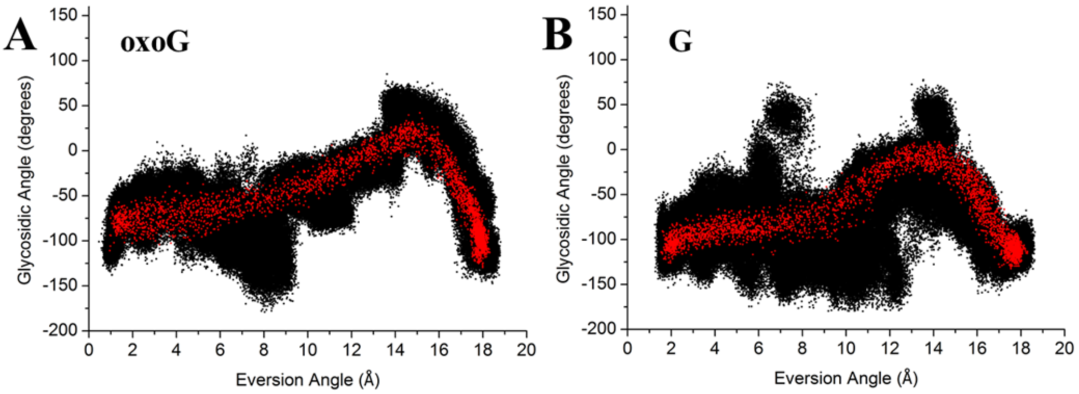


Figure 3-6. The glycosidic angle of the target base *versus* the eversion angle in the umbrella sampling simulations (black) and in NEB (red) for the oxoG-complex (A) and the G-complex (B). For each system, both of the two independent runs were included in the plot.

3.3. Results and discussion

3.3.1. Simulation strategy

To test our hypothesis that oxoG recognition by hOGG1 occurs before complete insertion of target base into the active site, we structurally and energetically compare the eversion of oxoG and G in hOGG1, which are the processes of base flipping from inside DNA (intrahelical state) to the active site of hOGG1 (in-pocket state). The base eversion process in hOGG1 has been found to occur on the millisecond timescale[28] and is currently unlikely to be sampled by conventional unrestrained MD simulations; therefore time-independent methods were used to sample the paths. Partial nudged elastic band (NEB)[50] and umbrella sampling simulations have been applied successfully to investigating the free energy of oxoG eversion in Fpg[38]. Here, we again used PNEB to generate the eversion paths for oxoG and G, and the energy profiles were obtained through the use of umbrella sampling. As mentioned above, crystallographic studies have proposed a major groove path for hOGG1[31, 33, 40, 106], and hOGG1's function analog Fpg also prefers base flipping through the major groove[38]. Thus, in this work we focus only on comparing the major groove paths.

3.3.2. Intrahelical structure – the starting point of the base flipping pathway

Modeling a path requires at least two endpoints; for base eversion the starting endpoint is the intrahelical structure, whereas the other endpoint is the in-pocket structure. The in-pocket structure has been solved by crystallography[13], while the intrahelical structure is more transient and, unlike Fpg, has not been captured yet. Disulfide cross-linking technique has been used to trap the putative intermediates of base flipping in hOGG1; however, as discussed above, the Cys149-Cyt crosslink sterically prevents the unopened conformation, while another crosslink which is distal from the target base pair has an artificial effect that biases the flipped base into the active site[32], and thus a fully intrahelical state has not been observed. Therefore, to accomplish our goal we first computationally modeled an intrahelical structure.

The published crystal structures of hOGG1, as well as the intrahelical Fpg structure, provide valuable insights into the conformation of the intrahelical hOGG1–DNA complex. In the hOGG1–DNA structures the "NNN" motif (containing residues Asn149, Asn150 and Asn151) occupies the void left by the everted base, and the Arg204 contacts the Watson-Crick face of the orphan cytosine[13, 31-33, 40, 107]. On the other hand, the apo structure of hOGG1 has the NNN motif and the Arg204 retracted, while the overall fold does not show significant difference from the bound structures[108]. We hypothesize that in the transient state in which hOGG1 interrogates an unopened target base pair, hOGG1 has not yet adjusted to the bound conformation that is seen in those everted hOGG1-DNA structures, but instead remains an apo-like conformation, with NNN motif retracted from the DNA helix and the Arg204 disengaged from the cytosine; otherwise, the target base pair would be disrupted. This assumption is reinforced by the observation that the Fpg–DNA complex shows similar structural differences between its everted state and intrahelical state, both of which have been crystallized. When Fpg binds to the damaged DNA and the oxoG is flipped out from the duplex, Met77 (the amino acid numeration for Fpg from *Geobacillus stearothermophilus*) intercalates into the DNA void left by oxoG, and Arg112 (the counterpart of Arg204 in hOGG1) interacts with the Watson-Crick edge of the orphan cytosine[74]. However, in the structure of Fpg bound to an intact, unopened DNA (stabilized by a disulfide crosslink between Fpg and DNA), the loop containing Met77 retracts from the DNA, and Arg112 also loses its contact with the orphan cytosine[30]. Furthermore, stopped-flow kinetics experiments also suggested that the void-filling of the hOGG1 residues into DNA likely occurs after the target base is everted[28]. In addition, the intrahelical Fpg-DNA complex showed a significant buckling at the target base pair induced by a wedge Phe114 residue[30], and since hOGG1 also has a counterpart wedge residue (Tyr203), target base pair buckling is also expected in the intrahelical hOGG1-DNA complex. Stopped-flow kinetics experiments also indicated destabilization of Watson-Crick or stacking interactions in the early stage nonspecific hOGG1-DNA complex[28]. Therefore, the intrahelical state model we created in this work is an apo-like hOGG1 (with the NNN motif retracted and the Arg204 disengaged from the cytosine of the target pair) interrogating an unopened but buckled base pair in a duplex DNA.

The intrahelical model was generated from 2I5W, whose target base is the least everted among all the hOGG1–DNA crystal structures. To obtain an unopened target base pair, the everted target base was forced into the helix to pair with the opposite C; to make space, the protein was simultaneously forced to adopt an apo conformation, using crystal structure 1KO9[108] as reference (see Methods for details). To test the stability of the restored target base pair, we calculated Watson-Crick hydrogen bond distances during 2ns unrestrained MD following the equilibration. No tendency of base opening was detected (Figure 3-7). Similar target pair stability was observed in the MD simulation starting from the crystal structure (2F5O) of an intrahelical Fpg complex. Around -20 degree of buckling was observed for the hOGG1 complexes, which is comparable with the buckling during the unrestrained MD on the intrahelical Fpg complex (Figure 3-8); other parameters of the target pair, including shear, stretch, propeller and opening, are also comparable between the hOGG1 and Fpg complexes (Table 3-1). This may due to the similarity in local geometry near the target base pair in these two enzymes. As shown in Figure 3-9, in the intrahelical structures of hOGG1, the target and 3' pairs are buckled, probably by the wedge residue, in a similar way as that in Fpg; while Arg204 and the NNN motif in hOGG1 are retracted from DNA, similar to Arg111 and Met76 in Fpg. Therefore, the intrahelical hOGG1 models we built have similar properties and stability as compared comparing to the Fpg simulation starting from the intrahelical crystal structure 2F5O. Overall, the built intrahelical hOGG1 structures appear to be reasonable models for base flipping path generation in this work.

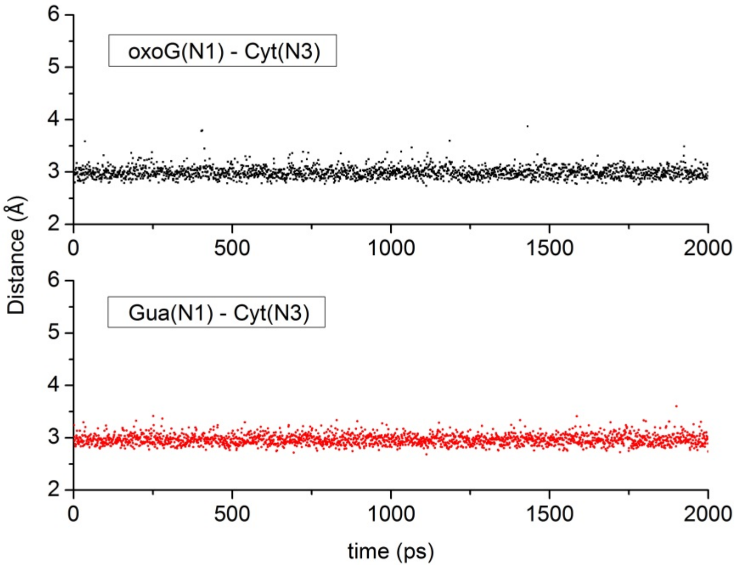


Figure 3-7. Distance of hydrogen bond oxoG/G(N^1) - C(N^3) of the target pair during the 2ns unrestrained MD of the equilibration, respectively.

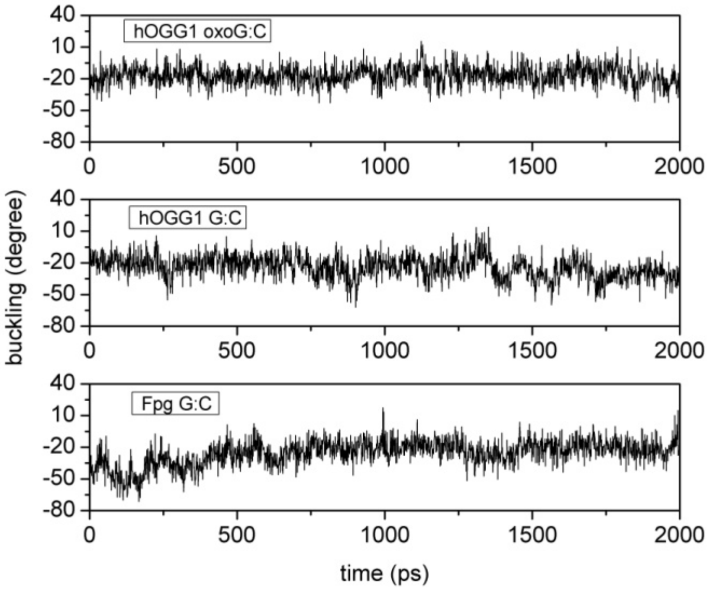


Figure 3-8. Comparison of buckling of the target pair in the hOGG1 intrahelical models and in the Fpg model during the 2ns unrestrained MD of the equilibration.

Table 3-1. Parameters of the target base pair in the intrahelical hOGG1 and Fpg systems.

	Shear	Stretch	Stagger	Propeller	Opening
hOGG1 oxoG-complex	0.0°	0.0°	0.5°	4.3°	-1.7°
hOGG1 G-complex	0.0°	-0.1°	0.1°	16.6°	-2.0°
Fpg G-complex	0.2°	0.1°	0.5°	9.4°	2.0°

Values were averaged over 2ns unrestrained MD equilibration of each system.

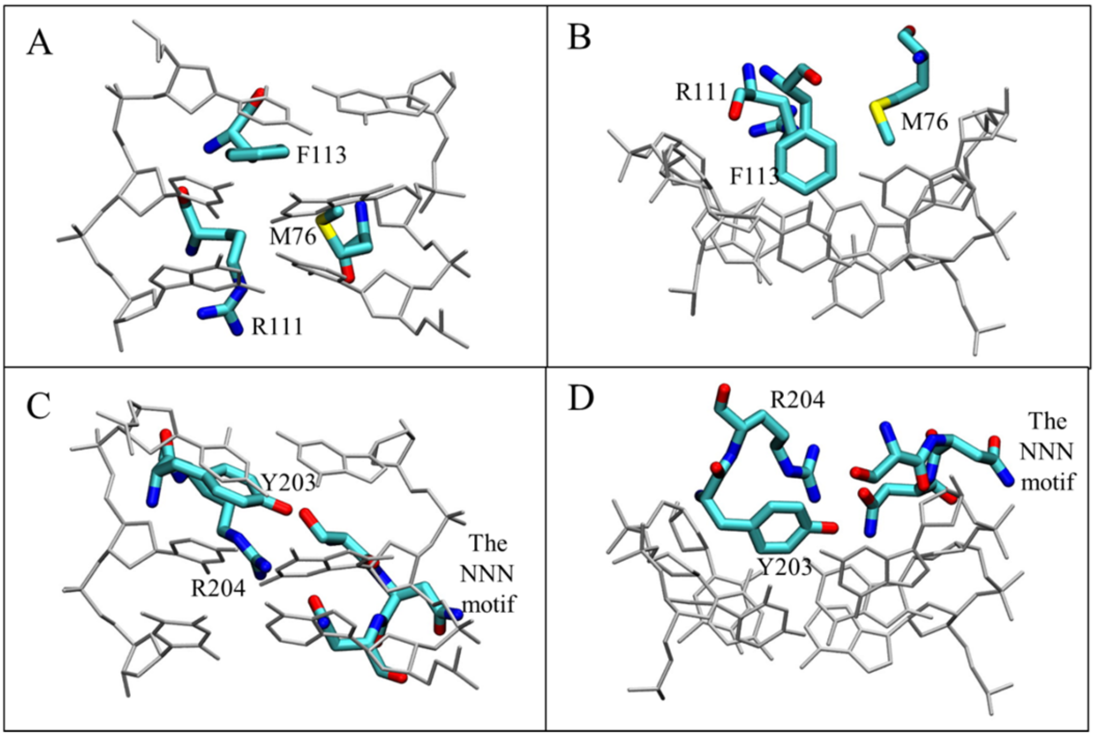


Figure 3-9. Comparison of intrahelical endpoint structures of Fpg (A,B) and hOGG1(C,D). B and D are same structure as A and C, respectively, but viewed from the top of the DNA, highlighting the depth of insertion. The DNA is in grey and the inserting residues are colored by atom.

