Computational Analysis of the Mode of Binding of 8-Oxoguanine to Formamidopyrimidine-DNA Glycosylase†

Kun Song,‡ Viktor Hornak,§ Carlos de los Santos,|| Arthur P. Grollman,|| and Carlos Simmerling*,‡,⊥

Department of Chemistry, Department of Pharmacological Sciences, and Center for Structural Biology, Stony Brook University, Stony Brook, New York 11794-3400, and Computational Science Center, Brookhaven National Laboratory, Upton, New York 11973

Received February 24, 2006; Revised Manuscript Received July 11, 2006

ABSTRACT: 8-Oxoguanine (8OG) is the most prevalent form of oxidative DNA damage. In bacteria, 8OG is excised by formamidopyrimidine glycosylase (Fpg) as the initial step in base excision repair. To efficiently excise this lesion, Fpg must discriminate between 8OG and an excess of guanine in duplex DNA. In this study, we explore the structural basis underlying this high degree of selectivity. Two structures have been reported in which Fpg is bound to DNA, differing with respect to the position of the lesion in the active site, one structure showing 8OG bound in the *syn* conformation and the other in the *anti* conformation. Remarkably, the results of our all-atom simulations are consistent with both structures. The *syn* conformation observed in the crystallographic structure of Fpg obtained from *Bacillus stearothermophilus* is stabilized through interaction with E77, a nonconserved residue. Replacement of E77 with Ser, creating the Fpg sequence found in *Escherichia coli* and other bacteria, results in preferred binding of 8OG in the *anti* conformation. Our calculations provide novel insights into the roles of active site residues in binding and recognition of 8OG by Fpg.

DNA is a major target of oxidative damage which, in turn, has been linked to human diseases associated with aging, including cancer (*1*, *2*). 8-Oxoguanine (8OG)[1] (*3*) is one of the most common forms of oxidative DNA damage (*4*); failure to repair this lesion prior to DNA replication leads to G:C to A:T transversion mutations in bacterial and mammalian cells (*5*).

In *Escherichia coli*, Fpg (MutM), MutY, and MutT work in concert to counter the potentially deleterious effects of 8OG (*6*, *7*). Fpg is an 8-oxoguanine-DNA glycosylase/AP lyase which excises 8OG from oxidatively damaged DNA. This multistep process is initiated by nucleophilic attack on C1′ by the N-terminal proline of Fpg, forming a Schiff base intermediate. Proton abstraction leads to β-elimination of the 3′-phosphate, followed by hydrolysis of the Schiff base and δ-elimination, generating a single-base gap (*8*, *9*).

In this context, damage recognition refers to the process by which a specific lesion, 8OG, is selected for excision by Fpg in a vast sea of unmodified DNA. The problem may be considered in two parts: how the enzyme "finds" the lesion embedded in DNA and how the modified base, once encountered, is accommodated as a Michaelis complex in the enzyme's active site. The current model for damage recognition involves (a) nonspecific binding to duplex DNA, (b) scanning the groove(s) by facilitated diffusion, (c) damage recognition mediated by hydrogen bonds or thermodynamic instability, (d) formation of a transient enzyme−DNA complex, (e) eversion of damaged nucleotide from helix, and (f) binding in the active site pocket (*10*).

While the catalytic mechanism by which 8OG is excised from DNA has been extensively investigated by structural methods (*11*−*17*), little is known with regard to the mechanism by which Fpg recognizes its cognate lesion and whether discrimination between the oxidized base and guanine occurs during one or several stages of binding. 8OG differs from guanine at the N7 and O8 positions in that N7 is protonated and the C8 hydrogen is replaced with oxygen (Figure 1). It seems likely that these structural differences account, at least in part, for the ability of Fpg to recognize and bind 8OG in its active site.

The initial challenge in determining structural requirements for binding specificity through X-ray crystallographic analysis was presented by the difficulty in trapping the Fpg−DNA complex before base excision occurs. Two approaches have been used to circumvent this problem. Gilboa et al. used sodium borohydride to trap the Schiff base intermediate by forming a covalently bound intermediate (*15*). The resulting crystal structure of the *E. coli* Fpg−DNA complex lacked the damaged base; moreover, the βFα10 loop (residues 217−224) does not appear in the resulting electron density map. Thus, direct evidence for substrate recognition was lacking.

More recently, Fromme and Verdine reported the crystal structure of the Fpg−DNA complex containing the damaged

† Support for this project was provided by NIH Grants GM6167803 (C.S.), CA047995 (C.d.l.S.), and CA17395 (A.P.G.) and by National Computational Science Alliance (NCSA) Grant MCA02N028 (C.S.), which provided computational resources at NCSA. C.S. is a Cottrell Scholar of Research Corp.

* To whom correspondence should be addressed. E-mail: carlos.simmerling@stonybrook.edu. Phone: (631) 632-1336. Fax: (631) 632-1555.

‡ Department of Chemistry, Stony Brook University.

§ Center for Structural Biology, Stony Brook University.

|| Department of Pharmacological Sciences, Stony Brook University.

⊥ Brookhaven National Laboratory.

[1] Abbreviations: Fpg, formamidopyrimidine-DNA glycosylase from *E. coli*; 8OG, 8-oxo-7,8-dihydroguanine; *B. st.*, *Bacillus stearothermophilus*; MM-GBSA, molecular mechanics-generalized Born-surface area method; PMF, potential of mean force; WHAM, weighted histogram analysis method.

10.1021/bi060380m CCC: $33.50
Published on Web 08/16/2006

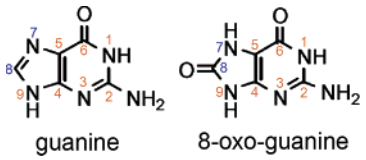

FIGURE 1: Structures of guanine and 8-oxoguanine.

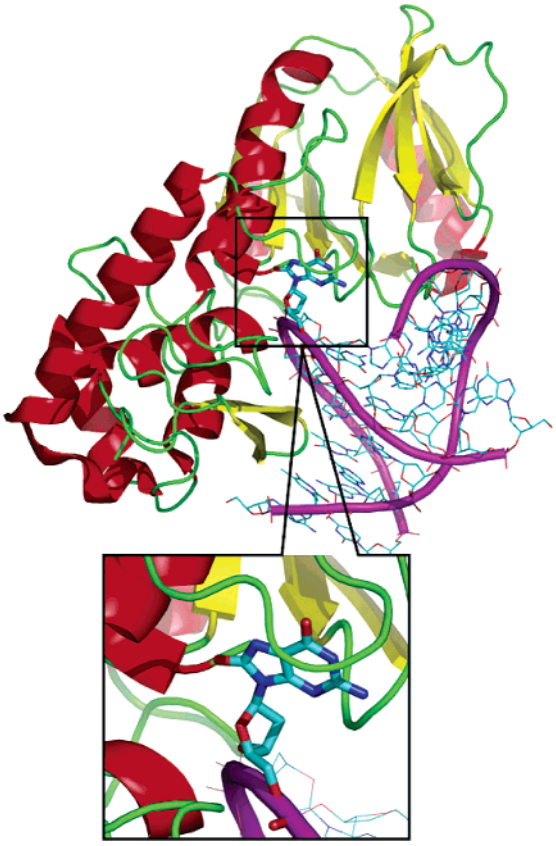

FIGURE 2: Structure of *B. st.* Fpg bound to 8OG DNA (PDB entry 1R2Y). Atomic detail is shown for the DNA duplex, and the protein is represented with a cartoon diagram.

base (*14*) (Figure 2). In this case, *Bacillus stearothermophilus* (*B. st.*) Fpg was used for the study with its catalytic activity eliminated by use of an E2Q mutant (*18*). This approach generated a structure with 8OG bound to Fpg in the *syn* conformation, in which a single hydrogen bond between the lesion N7 and Ser219's backbone carbonyl oxygen is formed. The mutation involves replacement of an ionizable (Glu) side chain with a neutral residue (Gln) that was observed to directly interact with the lesion; thus, the details of the resulting structure could differ from those of the active, wild-type enzyme.