3.3.3. Validation of the PNEB paths against crystal structure

As discussed above, the 1YQK crystal structure was reserved and not used as a starting structure for G eversion (see Methods); instead, it was used to compare with the intermediates of the generated PNEB path, so as to validate our model. The crystal structure 1YQK approximates a relatively stable intermediate for an everted G in the presence of a proximal crosslink[31]; but it is unclear whether the G would still stay in the "exo-site" of a wild type hOGG1 without the crosslink. Remarkably, this structure is sampled in our PNEB path (Figure 3-10) as well as in our umbrella sampling simulations (see discussions below), suggesting that the exo-site state is indeed an intermediate state in the G-flipping path in the absence of artificial crosslinks. This result also validates that our methods can reproduce experimentally verifiable intermediate structures during the base flipping process.

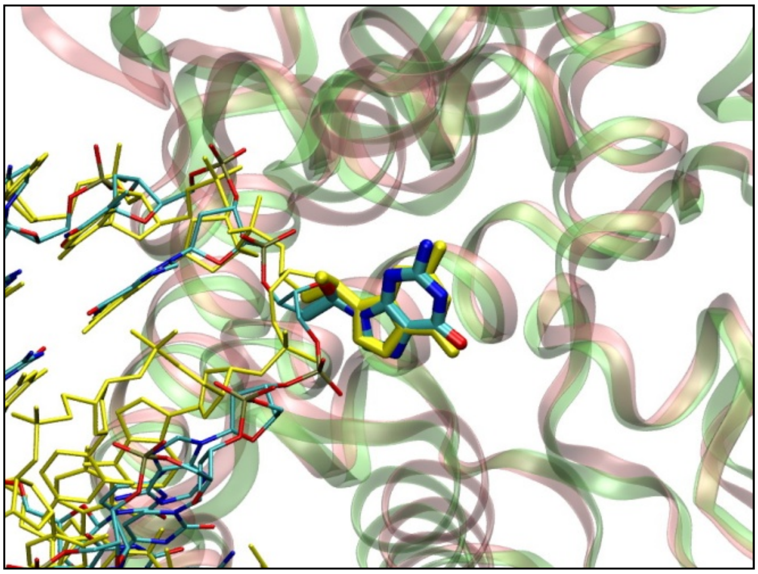


Figure 3-10. Superposition of the crystal structure 1YQK (in yellow and pink) and the structure sampled during the simulated base flipping (the DNA is colored by atom and the protein is in green). Both structures were superimposed by the heavy atoms of the protein part.

3.3.4. Free energy profiles of eversion of oxoG and G in hOGG1

To understand the energetic and structural basis of oxoG recognition, we compare the free energy profiles along the eversion pathways of oxoG and G (Figure 3-11) and investigate the discriminatory behavior of hOGG1 against oxoG and G. In Figure 3-11, the intra-helical position is around 1 Å of the eversion distance and the active site of hOGG1 is around 18 Å. The oxoG-complex favors the extrahelical endpoint by about 12 kcal/mol with respect to its intrahelical endpoint; for the G-complex, the “in-pocket” state is energetically similar to the intra-helical state. In addition, the energetic barrier to eversion of oxoG is only ~4 kcal/mol, comparable to what we observed in the Fpg-oxoG complex[38], while G has to overcome a higher energetic barrier (~7–8 kcal/mol) during the eversion. These overall energetic differences between the G and the oxoG eversion are consistent with oxoG being the cognate substrate for hOGG1 whereas G is not. The free energy profiles show significant differences between oxoG and G in four positions along the base flipping paths: 3–6 Å, 9–11 Å, 14–17 Å, and 17–18 Å of the eversion distance, respectively, and indicate multiple intermediate states of energetic and, presumably, structural or chemical differences between the oxoG and the G system, through which we may understand how damage recognition occurs. The details of these four states (labeled in Figure 3-11) are discussed below.

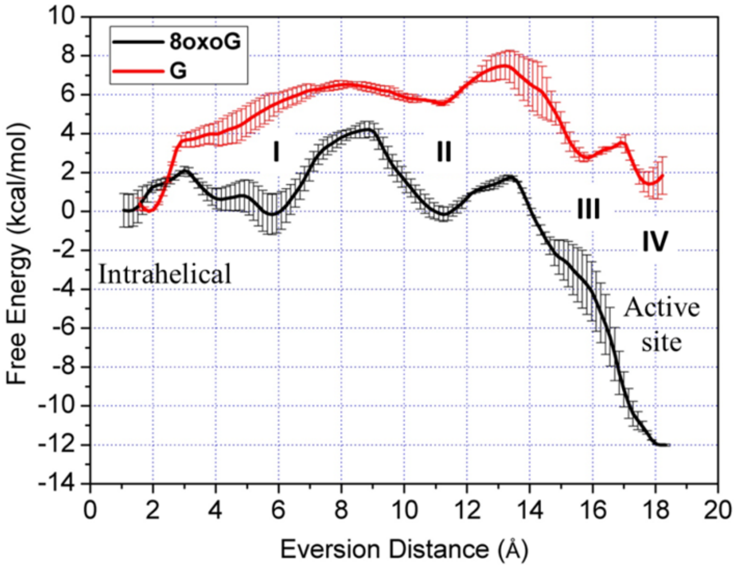


Figure 3-11. Comparison of free energy profiles for base eversion in the oxoG complex (black) and the G complex (red). Four potential oxoG-recognizing steps are labeled as I, II, III, and IV. The error bars reflect the difference between two independent US runs.

3.3.5. Initial opening of oxoG and G

The opening of the oxoG:C base pair occurs at 3 Å of the eversion distance (where oxoG loses all hydrogen bonds with the opposite C), and it requires only 1–3 kcal/mol (Figure 3-11), probably due to the destabilizing effects induced by the Tyr203 wedge, as observed in the intrahelical model. Similarly, our ongoing work on Fpg indicates that the wedge residue of Fpg can destabilize the intrahelical state (Kuznetsov *et al*, in preparation). Breaking the G:C pair costs comparable energy (3–4 kcal/mol), and no significant distinguishing interactions are apparent prior to base pair breaks. Notably, the energy cost for the oxoG:C opening in hOGG1 is comparable to that in Fpg,[38] and is comparable to the energy barriers of hOGG1/Fpg sliding along DNA (0.5 kcal/mol for hOGG1 and 2 kcal/mol for Fpg),[36] but is significantly lower than G opening from a B-DNA, which has a ~10 kcal/mol barrier.[109] Thus, hOGG1 probably promotes base eversion to achieve quick damage recognition during fast sliding.

3.3.6. Step I: target base slightly opened

The free energy landscapes of the two systems start to diverge after the base pair is open, indicating initial damage recognition (Figure 3-11, step I). When oxoG disengages from the opposite cytosine and is slightly everted (3~6 Å of the eversion distance), it is stabilized by a hydrogen bond between the H^7 and the second 5’ phosphate (the phosphate of the 5’ nucleotide, hereafter referred to as p^1) (Figure 3-12). This specific interaction corresponds to the free energy minimum at 6 Å of the eversion distance; whereas for the G-complex, the same phosphate exerts electrostatic repulsion to the lone pair of electrons with on N^7, and this interaction contributes to the energy rise at 4~6 Å of the eversion distance. To our knowledge, the common feature of the active substrates for hOGG1 is the absence of an electronegative group (i.e. a lone pair of electrons or a chlorine atom) on the N^7 position,[35, 110] so it is likely that hOGG1 selects against the electronegative N^7 position by rejecting it with an unfavorable interaction. The data suggest that the discriminatory interaction with oxoG and G provided by p^1 constructs a checkpoint for oxoG damage which could be more efficient than the active site selection, since the former occurs when the base is merely slightly everted. To pass through this unfavorable state, G has to overcome a ~6 kcal/mol energetic barrier (Figure 3-11), a significant impediment during the fast lesion search by hOGG1, and thus G probably returns to the duplex after a slightly eversion while hOGG1 continues sliding. In addition, recent studies have indicated that a mismatch at the 5’-neighboring position strongly decreased the rate of oxoG removal,[111, 112] and here we suggest that those mismatches may interrupt the favorable oxoG/p^1 interaction by repositioning the 5’ DNA backbone and thus hinder the oxoG eversion. It should be noted that the base-p^1

interactions here require a bent DNA; in a standard, unbent B-DNA, such interaction can hardly be formed due to the larger distance between the base and p^1 as compared to that in hOGG1 (Figure 3-13). Thus, DNA bending may be part of the oxoG recognition mechanism.

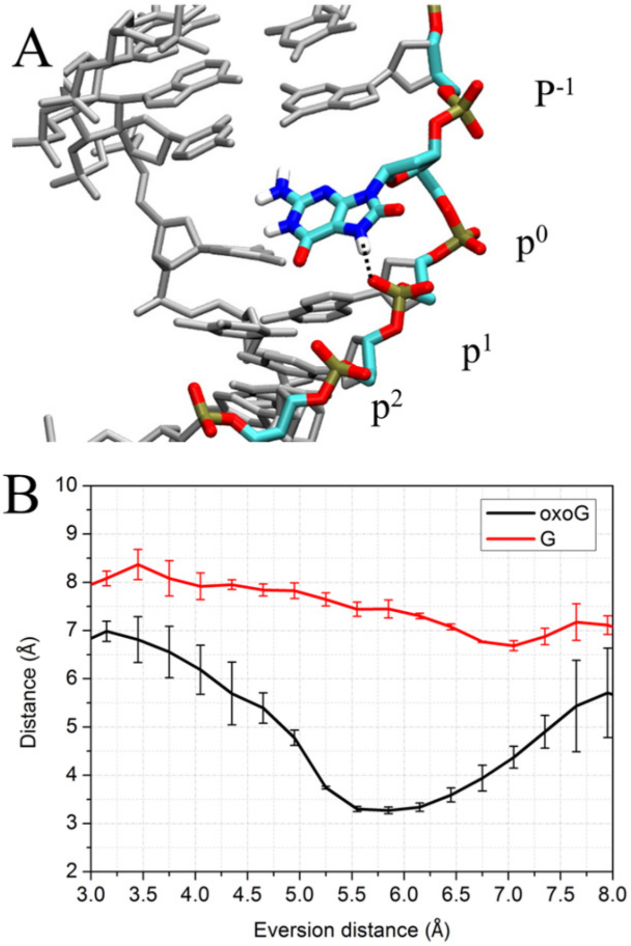


Figure 3-12. A, the DNA structure in step I. The black dots indicate the hydrogen bond between the N7 atom of oxoG and the OP2 atom of p1. Hydrogen atoms are not shown (the same hereinafter). B, comparison of distances between the OP2 atom of p^1 and the N7 atom of oxoG (black) and G (red). The error bars reflect the difference between two independent US runs.

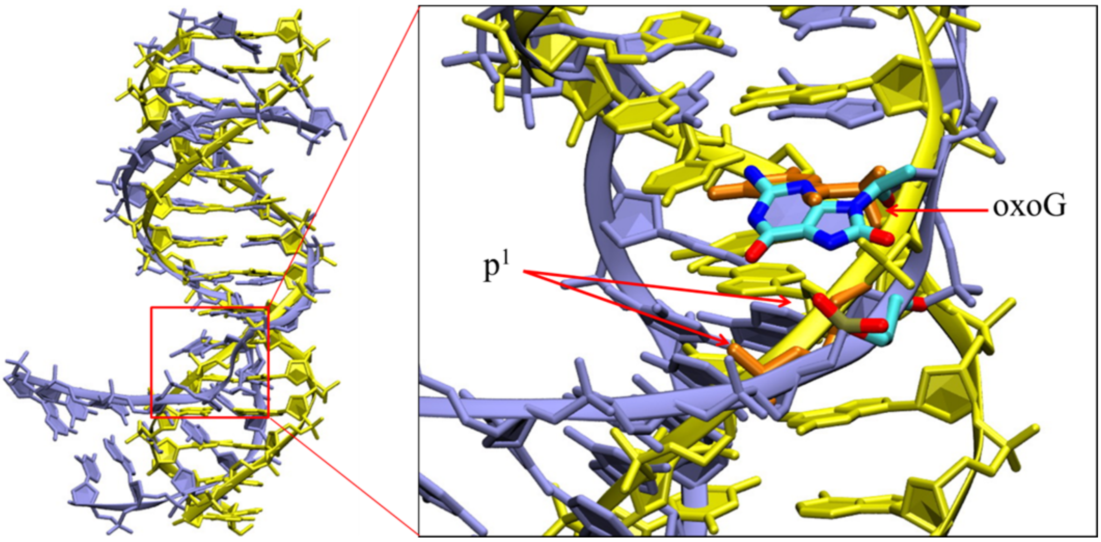


Figure 3-13. Superimposition of the bent DNA (purple) in Step I and a standard B-DNA with a same sequence (yellow), which was generated using 3DNA.[101] In the zoomed-in picture, the B-DNA is in yellow and its oxoG and p^1 are in orange; the DNA from our simulation is in purple and its oxoG and p^1 are colored by atom.

3.3.7. Step II: target base flips further into the major groove.

In our simulations, when oxoG is everted to a midway point (eversion distance ≈ 11 Å), it forms two hydrogen bonds to the p^1 (Figure 3-14A). G can also form the same hydrogen bonds to p^1, but the hydrogen-bonding distances between oxoG and p^1 are considerably shorter than those between G and p^1 (Figure 3-14B), indicating the p^1 keeps stronger interactions with oxoG than with G. This is probably because oxoG at the midway point is also stabilized by the hydrogen bond between O^8 of the oxoG and the backbone amide between His270 and Val269 (Figure 3-14A, C). On the other hand, His270 does not stabilize G in this state, since G does not possess an O^8. Thus, the energy minimum of G complex at the midway point is significantly shallower than that of oxoG complex (Figure 3-11), indicating another potential checkpoint for oxoG.