Simulation studies (*19*, *20*) have been conducted in which the missing base in the structure reported by Gilboa et al. was modeled into the crystallographic structure (*15*). On the basis of their data, the authors proposed a binding mode for 8OG different from that deduced for the E2Q mutant by Fromme and Verdine. These simulations suggested that 8OG binds to Fpg in the *anti* conformation, making hydrogen bonding contacts to both N7 and O8 of 8OG. In contrast, the crystal structure from Fromme and Verdine show the damaged base to be in the *syn* conformation, with only N7 involved in hydrogen bonding with Fpg (*14*).

Unlike its human analogue, hOGG, Fpg displays no significant $\pi-\pi$ interactions with 8OG, resulting in a more spacious binding pocket. Thus, one possible explanation for the apparently contradictory results is that 8OG may bind to Fpg in either an *anti* or a *syn* conformation, although this does not fully explain the divergent observations. We previously used a computational approach to study the preference for *anti* and *syn* 8OG conformations in a variety of sequence contexts in duplex DNA (*21*). In the study presented here, we employ all-atom molecular dynamics simulations and free energy calculations to gain insight into the interactions involved in the complex between 8OG and wild-type Fpg. We address two specific questions: what is the preferred conformation of 8OG in the Fpg active site, and what are the interactions between 8OG and Fpg that provide for lesion recognition or a secondary discrimination against guanine prior to base excision? We make the remarkable observation that our simulations are consistent with both models of the Fpg−8OG complex. Additional calculations provide insight into the role of specific Fpg residues in the binding of DNA containing 8OG.

MATERIALS AND METHODS

System Preparation. All initial structures were built using the Leap module of Amber (version 8) (*22*), based on the crystal structure of the *B. st.* Fpg−DNA complex with 8OG in the *syn* conformation (PDB entry 1R2Y) (*14*). The sequence of the DNA duplex was d[GTAGACCTGGAC]·[GTCCAG*GTCTAC] (where G* is 8OG). All water molecules in the crystal structure were retained. The *anti* conformation was built by rotating the glycosidic angle of 8OG using the molecular modeling software MOIL-VIEW (*23*). Protein mutants were generated by manual editing of the PDB file, with the new side chain built using Leap. These structures were minimized for 100 cycles of steepest descent and then solvated in truncated octahedron boxes with a minimum of 6 Å buffer between the box edge and the nearest protein atom. The TIP3P model (*24*) was used to explicitly represent water molecules. Following previous studies (*19*, *20*), the N-terminal proline was modeled as neutral to mimic the stage directly before the reaction. The parameters for neutral N-terminal proline were obtained from ref *20*. Force field parameters for 8OG were obtained from ref *25*. Zinc was modeled using the Stote nonbonded model ($q = 2e^-$, $\sigma = 1.7$ Å, $\epsilon = 0.67$ kcal/mol) (*26*). The remaining protein and nucleic acid parameters were determined by employing Amber ff99 (*27*, *28*), with modified protein backbone parameters to reduce the α-helical bias of those force fields (*29*). Previous calculations showed a large pK_a shift for glutamate in the active site of T4 endonuclease V (*30*). However, Fromme and Verdine suggested (*14*) that Glu2 is unlikely to be protonated since it is at the N-terminus of an α-helix; it has been shown that acidic residues in this position tend to have low pK_a values (*31*). In this study, we calculated the pK_a value of Glu2 residues using two approaches, Jensen's empirical method (*32*) and Onufriev's H++ method (*33*); these methods resulted in Glu2 pK_a values of 5.69 and 6.78, respectively. To mimic the crystallization pH (7.5) (*14*), we simulated Glu2 in its deprotonated state unless specified otherwise. We also repeated selected simulations and umbrella sampling calculations using protonated Glu2 to examine the sensitivity of the results to the ionization state.

Molecular Dynamics Simulations. All molecular dynamics simulations were carried out with the SANDER module in Amber. Solvated systems were minimized and equilibrated in three steps: (i) 50 ps MD simulation (*22*) with protein and DNA atoms constrained and movement allowed only for water, (ii) five 1000-step cycles of minimization, in which the positional restraints on the protein and DNA were gradually decreased, and (iii) four cycles of 5000 steps of MD simulation with decreasing restraints on protein and DNA. A final 5000 steps of MD were performed without restraints. The resulting structures were used in the production runs.