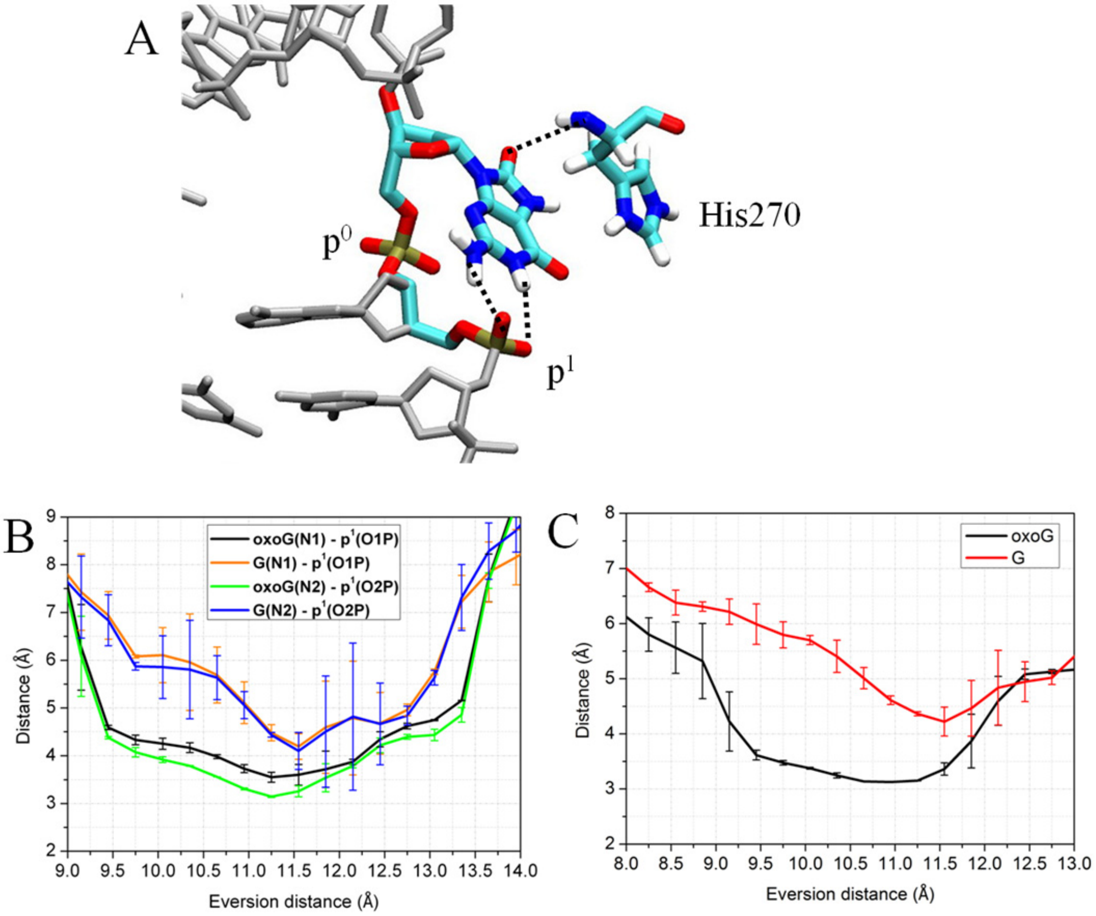


Figure 3-14. A, Structure of the oxoG-complex in step II. The hydrogen bonds stabilizing oxoG are depicted as dotted lines. B, comparison of distances between p^1 and the nitrogen atoms at the Watson-Crick face of the oxoG (black/green) and G (orange/blue). C, comparison of distances between the backbone nitrogen of H270 and the O8 of oxoG (black) or the H8 of G (red). The error bars reflect the difference between two independent US runs.

3.3.8. Step III: target base flips to the "exo-site".

As demonstrated in the crystal structure of the G-complex (1YQK), the exo-site is an alternative binding site outside the active pocket, and probably is a site the base would temporarily occupy before it enters the active site.[31] In our simulation, the base in the exo-site has an eversion distance of ~14.5–16.5 Å. As oxoG enters the exo-site, its O^8 atom forms a hydrogen bond to the ε-amino group of Lys249 (Figure 3-15A,B), and such O^8-Lys249 interaction may help to pull oxoG towards the active site. Since G does not possess an O^8 atom, the favorable O^8-Lys249 interaction is not observed in the G complex (Figure 3-15B) and thus helps to distinguish oxoG and G. It should be mention that the Lys249Gln and Lys249Cys/Cys253Lys mutants retain the ability of extruding oxoG to the active site[13, 34], suggesting that a lysine in position 249 may not be imperative for oxoG eversion; however, the side chain of Gln249, Cys249 or Lys253 in those mutants may still be able to form a hydrogen bond to the O^8 of oxoG in a similar way as does Lys249, facilitating the eversion of oxoG. In addition, pre-steady-state kinetic experiments indicated that the K249Q mutant of hOGG1 proceeds more slowly formation of the specific complex than does WT hOGG1,[113] in agreement with our results that Lys249 facilitates eversion of oxoG.

On the other hand, the intermediate structure in which G is in the exo-site has been captured in the crystal structure 1YQK[31], and we were able to sample this intermediate in independent simulations. In our simulation, the backbone amide of Ile152 forms a weak hydrogen bond to the N^7 atom of G in the exo-site, and the averaged distance is ~3.7 Å (Figure 3-15D), so it is considered a weak interaction. The same interaction has a similar distance (~4.2 Å) in 1YQK. In addition, our simulations showed that His270 interacts with the π-face of G in the exo-site (Figure 3-15C). In the crystal structure 1YQK, the high B-factor for the His270 side chain suggests that it only has a weak stacking interaction with G.[31] Nevertheless, when G breaks these weak interactions and enter the active site, it has to overcome a modest energetic barrier to leaving the exo-site (eversion distance ≈ 17 Å in Figure 3-11); whereas for oxoG, no obvious energy barrier is observed separating the exo-site state and the in-pocket state, probably because the active site starts to contact oxoG right after it loses the hydrogen-bonding interaction with Lys249 at ~ 16.7 Å of eversion distance (Figure 3-15B and Figure 3-16A, B).

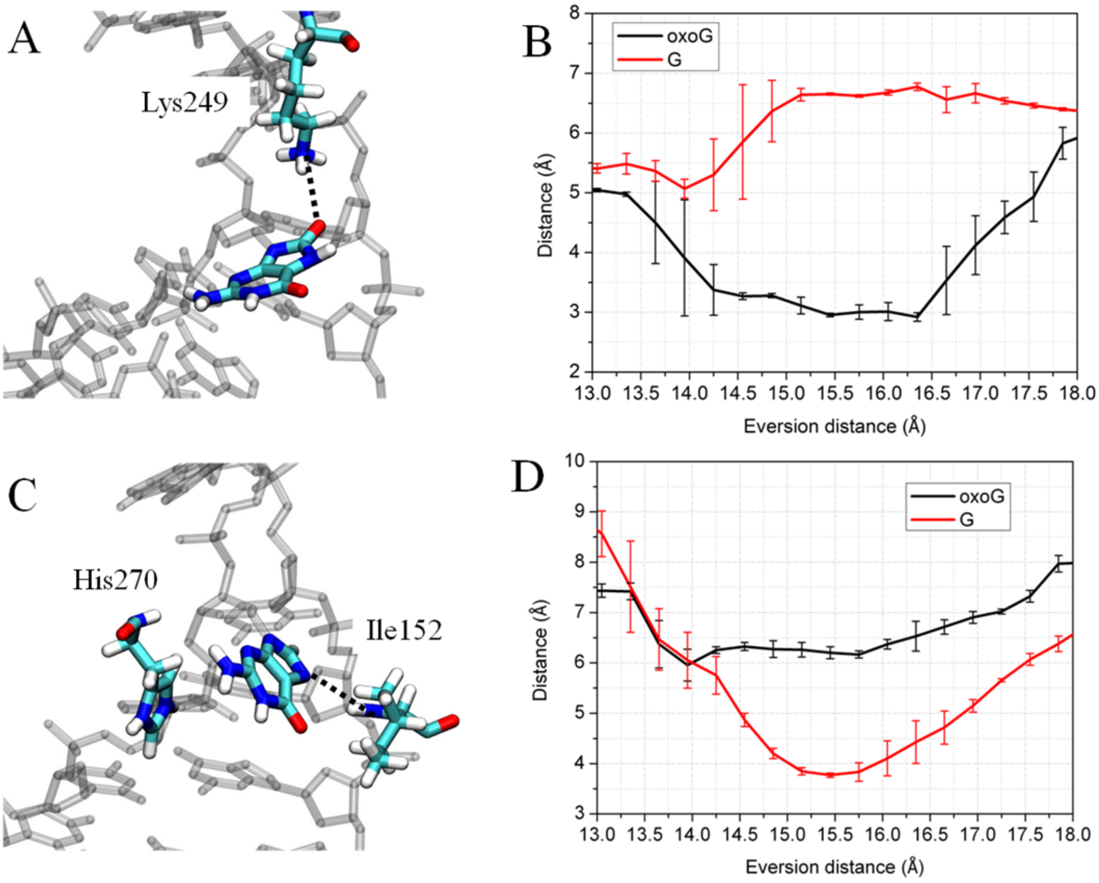


Figure 3-15. A, Structure of the oxoG-complex in step III. The oxoG-stabilizing hydrogen bond is depicted as a dotted line. B, comparison of distances between the epsilon-amino of Lys249 and the O8 of oxoG (black) or the H8 of G (red). C, the structure of the G complex in step III. The G was stabilized by the stacking interaction with His270 and a hydrogen bond with Ile152. D, comparison of distances between the backbone nitrogen of Ile152 and the N7 of G (red) or the N7 of oxoG (black). The error bars in B,D reflect the difference between two independent US

runs.

3.3.9. Step IV: target base enters the active pocket.

For oxoG, the free energy of the system drops drastically from step III (at the exo-site) to step IV (at the active site), suggesting that the oxoG is better stabilized in the active site than in the intermediate exo-site (Figure 3-11). This stable state has been captured in the crystal structure 1EBM, which has shown an important discriminatory hydrogen bond between the backbone carbonyl of Gly42 and the protonated N^7 of oxoG.[13] In our simulations, the region of 17~18 Å of eversion distance is considered as the in-pocket state because the hallmark Gly42/N7 hydrogen bond is formed in the oxoG-complex (Figure 3-16A, B). For the G complex, the in-pocket state was also captured experimentally (PDB 3IH7) with the help of a distal crosslink.[32] In the 3IH7 structure, G occupies almost the same position as does the in-pocket oxoG shown in 1EBM, regardless of the unfavorable interaction between the backbone carbonyl of Gly42 and the lone pair of electrons on N7 of G.[32] This is probably an artificial effect induced by the distal crosslink, since the unrestrained MD simulations starting from 3IH7 but without the artificial crosslink showed that G shifts away from its original in-pocket position.[32] In our simulations, the in-pocket endpoint for G was not generated from 3IH7; instead, it was generated from 1EBM with the in-pocket oxoG changed to G. Nevertheless, we also sampled that G is slightly retracted from the active site, and the backbone of Gly42 rotates to avoid contacting the lone pair of electrons on N^7 of the G (Figure 3-16C). These conformational changes of Gly42 and G have also been observed in the quantum mechanics/molecular mechanics (QM/MM) simulations previously performed to investigate the active site preference to oxoG and G.[31] Three other residues, Asp268, Gln315 and Phe319, contact oxoG and G in the active site in our simulations as well as in the crystal structures.[13, 32] The carboxyl of Asp268 contacts the N^1 atom of oxoG or G while Gln315 interacts with both the N^1 and the N^2 atoms, and Phe319 stacks with the six-member ring of oxoG (Figure 3-16A) or G. Therefore, Asp268, Gln315 and Phe319 do not provide direct damage recognition in the active site. Interestingly, even when G is forcibly presented in the active site, hOGG1 fails to catalyze base excision, and this catalytic checkpoint mechanism is still unclear[32]. We suggest that although Gly42 provides the sole interaction that structurally discriminates oxoG and G in the active site, it may be strong enough to keep G from entering the active site as deeply as does oxoG (Figure 3-16C). Experimental studies have shown that the Gln315Phe, Gln315Trp, Cys253Ile, and Cys253Leu mutations, which perturb the active site disposition of oxoG but not expel it altogether, can severely reduce the catalytic activity of hOGG1,[40, 114] thus suggesting that the catalysis of base excision by hOGG1 is geometrically demanding. Therefore, the unfavorable interaction of Gly42 may prevent G from achieving the optimal position for catalysis, and thus the active site constitutes the final damage checkpoint of hOGG1.

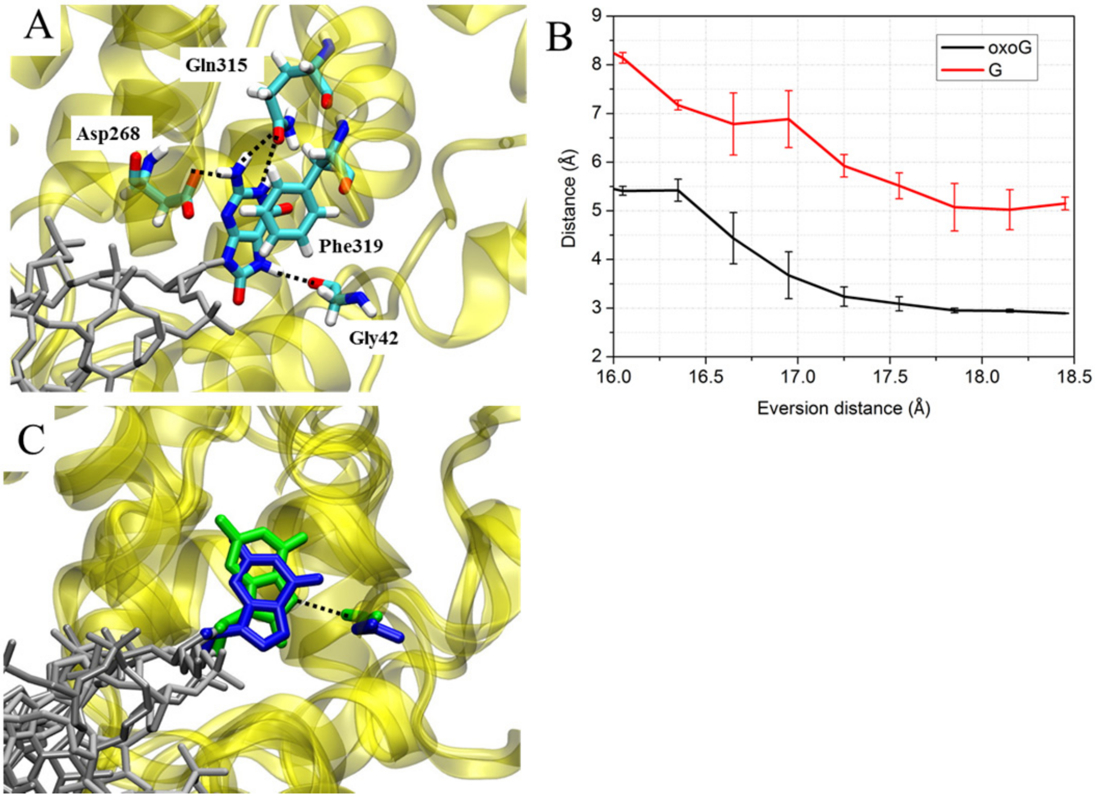


Figure 3-16. A. structure of the oxoG-complex in step IV. B, distance between the carbonyl of G42 and the N7 atom of G (red) or the N7 of oxoG (black). C, position of oxoG (green) and G (blue) in the active site, and the rest of DNA is colored in grey. The two structures are superimposed by the protein backbone (yellow). The error bars reflect the difference between two independent US runs.

3.3.10. Mechanism of oxoG recognition shared by hOGG1 and Fpg

Previous studies have found considerable similarities in damage recognition mechanism of hOGG1 and Fpg: they insert a bulky aromatic wedge into DNA and induce a kink to disrupt the stacking around the target base pair, and extrude oxoG into the active site where the N7/H7 of oxoG is recognized by the enzyme; they also insert several residues into the DNA to stabilize the extrahelical state, including an arginine residue that recognizes the Watson-Crick interface of the orphan cytosine.[13, 29] By comparing the data from this work and our previous work (Chapter 2), we found more significant evidence for them sharing a similar damage recognition mechanism. The free energy profiles for oxoG eversion in hOGG1 and Fpg show a similar pattern (Figure 3-17): they have two minima in the first half of eversion (Step I and II) and a global minimum at the extrahelical endpoint (step IV); between steps II and IV is an area with significant energy drop (step III). For both enzymes, these four steps correspond to four potential checkpoint for oxoG, because in each step specific interactions are made to the N7/H7 or the O8 of oxoG, which is the structural difference as compared to G. Notably, such specific interactions are similar in pattern between hOGG1 and Fpg (Figure 3-18): in step I, the slightly everted oxoG is stabilized by the N7/p^1 hydrogen bond; in step II, the Watson-Crick edge of oxoG hydrogen-bonds to p^1 while the O8 hydrogen-bonds to a residue (Asn173 in Fpg and H270 in hOGG1); in step III, the catalytic residue (Pro1 in Fpg and Lys249 in hOGG1) contacts the O8 of oxoG; in step IV, oxoG is specifically recognized by a hydrogen bond made to the protonated N7. Thus, although hOGG1 and Fpg are different in overall structure, they have similar mechanism for distinguishing between oxoG and G along the base eversion pathway.

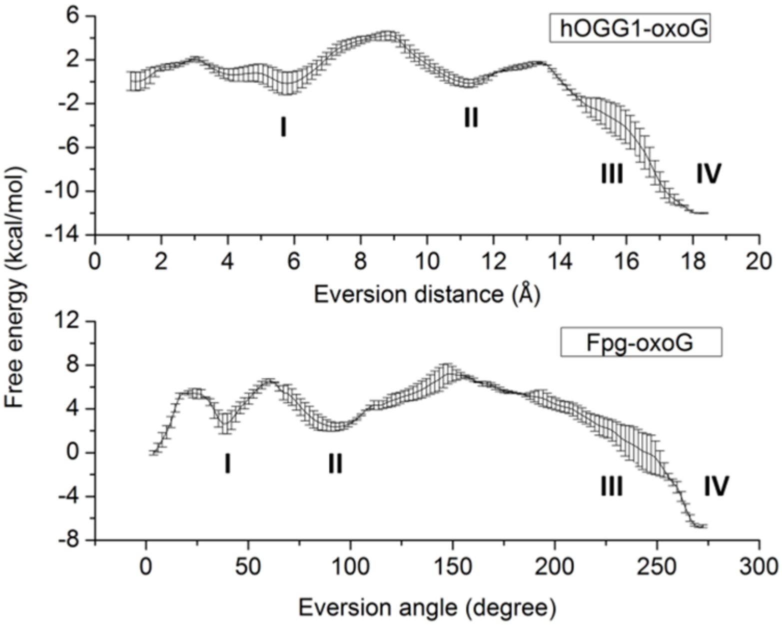


Figure 3-17. Free energy profiles for oxoG eversion in hOGG1 and Fpg. Four proposed steps of oxoG recognition are labelled as I, II, III and IV. The error bars reflect the difference between two independent US runs.

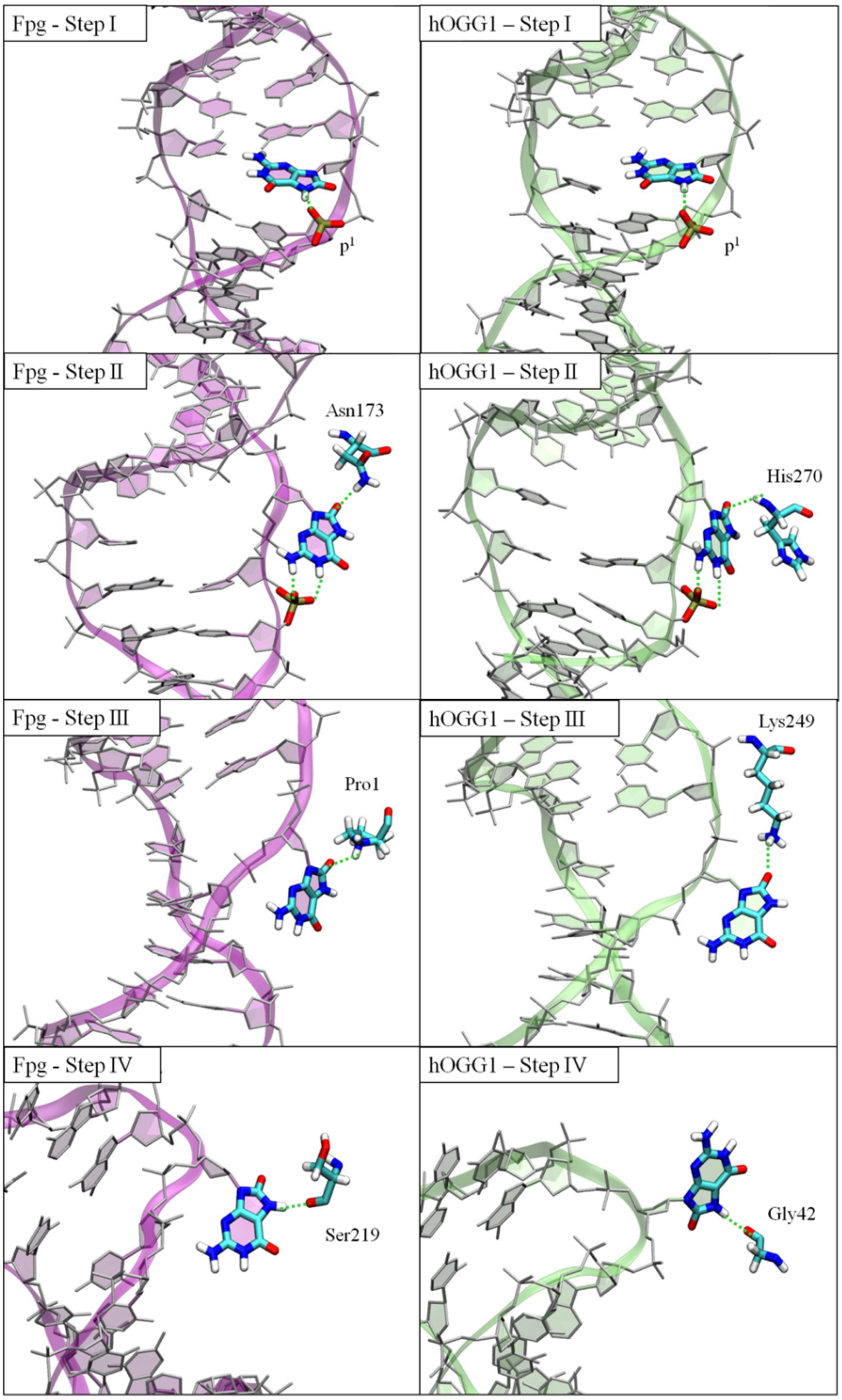


Figure 3-18. The structures of Step I, II, III, IV along the oxoG eversion pathway in Fpg (left) and hOGG1 (right). H-bonds that contact the O8 or the protonated N7 atom of oxoG are

indicated by green dots.

3.4. Conclusion

We have investigated the damage recognition mechanism of hOGG1 by energetically and structurally comparing the base eversion of oxoG and G. Based on the free energy profiles as well as the structural analysis that is linked to the energetic differences, we found that hOGG1 facilitates oxoG eversion while also hinders G eversion at multiple steps during base eversion, including stages earlier than suggested by crystal structures. Thus, during rapid scan of DNA damage, hOGG1 could discriminate oxoG from G even when the base is merely slightly everted. Interestingly, in comparable simulations for the Fpg system (showed in Chapter 2), the potential checkpoints for oxoG exhibit similar geometries compared to the hOGG1 system, suggesting that the damage recognition mechanism may be partly conserved between these two function analogs, even though they have no sequence or structural similarities.

4. Mutant studies reveal the role of Fpg's aromatic wedge in damage recognition

Acknowledgments

The material presented in this chapter contains data from the supplemental information of a manuscript that is in preparation by Nikita A. Kuznetsova, Christina Bergonzoc, Arthur J. Campbell, Haoquan Li, Grigory V. Mechetin, Carlos de los Santos, Arthur P. Grollman, Olga S. Fedorova, Dmitry O. Zharkov, and Carlos Simmerling. N.A.K. and G.V.M. performed kinetic experiments and analyzed data. C.B., H.L. and A.J.C. performed simulations and analyzed data. D.O.Z. and O.S.F. designed and supervised kinetic experiments and analyzed data. C.D.S. provided critical feedback on the design and analysis of the simulations. C.S. designed and supervised simulations and analyzed data. D.O.Z., A.P.G. and C.S. conceived and coordinated the study and wrote the manuscript.

4.1. Introduction

Glycosylases extrude their cognate DNA damage out of the DNA helix and insert it into the extrahelical active site where catalysis occurs. Base eversion is a crucial step in damage removal by glycosylases, but how glycosylases initially promote base eversion is still unclear. Glycosylases from different structural families appear to share a common feature that they use a bulky wedge residue to insert into the DNA and disrupt base stacking around the target base pair. Crystal structures of various glycosylase-DNA complexes show significant buckling of the base pairs around the inserted wedge, like Phe113 in *B.st* Fpg, Y203 in hOGG1 and L272 in human uracil-DNA glycosylase (UDG).[30, 33, 115]. Recently, Nelson *et al* found that Fpg, Nei and endonuclease III (Nth) may actively insert the wedge into DNA when they are scanning for damage because alanine substation of the wedge residue results in faster diffusion.[57] In addition, truncation of the wedge Phe can enhance Fpg's affinity to DNA by several folds.[71] Taken together, these findings suggest that the wedge in glycosylases may have a conserved function and wedge insertion is probably an active and crucial step in damage interrogation.

In the intrahelical structures for *B. st* Fpg, the aromatic wedge residue Phe113 is inserted in between the target base pair and the 3' pair, which are severely buckled.[30, 37] Such disruption of base pair stacking is probably induced by the unfavorable van der Waals interaction from the bulky wedge, since no buckling was observed in the F113A mutant Fpg bound with undamaged DNA.[71] Phe113 was also proposed to enforce DNA bending,[30] but this notion has been disproved since the F113A mutant of Fpg can bend DNA to a same degree as does WT.[71] The F113A mutation significantly diminishes the rate of oxoG excision but the AP lyase activity was not affected, suggesting F113A only impairs the early steps in oxoG repair which include base eversion and excision of the glycosidic bond.[71] Interestingly, in the crystal structure of F113A-oxoG, the oxoG is intrahelical but adopts a *syn* conformation and is disengaged from hydrogen bonding to the opposite C.[71] This *syn*-oxoG structure probably results from multiple cycles of DNA breathing during the long crystallization process. Overall, the structural data suggest that Fpg inserts Phe113 to disrupt the stacking around the target base pair and thus promotes base eversion, but such "insertion before eversion" event may be artificial, since those structures have an engineered disulfide crosslink that forces Fpg to bind to DNA and interrogate a specific base pair. A recent study shows that mutation of the wedge to alanine increases the rate of Fpg diffusion along DNA, suggesting that Fpg uses the wedge to probe for damage during processive sliding.[57]

Besides Phe113, Fpg also inserts two highly conserved residues, namely Met76 and Arg111, into DNA from the minor groove side (also known as "plugging"), forming important contacts at the protein-DNA interface. The plugging event probably occurs after base eversion, as suggested by pre-steady-state kinetic experiments.[27, 116] In the intrahelical structure, insertion of Met76 and Arg111 is blocked by the unopened target base pair and thus these two residues stay at the minor groove;[30] whereas in the extrahelical structure, Arg111 plugs into DNA to fill the void left by the everted oxoG and forms hydrogen bonds to the orphan C, and Met76 is positioned in the void adjacent to Arg111 (Figure 4-1).[29] The M76A-oxoG system has basically the same intrahelical structure as the V221P mutant (PDB 3GO8[37]), in which the Met76 side chain is intact.[71] Thus, truncation of the Met76 side chain does not alter the intrahelical structure of oxoG, suggesting Met76 probably does not direct the DNA conformation in the intrahelical state; on the other hand, truncation of the Met76 side chain alone is sufficient to convert the otherwise extrahelical oxoG to the intrahelical position, suggesting Met76, like V221, probably stabilizes the extrahelical state of oxoG. Therefore, it is likely that both Met76 and Arg111 are important residues for late stages of base eversion. Since Phe113 is positioned next to these two residues, it is possible that Phe113 also plays a role in stabilizing the extrahelical state by fine-tuning the inserting conformation of Met76 and Arg111.