SHAKE (*34*) was used to constrain bonds involving hydrogen atoms. The nonbonded cutoff was 8 Å. The particle mesh Ewald method (*35*, *36*) was used to calculate long-range electrostatics. Constant pressure (1 atm) and temperature (300 K) were maintained by the weak coupling algorithm (*37*).

Structural Analysis. The root-mean-square deviations (rmsd) of three regions of the complex were calculated separately: (i) protein, which is calculated using the Cα atom of all protein residues; (ii) DNA, which includes the backbone heavy atoms of only the 8OG·C and flanking base pairs; and (iii) loop, which includes the Cα atom of residues in the base binding loop (222−231), with the rmsd calculated for this region after a best fit of the coordinates of the entire protein. The glycosidic angle of 8OG is defined using atoms O4′, C1′, N9, and C4. The rmsd and glycosidic angle calculations, along with distance calculations, were carried out using the ptraj module.

Umbrella Sampling and Potential of Mean Force Calculations. Umbrella sampling (*38*−*41*) was used to calculate the potential of mean force (PMF) as a function of 8OG glycosidic angle in the binding site; 36 starting structures were generated using MOIL-VIEW (*23*) by rotating the 8OG glycosidic angle in 10° increments from 10° to 360°. These initial structures were energy-minimized, and one independent 200 ps simulation (i.e., one umbrella sampling window) was performed for each structure. The glycosidic angle was restrained to the initial value using a harmonic restraint with a force constant of 50 kcal mol^{-1} rad^{-2}. Residues more than 10 Å from 8OG were restrained using positional restraints (force constant of 2 kcal mol^{-1} Å^{-2}). The other parameters of these simulations were the same as those in the standard MD simulations. The resulting PMF was obtained by WHAM analysis (*39*−*41*) of the data using a program provided by A. Grossfield (freely available at dasher.wustl.edu/alan).

MM-GBSA. The MM-PBSA method has become widely used for calculation of free energies of binding (*42*). The MM-GBSA variant, in which the GB solvation model is used, was shown to successfully reproduce relative affinities and selectivities for a range of matrix metalloprotease inhibitors (*43*). In this study, we treated the protein and DNA (without 8OG) as the "receptor" and the 8OG nucleotide as the "ligand". The relative binding energy of 8OG in *anti* or *syn* conformations represents the contribution of protein−8OG interactions to the relative stability of these two binding modes. The absolute binding free energy is estimated with eq 1:

$$\Delta G_{\text{binding}} = G_{\text{COMP}} - G_{\text{receptor}} - G_{\text{ligand}}$$
$$G_{\text{x}} = E_{\text{VDW}} + E_{\text{EEL}} + G_{\text{pol}} + G_{\text{nonpol}} \quad (1)$$

The van der Waals energy (E_{VDW}) and intramolecular electrostatic energy (E_{EEL}) were calculated using SANDER (*22*, *44*). The polar part of the solvation free energy (G_{pol}) was calculated using the GB-OBC solvent model (*44*, *45*). The nonpolar solvation free energy (G_{nonpol}) was calculated by using the solvent accessible surface area (SASA) and a surface tension of 5 cal mol^{-1} Å^{-2} (*46*). Energy calculations for the isolated ligand (8OG) and receptor (Fpg) used coordinate sets obtained from the trajectory of the complex.

The images in this article were made using VMD (*47*) and pymol (*48*).

RESULTS AND DISCUSSION

Effect of the E2Q Mutation. In the crystal structure of *B. st.* E2Q-inactivated Fpg with DNA, the damaged base binds to the active site in the *syn* conformation (*14*) and the βFα10 binding loop is ordered. However, because of the proximity of the Glu2 side chain to the 8OG base (∼3 Å), the electrostatic effects arising from the E2Q mutation may be significant. Perlow-Poehnelt et al. (*20*) suggested this mutation as one possible reason for the difference in the 8OG conformation in the E2Q *B. st.* crystal structure and the *E. coli* computational model. We used two sets of simulations to test this possibility and also to investigate the general influence of the mutation on the structure of the complex. One simulation used the same E2Q mutant *B. st.* Fpg−DNA complex that was studied crystallographically, and the other employed wild-type *B. st.* Fpg in the complex. Three independent 2 ns simulations were used for each sequence to evaluate precision.