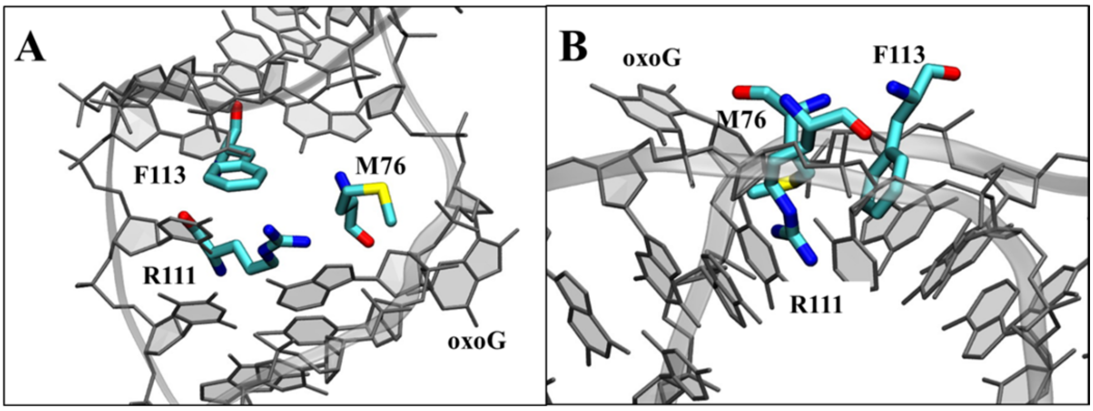


Figure 4-1. A, structure of the extrahelical state (PDB ID: 1R2Y[29]). DNA is colored in grey; Met76, Arg111 and Phe113 are colored by atom. B, same structure viewed from a different perspective, highlighting the depth of insertion.

The process of Fpg sliding over one base pair occurs in an ns timescale, and is much faster than base eversion which occurs in an ms timescale.[27, 36] Thus, Fpg has limited time to detect DNA damage during fast sliding. It raises questions of how Fpg probes the encountered bases, and how scrupulous the inspection is. We have been investigating these questions using stopped-flow kinetics and molecular dynamics. We have explored the damage probing events by *E. coli* Fpg and its two mutants F110A (wedge truncated) and F110W (wedge chaged to a fluorescent reporter tryptophan) (Kuznetsov *et al*, in preparation). We have found that Fpg sampled both oxoG and G during lesion search, as indicated by several fluorescently discernible stages in the F110W stopped-flow traces, whereas the F110A Fpg failed to process oxoG. Simulations of the oxoG and G eversion and energy decomposition analysis have suggested that the phenylalanine wedge in Fpg can destabilize the intrahelical state by inducing buckling to both base pairs flanking to the target base. We have also found that F110A Fpg maintained residual activity on abasic (AP) site, probably because no wedge insertion is required for directing the AP substrate to the catalytically competent conformation, as suggested by our simulations. Overall, our results have suggested that Fpg actively destabilizes its substrates by wedge insertion and efficiently samples normal and damaged bases for substrate discrimination, and this mechanism may be common for glycosylases with a bulky wedge residue. However, our previous simulations focused on the first half of the eversion only, and thus whether and how the wedge may play a role in late stages of base eversion is still unknown. The goal of this study is to provide insights into the base probing mechanism of Fpg, by understanding the function of the wedge residue in full oxoG eversion using MD simulations.

In this work, to isolate the function of Phe wedge, we investigated the base eversion in the F113A mutant of *B. st* Fpg. As discussed in Chapter 2, we have successfully used PNEB and umbrella sampling to generate the base eversion pathway and calculate the PMF along the pathway, and thus understood the oxoG/G discrimination mechanism. Here we again use umbrella sampling to explore the energetic and structural details of base eversion in the F113A mutant. By comparing the free energy profiles and the dynamics of the mutant and WT system, we can learn how Phe113 plays a role in damage recognition.

4.2. Methods

The Amber 12 package[75] was used for all simulations and structural analysis. The ff99SB force field[46] and the parmbsc0 corrections for nucleic acids[49] were used for all systems. The parameters for oxoG were obtained from Miller *et al.*[76] Two systems are under study: WT Fpg bound with an oxoG-containing DNA (the WT system) and F113A mutant Fpg bound with an oxoG-containing DNA (the F113A system). Both systems were solvated in 13,394 TIP3P[97] waters. For all simulations, SHAKE was used to constrain bonds with hydrogen.[79]

4.2.1. Mutation of F113A and equilibration

Previously we have generated the oxoG eversion pathway using PNEB (discussed in Chapter 2). To isolate the function of F113, we are interested in learning how the F113A mutation would affect the energy and structure along the WT path, so in this work we did not regenerate a new pathway for F113A using PNEB; instead, we made mutation on the structures from the WT PNEB pathway and performed equilibration, followed by umbrella sampling simulations. The initial F113A structures were mutated from the structures taken from the PNEB trajectories of oxoG eversion in WT Fpg. The side chain of Phe113 was deleted, keeping the backbone and the CB atom, and the rest of the substituting alanine was generated using the LEAP module. Thus, the F113A mutant system has the exact same number of explicit waters and same sequence as the WT system except for residue 113.

Each mutated structure (or window of umbrella sampling) was subjected to a 100ps equilibration in the NVT ensemble, with 100 $\text{kcal}\cdot\text{mol}^{-1}\cdot\text{A}^{-2}$ positional restraints on heavy atoms of the solute, allowing the water and hydrogens to equilibrate. The time step was set to 1 fs, and the temperature was maintained using the Berendsen thermostat[82] at 330K, reflecting the biological temperature of *Bacillus stearothermophilus*.

4.2.2. Umbrella sampling

Since we built the F113A system based on the WT PNEB pathway, and we need to compare the PMF between the WT and the mutant, we used the eversion angle as the reaction coordinate for oxoG eversion in F113A, which is same as that for the WT system as described in Chapter 2. The procedure for performing umbrella sampling simulations is basically the same as that for the WT, which has been described in Chapter 2. Two sets of independent umbrella sampling simulations were performed, each of which has 64 windows evenly spaced along the eversion angle. Each window was simulated in an NVT ensemble, with 0.183 $\text{kcal}\cdot\text{mol}^{-1}\cdot\text{deg}^{-2}$ umbrella restraints. The time step was set to 2 fs and the temperature was maintained at 330K using the Langevin thermostat[85] with 75.0 ps^{-1} collision frequency. The umbrella sampling simulations were performed over 2700ps, but only the last 2500ps was included in PMF calculation, and thus the first 200ps was an equilibration step allowing the system to relax. PMF was calculated using WHAM.[55] The error bars for PMF were estimated from the difference of the two independent runs.

To check how well the system is restrained by the reaction coordinate to the desired pathway, we measure the RMSD of the oxoG to estimate how far it shifts away from its original position. The RMSD was calculated in the same way as that in Chapter 2. As shown in Figure 4-2, the majority of the windows have RMSD lower than 2.5Å, comparable to that of the normal fluctuation of the intrahelical oxoG during equilibration. Some windows have higher RMSD (~4 Å), indicating potential shifting from the desired pathway. The windows with an averaged RMSD higher than 2.5 Å were listed in Table 2-1. It appears that the area around the eversion angle of 50°~80° have high RMSD, indicating possible structural shifting, which may cause inaccuracy in PMF calculation.

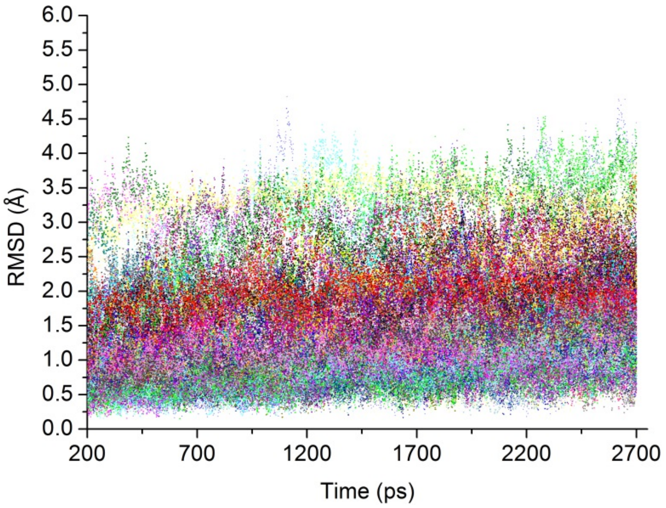


Figure 4-2. RMSD of the oxoG base in F113A Fpg druing the last 2.5 ns of umbrella sampling, which was included in PMF calculation. Structures were fit to the protein backbone and were referenced to the starting structure of each window. All of the 128 windows were included.

Table 4-1. Windows that have high RMSD of oxoG

Target eversion angle of the window, °	58.5	62.8	63.8	72.4	161.7	59.5	67.1	76.7	252
Average RMSD of oxoG, Å	2.78	2.80	2.83	2.85	2.90	3.01	3.15	3.24	3.34

4.3. Results and discussion

4.3.1. Phe113 may play a role in early and late stages of oxoG eversion

Along with our collaborators, we have been combining MD simulations and stopped-flow kinetics to investigate the role of the wedge phenylalanine of Fpg in oxoG recognition. We have shown that Fpg actively uses the wedge residue to destabilize the intrahelical state of oxoG by buckling the surrounding base pairs (Kuznetsov *et al*, in preparation). However, our previous simulation only focused on the early stages of the base eversion pathway, and thus it is unknown whether the wedge residue plays a role in the late stages of base eversion. In Chapter 2 we have introduced an improved reaction coordinate, the eversion angle, for base eversion in Fpg. This eversion angle can well describe the full base eversion from intrahelical to the active site, and thus we applied this reaction coordinate for sampling of the oxoG eversion in the F113A mutant. Figure 4-3 shows the free energy profiles for the complete oxoG eversion in WT and F113A Fpg. The two profiles are aligned by the energetic maximum at ~150° of eversion angle so as to highlight the free energy differences at the intrahelical and late extrahelical areas (discussed below). Overall, the PMF in this work shows that the F113A mutation eliminates the energetic difference between the intrahelical and extrahelical states and increases the energetic barrier to eversion, and thus it thermodynamically and kinetically hinders eversion of oxoG. The F113A mutation increases the energy barrier for opening the oxoG:C base pair, suggesting that Phe113 may help to promote initial oxoG eversion, and is consistent with our previous simulation and experimental data. In addition to the difference at the intrahelical region, the free energy profiles for the WT and the mutant are different at the second half of the path (from ~150° to ~275° of eversion angle), suggesting the F113A mutation may also affect the late stage of oxoG eversion. Therefore, Phe113 may play a role in both the intrahelical and the late extrahelical state of oxoG eversion.

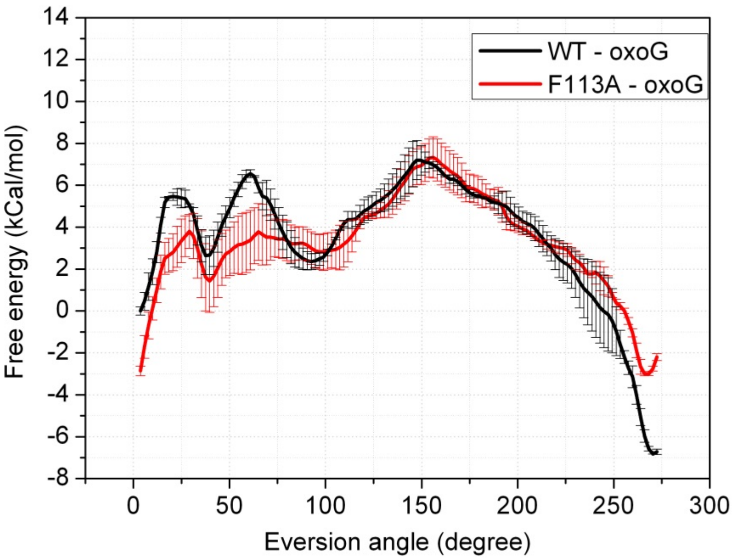


Figure 4-3. Free energy profiles for oxoG eversion in WT and F113A mutant of Fpg. The error bars reflect the difference between two independent US runs.

4.3.2. How does Phe113 promote eversion of oxoG?

The energy barrier for initial base opening in WT Fpg is 5.5 ± 0.5 kcal/mol (from 0° to 25° of eversion angle), whereas in the F113A mutant such barrier is 7 ± 1 kcal/mol. To explain this possible energetic difference, we inspect the intrahelical state of the oxoG eversion pathway and observed that the F113A mutation changes the intrahelical stacking around the target base pair. In WT Fpg, the oxoG:C pair and its 3' pair are buckled to opposite directions (Figure 4-4), which mirrors the significant separation of the buckling angles of these two pairs shown in Figure 4-5; whereas in the F113A mutant, the two base pairs are slightly buckled to the same direction (Figure 4-5, bottom panel) . This is consistent with the observation in the crystal structures of WT and F113A Fpg cross-linked to undamaged DNA (PDB ID: 2F5O and 4G4Q, respectively).[30, 71] In addition, in WT Fpg the rise between oxoG:C and the 3' pair is ~4 Å, and the F113A mutation shifts this value to 3.3 Å, close to the regular value in B-DNA (Figure 4-6).

Pairwise energy decomposition analysis indicates that the F113A mutation lowers the nonbonded energy between oxoG and two flanking base pairs at the eversion angle of 0°~20° (Figure 4-7), stabilizing the intrahelical state. These findings indicate that Phe113 disrupts the stacking around oxoG and thus destabilizes the intrahelical state of oxoG and promotes eversion. In the region where oxoG is slightly everted (around 30° to 80° of eversion angle), due to the relatively large uncertainty of the PMF, we cannot exactly quantify the energetic difference between the WT and mutant systems, and structural analysis shows no significant changes to the interactions that are made to oxoG.

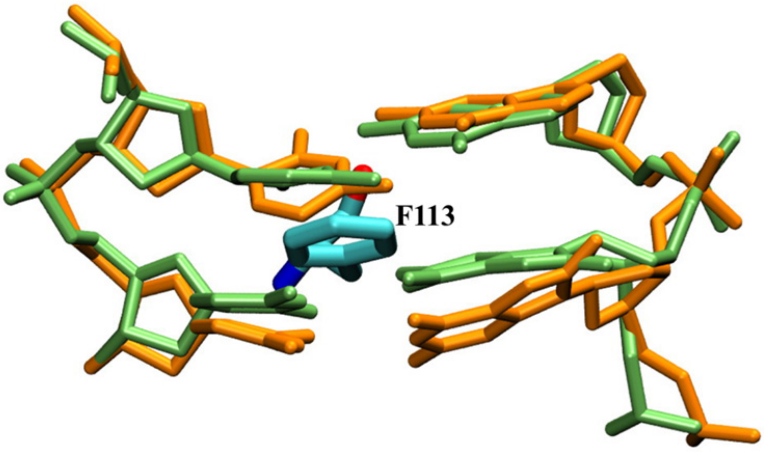


Figure 4-4. Overlap of the structures of the OG:C and 3’ pair in WT Fpg (orange) and F113A mutant (green). Phe113 is colored by atom.

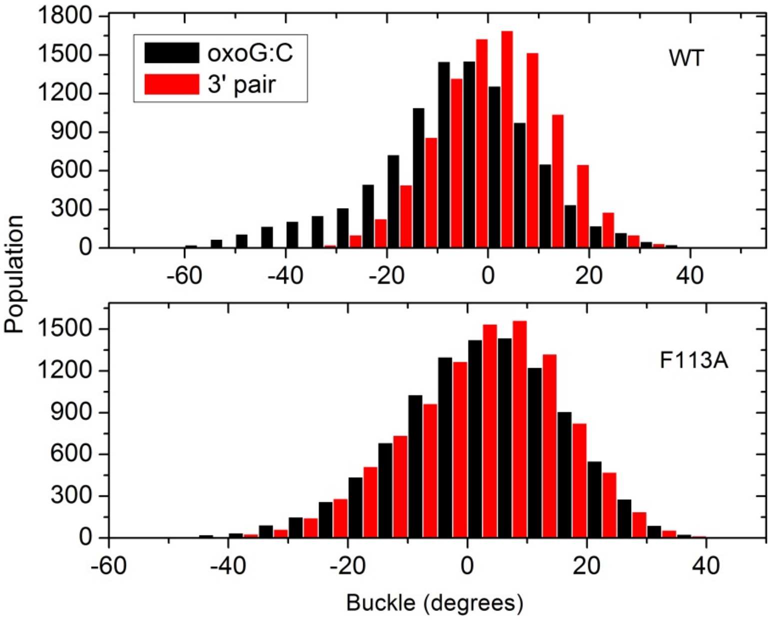


Figure 4-5. Comparison of buckling of the oxoG:C and the 3’ pair during the intrahelical stage for F113A (top) and WT Fpg (bottom), calculated from umbrella sampling windows with target values of eversion angle less than 10°.

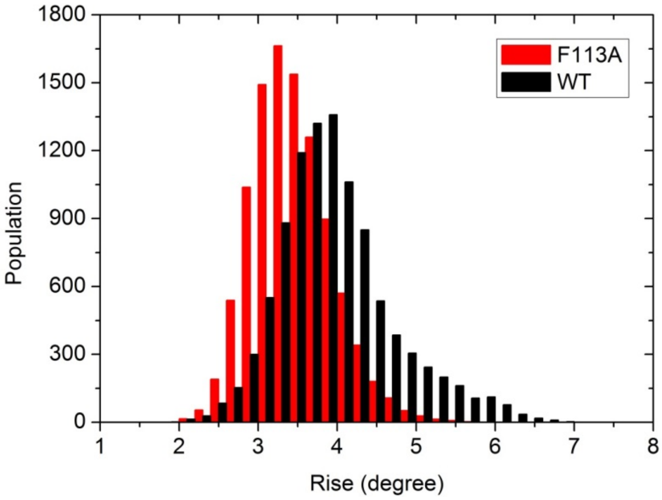


Figure 4-6. Comparison of rise between the oxoG:C and its 3’ pair during the intrahelical stage for F113A and WT Fpg, calculated from umbrella sampling windows with target values of eversion angle less than 10°.

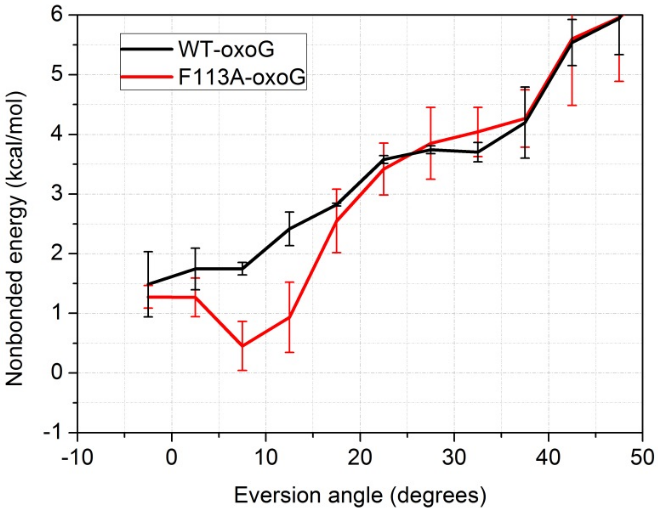


Figure 4-7. Comparison of pairwise nonbonded energies between oxoG and the two flanking base pairs in WT Fpg (black) and the F113A mutant (red). Phosphate groups were not included in calculation. The error bars reflect the difference between two independent US runs.

4.3.3. Phe113 facilitates late oxoG eversion by stabilizing the extrahelical state

The F113A mutation lowers the energetic difference between the maximum (eversion angle ≈ 150°) and the extrahelical endpoint by about 4 ± 2 kcal/mol (Figure 4-3), suggesting that Phe113 plays a role in the late stages of base eversion. We found several interactions that may contribute to the change of the PMF. First, in the late stages of oxoG eversion, Met76 makes weaker van der Waals interaction to oxoG in the F113A mutant than in the WT Fpg (Figure 4-8). This is probably because loss of the wedge creates a void so that Met76 can move away from the sugar-phosphate backbone of oxoG (Figure 4-9) and thus the van der Waals interaction between Met76 and oxoG is weakened. Consistent with this notion, the M76A mutation can convert the otherwise extrahelical oxoG to an intrahelical conformation, which has the same structure as the V221P mutant with an intact Met76, suggesting that Met76 side chain does not play an important role in directing the specific intrahelical conformation of DNA, but rather stabilizes the extrahelical state of oxoG, therefore loss of the Met76 side chain leads to an intrahelical oxoG.[71] Second, the interaction between oxoG and the second 5' phosphate (p^1) is weakened by the F113A mutation. As shown in Figure 4-10A, the nonbonded energy of oxoG- p^1 is lower in the WT than in the mutant in most areas after the eversion angle of 160°. Figure 4-10B indicates that the N2 atom of oxoG makes a weak hydrogen bond to the O1P atom of p^1 at the areas around the eversion angle of 175° and 235° in the WT system (averaged bond distance ≈ 3.5 Å) but such hydrogen bond is disrupted (averaged bond distance increased to ~4 Å) in the F113A mutant. Figure 4-11 shows that in the F113A mutant, Arg111 can move to the void otherwise occupied by the flanking wedge side chain, leaving more room for p^1 to move than in the WT. We thus suggest that p^1 in the F113A mutant may gain more flexibility and thus can move away from the oxoG base. Overall, the F113A mutation affects the residues that plug into the DNA when oxoG is extrahelical, namely Met76 and Arg111, weakening their favorable interaction with oxoG and thus destabilizes the extrahelical state. Therefore our findings indicate that Phe113 may play a role in facilitating oxoG eversion by stabilizing the extrahelical state.

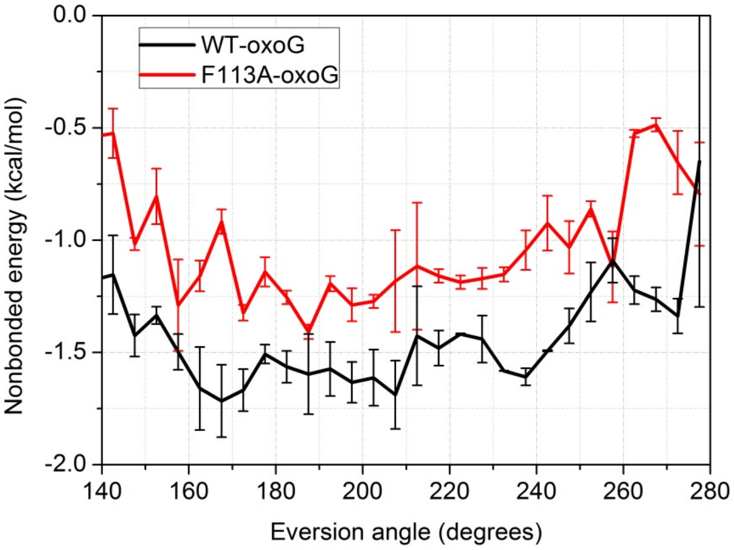


Figure 4-8. Comparison of Van der Waals energies between oxoG and Met76 in WT Fpg (black) and the F113A mutant (red). The error bars reflect the difference between two independent US runs.

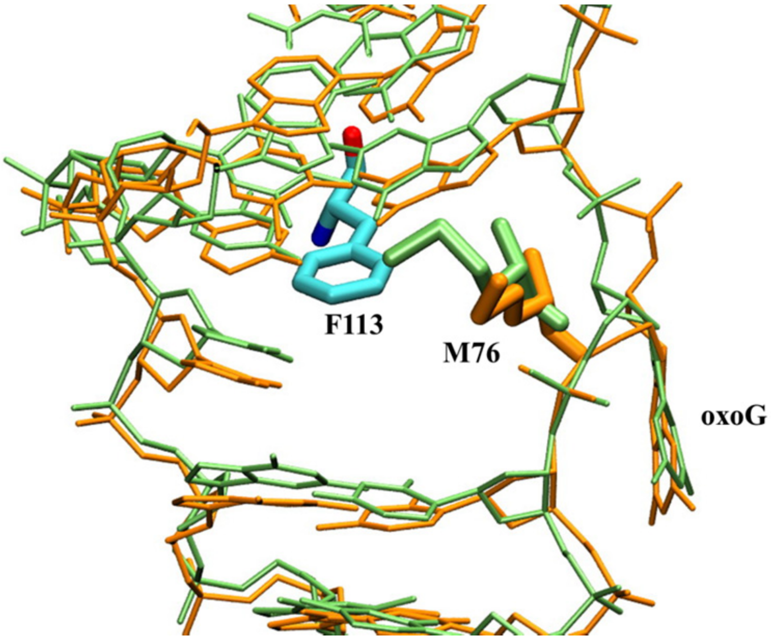


Figure 4-9. Overlap of the structures of the WT Fpg (orange) and the F113A mutant (green). Phe113 is colored by atom. The oxoG here has ~230° of eversion angle.

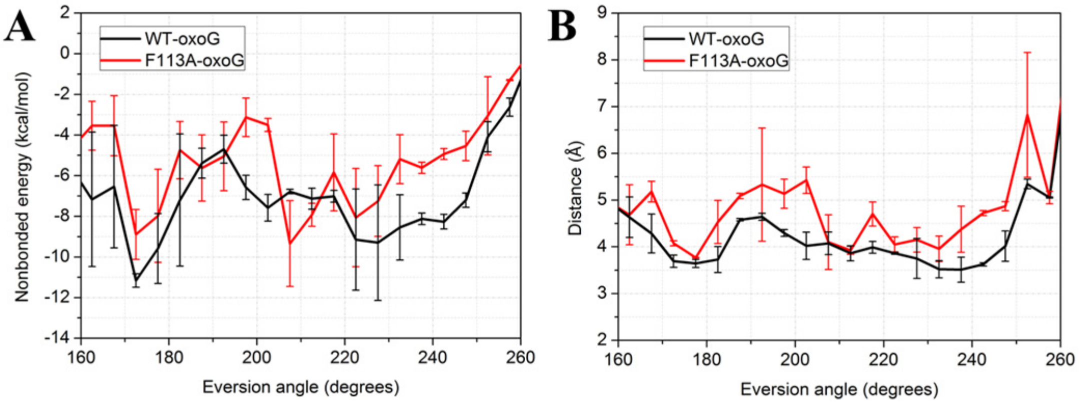


Figure 4-10. A, comparison of pairwise nonbonded energies between the oxoG base and p^1 in WT Fpg (black) and the F113A mutant (red). B, comparison of distances between N2 of oxoG and the O1P atom of p^1 in WT Fpg (black) and the F113A mutant (red). The error bars reflect the difference between two independent US runs.