Stability of the Wild-Type (wt) and E2Q Systems. To analyze the stability of the structures under the simulation conditions, the rmsds of protein Cα atoms, lesion site residues (8OG·C and flanking base pairs), and the βFα10 loop were computed for each of the six independent simulations. The flexibility of the βFα10 loop was proposed to play an important role in the recognition and excision of the damaged base (*12*−*16*, *19*, *20*, *49*). Both systems are quite stable in all six 2 ns simulations. For the DNA lesion site, the rmsd plateau values are 0.5−1.0 Å with either protein sequence, while the rmsds of the entire protein and the βFα10 loop region are both slightly larger at ∼1.0 Å (Figure S1). The loop shows greater fluctuations about this average than the protein or DNA. We thus conclude that the simulations provide stable dynamics and that the mutation does not have any dramatic effect on the overall structure and stability of the complex.

Influence of the E2Q Mutation on the Active Site Geometry. Detailed analysis of the trajectories brought to light several small but interesting differences in the behavior of the two sequences. In the crystal structure of E2Q Fpg, a hydrogen bond forms between Nε of Gln2 and O8 of 8OG. This interaction is reproduced in the E2Q simulations (shown in Figure 3), with an average distance between these atoms of 3.04 Å (Figure S2), in excellent agreement with the value of 3.08 Å observed in the crystal structure (*14*). However, this hydrogen bond cannot form in the wild-type sequence since the amide nitrogen is not present. In particular, unfavorable electrostatic interactions are present in the wild type between the Glu2 carboxylate and O8 of 8OG, resulting in a shift of the protein strand containing Glu2 away from 8OG in the wt simulations (Figure 3).