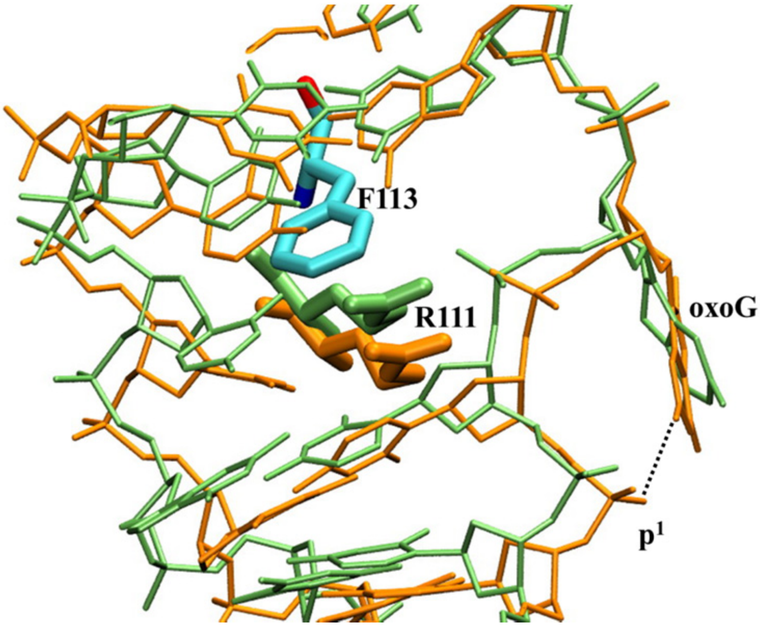


Figure 4-11. Overlap of the structures of the WT Fpg (orange) and the F113A mutant (green). Phe113 is colored by atom. The oxoG here has ~230° of eversion angle. The hydrogen bond between oxoG and p^1 is denoted as a dotted line.

4.4. Conclusion

In this work, we have compared the PMF for the full eversion of oxoG in the WT and the F113A mutant of Fpg and showed that eversion of oxoG is hindered by the mutation. We found that the F113A mutation probably affects both the intrahelical and the late extrahelical states of oxoG. Structural analysis and energy decomposition analysis indicates that loss of the wedge side chain strengthens interactions between oxoG and the flanking pairs by abating buckling of the target and the 3’ pair and reduces the rise between these two pairs to a regular value, and thus suggests that Phe113 in WT can destabilize the intrahelical state of oxoG and promote eversion. In addition, the F113A mutation also affects the plugging residues (Met76and Arg111) and thus weakens the stabilizing interactions that made to the extrahelical oxoG. Therefore, Phe113 probably plays a role in oxoG eversion by destabilizing the intrahelical state and stabilizing the extrahelical state of oxoG.

5. Summary

The present work involves research that covers repair of DNA damage, which is crucial to the survival of all living organisms.[2] Selecting oxoG as model DNA damage, we explored the mechanism by which DNA repair proteins process oxidatively damaged DNA. The proteins under study are Fpg and hOGG1, which are representative glycosylases from two different structural superfamilies and have a same function of removing oxoG from DNA. The mechanism of Fpg and hOGG1 efficiently recognizing oxoG from a sea of undamaged DNA bases was investigated in this research. The guiding hypothesis is that oxoG recognition by DNA glycosylases is not limited to the recognizing interactions in the enzyme's active site but, rather, is part of a dynamic process which is activated upon binding of glycosylases to damaged DNA, and involves multiple checkpoints for oxoG during eversion. The intermediates of the oxoG-recognizing process usually have high energies and low populations, and thus are difficult to probe with experimental approaches; therefore, MD simulations were employed to investigate the dynamic of the oxoG recognition events.

Eversion of oxoG and G in Fpg were structurally and energetically characterized, and the oxoG recognition mechanism was discussed in Chapter 2. The free energy profiles indicate that eversion of oxoG is more thermodynamically and kinetically favorable than that of G. Our simulations reveal specific interactions that facilitate eversion of oxoG and that hinder eversion of G during early stages of eversion. Four steps of oxoG/G discrimination were found along the base eversion pathway, three of which occur before the base enters the active site, suggesting that Fpg can recognize oxoG during early base eversion. The critical roles of Arg111, Asn173 and Arg263 in damage recognition were validated by biochemical mutation studies.

In Chapter 3, we have compared the oxoG recognition process during base eversion in hOGG1 and Fpg, and found significant similarities in the geometries of the oxoG checkpoints, even though these two enzymes are structurally unrelated. Fpg and hOGG1 recognize oxoG by contacting the structural differences between oxoG and G -- the O8 and the H7 atoms; and this damage recognition occurs during early base flipping before the base enters the active site. Therefore, hOGG1 and Fpg can "sense" the difference between oxoG and G when the base is only slightly opened from the intrahelical position, and thus allowing the enzyme to efficiently recognize damage during rapid sliding along DNA. Among the oxoG-recognizing interactions observed in our simulations, the one by the second 5' phosphate (p^1) is the earliest along the eversion pathway and is strictly conserved between Fpg and hOGG1, thus it would be interesting to test this interaction with experiments, such as to chemically substitute p^1 with a neutral methylphosphonate analog and then compare the base flipping in this mutant to that in the WT system.

Chapter 4 discusses the function of the aromatic phenylalanine wedge by investigating the consequences of removing Phe113 from Fpg. Active wedge insertion has been proposed as a critical early stage for oxoG repair by Fpg.[57, 71] Our simulation data confirm the important role of Phe113 in initial destabilization of the intrahelical base pair, and also indicate that Phe113 can help to stabilize the extrahelical endpoint.

In addition to oxoG, Fpg and hOGG1 can also excise 2,6-diamino-4-hydroxy-5-formamidopyrimidine (Fapy-G) and 4,6-diamino-5-formamidopyrimidine (Fapy-A),[117-119] which are common ring-opened lesions, and have notably different structure as compared to oxoG. It would be interesting to investigate how Fpg and hOGG1 recognize Fapy-G/Fapy-A and then compare to that for oxoG. The methods in the present work, including PNEB and umbrella sampling, could be employed to characterize the processes of Fapy-G/Fapy-A recognition and reveal the mechanism of lesion specificity.

References

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

2. Friedberg, E. C., Walker, G. C., and Siede, W. (1995) *DNA repair and mutagenesis*, American Society for Microbiology (ASM).

3. Lindahl, T. (1993) Instability and decay of the primary structure of DNA, *Nature 362*, 709-715.

4. Wood, R. D., Mitchell, M., Sgouros, J., and Lindahl, T. (2001) Human DNA repair genes, *Science 291*, 1284-+.

5. Halford, S. E., and Marko, J. F. (2004) How do site-specific DNA-binding proteins find their targets?, *Nucleic Acids Res 32*, 3040-3052.

6. Hu, T., Grosberg, A. Y., and Shklovskii, B. I. (2006) How proteins search for their specific sites on DNA: the role of DNA conformation, *Biophysical journal 90*, 2731-2744.

7. Barnes, D. E., Lindahl, T., and Sedgwick, B. (1993) DNA repair, *Curr Opin Cell Biol 5*, 424-433.

8. Sun, B., Latham, K. A., Dodson, M. L., and Lloyd, R. S. (1995) Studies on the catalytic mechanism of five DNA glycosylases. Probing for enzyme-DNA imino intermediates, *J Biol Chem 270*, 19501-19508.

9. Zharkov, D. O., Rieger, R. A., Iden, C. R., and 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, *Journal of Biological Chemistry 272*, 5335-5341.

10. Nash, H. M., Lu, R., Lane, W. S., and Verdine, G. L. (1997) The critical active-site amine of the human 8-oxoguanine DNA glycosylase, hOgg1: direct identification, ablation and chemical reconstitution, *Chem Biol 4*, 693-702.

11. Stivers, J. T., and Jiang, Y. L. (2003) A mechanistic perspective on the chemistry of DNA repair glycosylases, *Chem Rev 103*, 2729-2759.

12. David, S. S., and Williams, S. D. (1998) Chemistry of Glycosylases and Endonucleases Involved in Base-Excision Repair, *Chem Rev 98*, 1221-1262.

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

14. Gedik, C. M., and Collins, A. (2005) Establishing the background level of base oxidation in human lymphocyte DNA: results of an interlaboratory validation study, *The FASEB journal 19*, 82-84.

15. Lunec, J. (1998) ESCODD: European Standards Committee on oxidative DNA damage, *Free radical research 29*, 601-608.

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

17. Cheng, X., Kelso, C., Hornak, V., de los Santos, C., Grollman, A. P., and Simmerling, C. (2005) Dynamic behavior of DNA base pairs containing 8-oxoguanine, *J Am Chem Soc 127*, 13906-13918.

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

19. Lindahl, T., and Wood, R. D. (1999) Quality control by DNA repair, *Science 286*, 1897-1905.

20. Michaels, M. L., and Miller, J. H. (1992) The Go System Protects Organisms from the Mutagenic Effect of the Spontaneous Lesion 8-Hydroxyguanine (7,8-Dihydro-8-Oxoguanine), *J Bacteriol 174*, 6321-6325.

21. Barnes, D. E., and Lindahl, T. (2004) Repair and genetic consequences of endogenous DNA base damage in mammalian cells, *Annu Rev Genet 38*, 445-476.

22. Maga, G., Villani, G., Crespan, E., Wimmer, U., Ferrari, E., Bertocci, B., and Hubscher, U. (2007) 8-oxo-guanine bypass by human DNA polymerases in the presence of auxiliary proteins, *Nature 447*, 606-+.

23. Tchou, J., and Grollman, A. P. (1995) The Catalytic Mechanism of Fpg Protein - Evidence for a Schiff-Base Intermediate and Amino-Terminus Localization of the Catalytic Site, *Journal of Biological Chemistry 270*, 11671-11677.

24. Bailly, V., Verly, W. G., Oconnor, T., and Laval, J. (1989) Mechanism of DNA Strand Nicking at Apurinic Apyrimidinic Sites by Escherichia-Coli [Formamidopyrimidine]DNA Glycosylase, *Biochem J 262*, 581-589.

25. Bhagwat, M., and 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-Us 35*, 659-665.

26. Bjoras, M., Luna, L., Johnsen, B., Hoff, E., Haug, T., Rognes, T., and Seeberg, E. (1997) Opposite base-dependent reactions of a human base excision repair enzyme on DNA containing 7,8-dihydro-8-oxoguanine and abasic sites, *EMBO J 16*, 6314-6322.

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

28. Kuznetsov, N. A., Koval, V. V., Nevinsky, G. A., Douglas, K. T., Zharkov, D. O., and Fedorova, O. S. (2007) Kinetic conformational analysis of human 8-oxoguanine-DNA glycosylase, *J Biol Chem 282*, 1029-1038.

29. Fromme, J. C., and Verdine, G. L. (2003) DNA lesion recognition by the bacterial repair enzyme MutM, *J Biol Chem 278*, 51543-51548.

30. Banerjee, A., Santos, W. L., and Verdine, G. L. (2006) Structure of a DNA glycosylase searching for lesions, *Science 311*, 1153-1157.

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

32. Crenshaw, C. M., Nam, K., Oo, K., Kutchukian, P. S., Bowman, B. R., Karplus, M., and Verdine, G. L. (2012) Enforced presentation of an extrahelical guanine to the lesion-recognition pocket of the human 8-oxoguanine DNA glycosylase, hOGG1, *J Biol Chem*.

33. Banerjee, A., and Verdine, G. L. (2006) A nucleobase lesion remodels the interaction of its normal neighbor in a DNA glycosylase complex, *Proc Natl Acad Sci U S A 103*, 15020-15025.

34. Dalhus, B., Forsbring, M., Helle, I. H., Vik, E. S., Forstrom, R. J., Backe, P. H., Alseth, I., and Bjoras, M. (2011) Separation-of-function mutants unravel the dual-reaction mode of human 8-oxoguanine DNA glycosylase, *Structure 19*, 117-127.

35. Hamm, M. L., Gill, T. J., Nicolson, S. C., and Summers, M. R. (2007) Substrate specificity of Fpg (MutM) and hOGG1, two repair glycosylases, *J Am Chem Soc 129*, 7724-7725.

36. Blainey, P. C., van Oijen, A. M., Banerjee, A., Verdine, G. L., and Xie, X. S. (2006) A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA, *Proc Natl Acad Sci U S A 103*, 5752-5757.

37. Qi, Y., Spong, M. C., Nam, K., Banerjee, A., Jiralerspong, S., Karplus, M., and Verdine, G. L. (2009) Encounter and extrusion of an intrahelical lesion by a DNA repair enzyme, *Nature 462*, 762-U779.

38. Bergonzo, C., Campbell, A. J., de Los Santos, C., Grollman, A. P., and Simmerling, C. (2011) Energetic Preference of 8-oxoG Eversion Pathways in a DNA Glycosylase, *J Am Chem Soc 133*, 14504-14506.

39. Lee, S., Radom, C. T., and Verdine, G. L. (2008) AEI 14-Use of photocaging, cryotrapping and X-Ray crystallography to observe a very late intermediate in the lesion-extrusion pathway of hOGG1, *Abstr Pap Am Chem S 236*, -.

40. Radom, C. T., Banerjee, A., and Verdine, G. L. (2007) Structural characterization of human 8-oxoguanine DNA glycosylase variants bearing active site mutations, *J Biol Chem 282*, 9182-9194.

41. Karplus, M., and McCammon, J. A. (2002) Molecular dynamics simulations of biomolecules, *Nature structural biology 9*, 646-652.

42. Simmerling, C., and Kollman, P. (1996) MOIL-View: A program for visualization of structure and dynamics of biomolecules., *Abstr Pap Am Chem S 211*, 92-Comp.

43. Schrodinger, LLC. (2010) The PyMOL Molecular Graphics System, Version 1.3.

44. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics, *J Mol Graph Model 14*, 33-38.

45. 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., and Kollman, P. A. (1996) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules (vol 117, pg 5179, 1995), *Journal of the American Chemical Society 118*, 2309-2309.

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

47. Lange, O. F., van der Spoel, D., and de Groot, B. L. (2010) Scrutinizing Molecular Mechanics Force Fields on the Submicrosecond Timescale with NMR Data, *Biophysical journal 99*, 647-655.

48. Wickstrom, L., Okur, A., and Simmerling, C. (2009) Evaluating the Performance of the ff99SB Force Field Based on NMR Scalar Coupling Data, *Biophysical journal 97*, 853-856.

49. Perez, A., Marchan, I., Svozil, D., Sponer, J., Cheatham, T. E., 3rd, Laughton, C. A., and Orozco, M. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers, *Biophys J 92*, 3817-3829.

50. Bergonzo, C., Campbell, A. J., Walker, R. C., and Simmerling, C. (2009) A Partial Nudged Elastic Band Implementation for Use With Large or Explicitly Solvated Systems, *Int J Quantum Chem 109*, 3781-3790.

51. D.A. Case, T. A. D., T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and, and Kollman, P. A. (2011) AMBER 11, *University of California, San Francisco*.

52. Torrie, G. M., and Valleau, J. P. (1977) Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling, *Journal of Computational Physics 23*, 187-199.

53. Roux, B. (1995) The Calculation of the Potential of Mean Force Using Computer-Simulations, *Comput Phys Commun 91*, 275-282.

54. Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H., and Kollman, P. A. (1995) Multidimensional Free-Energy Calculations Using the Weighted Histogram Analysis Method, *J Comput Chem 16*, 1339-1350.

55. Kumar, S., Bouzida, D., Swendsen, R. H., Kollman, P. A., and Rosenberg, J. M. (1992) The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules .1. The Method, *J Comput Chem 13*, 1011-1021.

56. Li, H. J., Lai, C. T., Pan, P., Yu, W., Liu, N., Bommineni, G. R., Garcia-Diaz, M., Simmerling, C., and Tonge, P. J. (2014) A Structural and Energetic Model for the Slow-Onset Inhibition of the Mycobacterium tuberculosis Enoyl-ACP Reductase InhA, *ACS chemical biology 9*, 986-993.

57. Nelson, S. R., Dunn, A. R., Kathe, S. D., Warshaw, D. M., and Wallace, S. S. (2014) Two glycosylase families diffusively scan DNA using a wedge residue to probe for and identify oxidatively damaged bases, *Proc Natl Acad Sci U S A*.

58. Bjelland, S., and Seeberg, E. (2003) Mutagenicity, toxicity and repair of DNA base damage induced by oxidation, *Mutat Res-Fund Mol M 531*, 37-80.

59. Shibutani, S., Takeshita, M., and Grollman, A. P. (1991) Insertion of Specific Bases during DNA-Synthesis Past the Oxidation-Damaged Base 8-Oxodg, *Nature 349*, 431-434.

60. von Sonntag, C. (2006) *Free-radical-induced DNA damage and its repair*, Springer.

61. Kasai, H., and Nishimura, S. (1984) Hydroxylation of Deoxyguanosine at the C-8 Position by Ascorbic-Acid and Other Reducing Agents, *Nucleic Acids Research 12*, 2137-2145.

62. Gutteridge, J. M. C., and Halliwell, B. (2000) Free radicals and antioxidants in the year 2000 - A historical look to the future, *Ann Ny Acad Sci 899*, 136-147.

63. Brieba, L. G., Eichman, B. F., Kokoska, R. J., Doublie, S., Kunkel, T. A., and Ellenberger, T. (2004) Structural basis for the dual coding potential of 8-oxoguanosine by a high-fidelity DNA polymerase, *Embo J 23*, 3452-3461.

64. Friedberg, E. C., Walker, G. C., Siede, W., Wood, R. D., Schultz, R. A., and Ellenberger, T. (2006) DNA Repair and Mutagensis, *ASM Press Washington, DC.*

65. Hegde, M. L., Hazra, T. K., and Mitra, S. (2008) Early steps in the DNA base excision/single-strand interruption repair pathway in mammalian cells, *Cell Res 18*, 27-47.

66. Kunkel, T. A. (1999) The high cost of living. American Association for Cancer Research Special Conference: endogenous sources of mutations, Fort Myers, Florida, USA, 11-15 November 1998, *Trends Genet 15*, 93-94.

67. Savva, R., Mcauleyhecht, K., Brown, T., and Pearl, L. (1995) The Structural Basis of Specific Base-Excision Repair by Uracil-DNA Glycosylase, *Nature 373*, 487-493.

68. Nash, H. M., Bruner, S. D., Scharer, O. D., Kawate, T., Addona, T. A., Sponner, E., Lane, W. S., and Verdine, G. L. (1996) Cloning of a yeast 8-oxoguanine DNA glycosylase reveals the existence of a base-excision DNA-repair protein superfamily, *Curr Biol 6*, 968-980.

69. Dodson, M. L., Michaels, M. L., and Lloyd, R. S. (1994) Unified Catalytic Mechanism for DNA Glycosylases, *Journal of Biological Chemistry 269*, 32709-32712.

70. Fromme, J. C., and Verdine, G. L. (2002) Structural insights into lesion recognition and repair by the bacterial 8-oxoguanine DNA glycosylase MutM, *Nat Struct Biol 9*, 544-552.

71. Sung, R. J., Zhang, M., Qi, Y., and Verdine, G. L. (2013) Structural and Biochemical Analysis of DNA Helix Invasion by the Bacterial 8-Oxoguanine DNA Glycosylase MutM, *J Biol Chem 288*, 10012-10023.

72. Sung, R. J., Zhang, M., Qi, Y., and Verdine, G. L. (2012) Sequence-dependent structural variation in DNA undergoing intrahelical inspection by the DNA glycosylase MutM, *J Biol Chem 287*, 18044-18054.

73. Qi, Y., Spong, M. C., Nam, K., Karplus, M., and Verdine, G. L. (2010) Entrapment and structure of an extrahelical guanine attempting to enter the active site of a bacterial DNA glycosylase, MutM, *J Biol Chem 285*, 1468-1478.

74. Fromme, J. C., and Verdine, G. L. (2003) DNA lesion recognition by the bacterial repair enzyme MutM, *Journal of Biological Chemistry 278*, 51543-51548.

75. D.A. Case, T. A. D., T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A.W. Götz, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, R.M. Wolf, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, P.A. Kollman. (2012) AMBER 12, *University of California, San Francisco*.

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

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

78. Anandakrishnan, R., Aguilar, B., and Onufriev, A. V. (2012) H++3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations, *Nucleic Acids Res 40*, W537-W541.

79. Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes, *J Comput Phys 23*, 327-341.

80. Cheatham, T. E., Miller, J. L., Fox, T., Darden, T. A., and 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.

81. Darden, T., York, D., and Pedersen, L. (1993) Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems, *J Chem Phys 98*, 10089-10092.

82. Berendsen, H. J. C., Postma, J. P. M., Vangunsteren, W. F., Dinola, A., and Haak, J. R. (1984) Molecular-Dynamics with Coupling to an External Bath, *J Chem Phys 81*, 3684-3690.

83. Song, K., Campbell, A. J., Bergonzo, C., de los Santos, C., Grollman, A. P., and Simmerling, C. (2009) An Improved Reaction Coordinate for Nucleic Acid Base Flipping Studies, *J Chem Theory Comput 5*, 3105-3113.

84. Song, K., Kelso, C., de los Santos, C., Grollman, A. P., and 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-+.

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

86. Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) WebLogo: A sequence logo generator, *Genome Res 14*, 1188-1190.

87. Schneider, T. D., and Stephens, R. M. (1990) Sequence Logos - a New Way to Display Consensus Sequences, *Nucleic Acids Res 18*, 6097-6100.

88. Gilboa, R., Zharkov, D. O., Golan, G., Fernandes, A. S., Gerchman, S. E., Matz, E., Kycia, J. H., Grollman, A. P., and Shoham, G. (2002) Structure of formamidopyrimidine-DNA glycosylase covalently complexed to DNA, *Journal of Biological Chemistry 277*, 19811-19816.

89. Drsata, T., Kara, M., Zacharias, M., and Lankas, F. (2013) Effect of 8-Oxoguanine on DNA Structure and Deformability, *J Phys Chem B*.

90. Qi, Y., Spong, M. C., Nam, K., Karplus, M., and Verdine, G. L. (2010) Entrapment and Structure of an Extrahelical Guanine Attempting to Enter the Active Site of a Bacterial DNA Glycosylase, MutM, *Journal of Biological Chemistry 285*, 1468-1478.

91. Lukin, M., and de los Santos, C. (2006) NMR structures of damaged DNA, *Chemical Reviews 106*, 607-686.

92. Chen, L., Haushalter, K. A., Lieber, C. M., and Verdine, G. L. (2002) Direct visualization of a DNA glycosylase searching for damage, *Chem Biol 9*, 345-350.

93. Stivers, J. T. (2004) Site-specific DNA damage recognition by enzyme-induced base flipping, *Prog Nucleic Acid Res Mol Biol 77*, 37-65.

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

95. Asagoshi, K., Yamada, T., Terato, H., Ohyama, Y., Monden, Y., Arai, T., Nishimura, S., Aburatani, H., Lindahl, T., and Ide, H. (2000) Distinct repair activities of human 7,8-dihydro-8 oxoguanine DNA glycosylase and formamidopyrimidine DNA glycosylase for formamidopyrimidine and 7,8-dihydro-8-oxoguanine, *Journal of Biological Chemistry 275*, 4956-4964.

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

97. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of Simple Potential Functions for Simulating Liquid Water, *J Chem Phys 79*, 926-935.

98. Guex, N., and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling, *Electrophoresis 18*, 2714-2723.

99. Sali, A., and Blundell, T. L. (1993) Comparative Protein Modeling by Satisfaction of Spatial Restraints, *Journal of Molecular Biology 234*, 779-815.

100. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF chimera - A visualization system for exploratory research and analysis, *J Comput Chem 25*, 1605-1612.

101. Lu, X. J., and Olson, W. K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures, *Nucleic Acids Research 31*, 5108-5121.

102. Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005) The Amber biomolecular simulation programs, *J Comput Chem 26*, 1668-1688.

103. Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography, *Acta Crystallogr D 66*, 12-21.

104. Dashti, D. S., and Roitberg, A. E. (2013) Optimization of Umbrella Sampling Replica Exchange Molecular Dynamics by Replica Positioning, *J Chem Theory Comput 9*, 4692-4699.

105. MacKerell, A. D., and Banavali, N. K. (2002) Free energy and structural pathways of base flipping in a DNA GCGC containing sequence, *Journal of Molecular Biology 319*, 141-160.

106. Lee, S., Radom, C. T., and Verdine, G. L. (2008) Trapping and structural elucidation of a very advanced intermediate in the lesion-extrusion pathway of hOGG1, *Journal of the American Chemical Society 130*, 7784-+.

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

108. Bjoras, M., Seeberg, E., Luna, L., Pearl, L. H., and Barrett, T. E. (2002) Reciprocal "flipping" underlies substrate recognition and catalytic activation by the human 8-oxo-guanine DNA glycosylase, *J Mol Biol 317*, 171-177.

109. Giudice, E., Varnai, P., and Lavery, R. (2003) Base pair opening within B-DNA: free energy pathways for GC and AT pairs from umbrella sampling simulations (vol 31, pg 1434, 2003), *Nucleic Acids Res 31*, 2703-2703.

110. McKibbin, P. L., Kobori, A., Taniguchi, Y., Kool, E. T., and David, S. S. (2012) Surprising Repair Activities of Nonpolar Analogs of 8-oxoG Expose Features of Recognition and Catalysis by Base Excision Repair Glycosylases, *J Am Chem Soc 134*, 1653-1661.

111. Sassa, A., Beard, W. A., Prasad, R., and Wilson, S. H. (2012) DNA Sequence Context Effects on the Glycosylase Activity of Human 8-Oxoguanine DNA Glycosylase, *J Biol Chem*.

112. Kirpota, O. O., Endutkin, A. V., Ponomarenko, M. P., Ponomarenko, P. M., Zharkov, D. O., and Nevinsky, G. A. (2011) Thermodynamic and kinetic basis for recognition and repair of 8-oxoguanine in DNA by human 8-oxoguanine-DNA glycosylase, *Nucleic Acids Res*.

113. Kuznetsova, A. A., Kuznetsov, N. A., Ishchenko, A. A., Saparbaev, M. K., and Fedorova, O. S. (2014) Step-by-step mechanism of DNA damage recognition by human 8-oxoguanine DNA glycosylase, *Biochim Biophys Acta 1840*, 387-395.

114. Lukina, M. V., Popov, A. V., Koval, V. V., Vorobjev, Y. N., Fedorova, O. S., and Zharkov, D. O. (2013) DNA Damage Processing by Human 8-Oxoguanine-DNA Glycosylase Mutants with the Occluded Active Site, *Journal of Biological Chemistry 288*, 28936-28947.

115. Parikh, S. S., Mol, C. D., Slupphaug, G., Bharati, S., Krokan, H. E., and 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.

116. Koval, V. V., Kuznetsov, N. A., Zharkov, D. O., Ishchenko, A. A., Douglas, K. T., Nevinsky, G. A., and 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*, 926-935.

117. Krishnamurthy, N., Haraguchi, K., Greenberg, M. M., and David, S. S. (2008) Efficient removal of formamidopyrimidines by 8-oxoguanine glycosylases, *Biochemistry-Us 47*, 1043-1050.

118. Boiteux, S., Gajewski, E., Laval, J., and Dizdaroglu, M. (1992) Substrate-Specificity of the Escherichia-Coli Fpg Protein (Formamidopyrimidine DNA Glycosylase) - Excision of Purine Lesions in DNA Produced by Ionizing-Radiation or Photosensitization, *Biochemistry-Us 31*, 106-110.

119. Chetsanga, C. J., and 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.