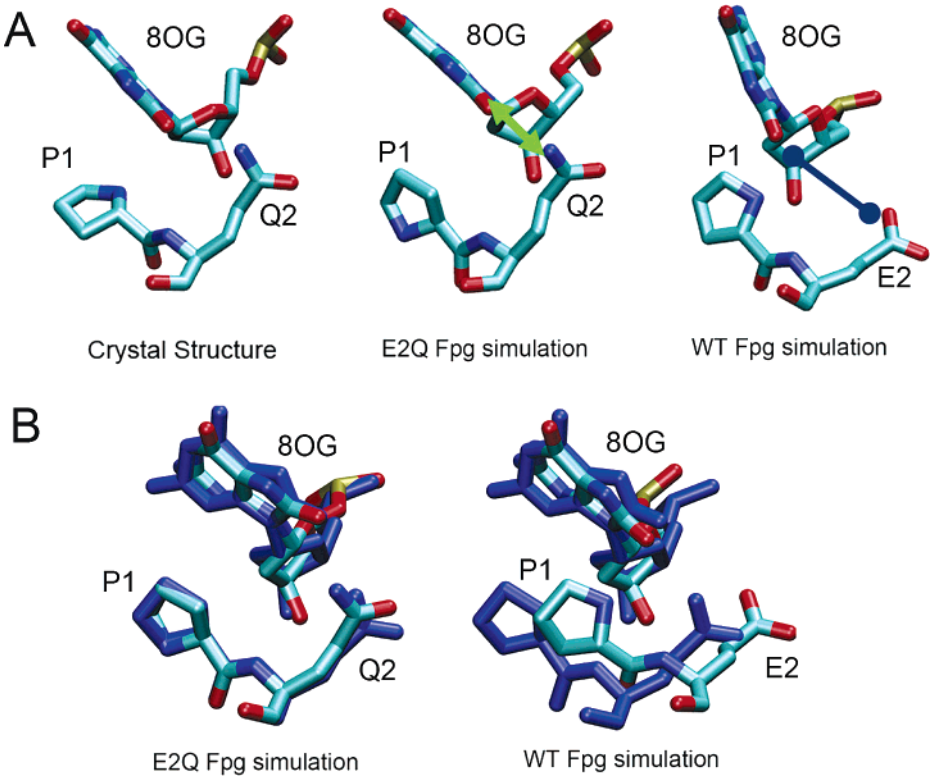

FIGURE 3: Comparison of conformation of the region containing 8OG, Pro1 and Gln2 (Glu2 in wt) between the crystal structure, the E2Q Fpg simulation, and the wild-type Fpg simulation. The favorable interaction between Q2 and O8 of 8OG is indicated by a green line, and the unfavorable interaction between WT E2 and O8 of 8OG is indicated with a blue line. Panel B shows the region after the best fit of 8OG, in which the crystal structure is colored dark blue and simulation structures are colored by atom. The E2Q simulations reproduce the E2Q crystal structure, but a shift in P1/E2 relative to 8OG is apparent in the wild-type simulation.

This difference between the E2Q mutant and wt Fpg causes two obvious effects. First, due to repulsion between 8OG and Glu2, the purine ring rotates slightly, increasing the distance between O8 and the Glu2 carboxylate. The glycosidic angle of 8OG in E2Q simulations is 108 ± 21° (uncertainty denotes the standard deviation), similar to the value of 101° in the crystal structure (*14*), suggesting that the simulation is reasonable. In the wild-type simulations, however, this angle is significantly different (57 ± 14°).

The second observed effect is that the average distance between N of Pro1 and C1′ of 8OG is shorter in the wild type than in the E2Q mutant. In the proposed mechanism of 8OG excision, C1′ of 8OG undergoes nucleophilic attack by the N of Pro1 (*9*, *50*). In simulations of wild-type Fpg, the average distance is ~4.2 Å. The E2Q mutant samples two minima at 4.5 and 5.8 Å, with a preference for the longer distance resulting in an average of ~5.3 Å (Figure S1). In the crystal structure, this value is 3.4 Å, much shorter than what we observe in the simulation of the same sequence. This apparent inconsistency arises from an ~180° rotation of the Pro1 ring in the simulation such that the nitrogen moves farther from 8OG. Interestingly, the rotated ring from the E2Q Fpg simulation still overlaps well with the crystal structure (but with exchange of the positions of N and C), and thus, the structure would be expected to remain in good agreement with the electron density (Figure 3B). Therefore, the different orientation of the Pro1 ring in our E2Q Fpg simulation and X-ray structure could result from the difficulty of distinguishing between carbon and nitrogen from the electron densities.

Thus, while the control simulations of the E2Q mutant reproduced the crystallographic data with high accuracy, analogous simulations on the wild-type sequence resulted in a slight rearrangement of the residues in contact with the lesion. These include a loss of a hydrogen bond present with Glu2 and a shift of Glu2 away from 8OG, resulting in shortening of the average distance between the Pro1 nucleophile and the 8OG sugar ring. These observations are consistent with the proposed catalytic mechanism.

Previous calculations have suggested that the pK_a of glutamate in the active site of T4 endonuclease V can be shifted to a higher value (*30*). However, Fromme and Verdine have suggested that Glu2 in Fpg is unlikely to be protonated at neutral pH (*14*), due to the observation that glutamates at N-termini of α-helices (such as Glu2 in Fpg) are less likely to be protonated (*31*). We calculated the pK_a value for Glu2 and obtained values of 5.7 and 6.8 depending on the method used for the calculation (see Materials and Methods). Since this value is reasonably close to neutral pH, we repeated the simulation using Glu2 with a protonated side chain carboxyl group. The resulting structure was similar to that observed in both the E2Q X-ray structure and the E2Q mutant simulation (Figure S1), with a hydrogen bond between O8 of 8OG and the Glu2 acidic hydrogen, which is analogous to the hydrogen bond observed between O8 of 8OG and Nϵ2 of Gln2 in the E2Q mutant. Due to the uncertainty of the ionization state of Glu2 and the lack of experimental structure data for the wild-type sequence with 8OG bound in the active site, we conclude that it is not possible from these simulations to determine which of these active site hydrogen bonding patterns is adopted under physiological conditions.

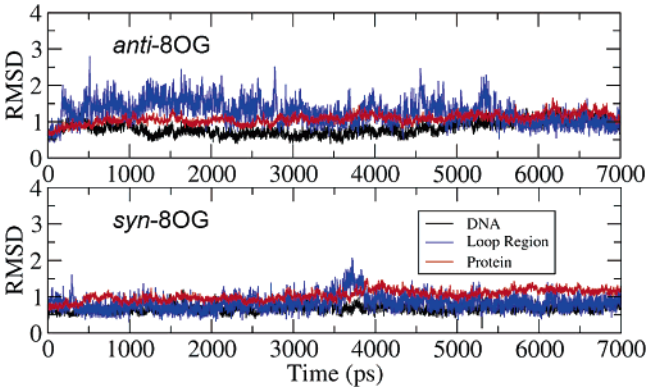

FIGURE 4: Root-mean-square deviations in Å for the protein, DNA, and binding loop during simulations of the wt Fpg−DNA complex with *anti* (top) and *syn* (bottom) conformations for 8OG.

Structural and Energetic Analysis of the Anti and Syn 8OG Binding Modes. The glycosidic angle of 8OG in the simulations described above was observed to depend somewhat on the sequence; the glycosidic angle changed from high *syn* (108°) in the E2Q mutant simulation (101° in the X-ray crystal structure) to *syn* (57°) in the wild-type sequence. However, a full transition to *anti* was not observed in any of the six simulations. Two scenarios could account for this observation. Either the *syn* 8OG conformation is thermodynamically favored, or 8OG also can bind in an *anti* conformation; however, the energy barrier between these two conformations is sufficiently large that transitions occur more slowly than the nanosecond time scale of our simulations. Thus, we cannot infer from solely these data whether the 8OG's conformation is more favorable in the *syn* or *anti* conformation.

Stability of the Anti 8OG and Syn 8OG Binding Modes. Simulations were performed for 7 ns for each of the two systems (*anti* and *syn* for wt Fpg). As shown in Figure 4, all four conformations are stable during the fully unrestrained simulations, with plateau rmsd values of ∼1 Å for the proteins and DNA fragments. The rmsds of the loop region (residues 222−231) fluctuate between 1 and 2 Å, with greater flexibility apparent in the *anti* 8OG systems.

Specific Interactions between 8OG and Fpg in the Two Binding Modes. Figure 5 shows the *anti* and *syn* binding modes of 8OG in the final structures from the simulations. Comparison of the structures reveals surprising similarity in the hydrogen bonding patterns of the two binding modes. Both conformations of 8OG fit in the binding pocket without significantly changing the conformation of the βFα10 loop. In the *syn* conformation (Figure 5a), 8OG forms four hydrogen bonds to Fpg: between N1 of 8OG and Oγ of T223, between N7 of 8OG and the backbone O in S219, between N2 of 8OG and Oϵ of E77, and a network of hydrogen bonds between O6 of 8OG and the N atoms of residues 221−224.

In the *anti* conformation (Figure 5b), because of the rotation of the base about the glycosidic bond, N1 and N7 swap their respective partners, with N7 now forming a hydrogen bond with T223's Oϵ and N1 hydrogen bonding to the O of S219. N2 acts as hydrogen bond donor to a different Glu residue, E5. O6 maintains its hydrogen bonds with the four residues (221−224).

To examine the stability of these interactions during our simulations, we used histogram analysis to compute the distributions of the distances corresponding to these contacts. The results, shown in Figure 6, confirm that these hydrogen bond interactions are stable throughout the simulations. In both conformations, O6 of 8OG forms a hydrogen bond network with the backbone N atoms of V221, R222, T223, and Y224. For simplicity, the average distance between O6 and these four N atoms is denoted "VRTY". The distributions of the average distances between O6 and the four VRTY residues are similar in the two binding modes.

Figure 6 also shows that the distribution of the distance between S219 and N7 in *syn* 8OG is similar to that for S219 and N1 in the *anti* 8OG binding mode. This also is the case for the distances involving T223, reflecting the exchange of these hydrogen bonds resulting from rotation about the glycosidic bond (Figure 5). N2 of 8OG forms a slightly shorter hydrogen bond with E77 in *syn* 8OG than it does with E5 in *anti* 8OG. The longer 8OG−E5 hydrogen bond distance may arise from the interaction of E5 with S219 in addition to 8OG, while E77 does not form hydrogen bonds with residues other than 8OG and therefore has more freedom to optimize its interaction with 8OG.

8OG differs from guanine in hydrogen bonding ability by having an acceptor O8 and conversion of N7 from acceptor to donor. Specific interactions with either group could be employed by Fpg as a final step in allowing the enzyme to discriminate between 8OG and deoxyguanine. In the crystal structure reported by Fromme and Verdine (*14*), N7 of 8OG in the *syn* conformation forms a specific hydrogen bonding interaction with S219. In contrast, previous simulations of *E. coli* Fpg (*19*, *20*) suggested that O8 of *anti* 8OG interacts with K217 (R222 in *B. st.* Fpg).

Interestingly, the binding pocket in our simulations appears to be able to accommodate both *anti* and *syn* 8OG, and a similar set of specific hydrogen bonds is formed in each. Both conformations have hydrogen bonds that could be used to differentiate 8OG from guanine. In our simulations, the hydrogen bond interaction formed with N7 of *syn* 8OG involves S219, consistent with the crystal structure reported by Fromme and Verdine (*14*). In the *anti* conformation, T223 is the key residue in recognizing the oxidized base, and not R222. Experimental data indicate that 8OG binding affinity is reduced approximately 4-fold in the K217T *E. coli* Fpg mutant (*19*). We calculated the distance between R222 and O8 during the simulations and found no significant interaction. However, since R222 is next to T223, mutation of the basic residue in this position may affect the conformation of T223 and thus the binding affinity of 8OG.

We also examined the effect of the *anti* versus *syn* change on the distance between the P1 nitrogen and C1′ of 8OG. As described above, the P1 N acts as a nucleophile attacking C1′ of the 8OG in the initial step of the glycosylase reaction (*9*, *50*). The distance between these two atoms is significantly different in the two binding modes. In *syn* 8OG, this distance is 4.2 ± 0.3 Å. In *anti* 8OG, this distance is 3.5 ± 0.3 Å. The shorter distance in *anti* 8OG could suggest that this conformation may be more reactive; however, other factors are likely to be involved.

Does Fpg Preferentially Bind Syn or Anti 8OG, and What Is the Role of Sequence Conservation? The MD simulations described above demonstrate that 8OG can adopt both *anti* or *syn* conformations in the active site. However, no transitions between these two conformations are observed in these

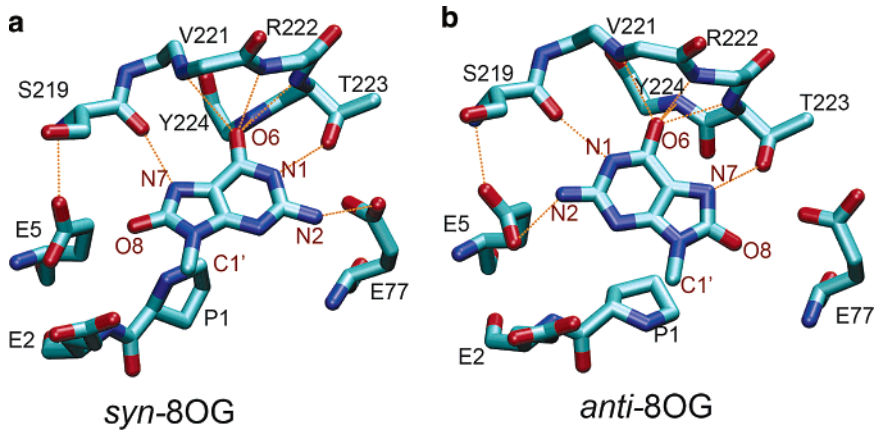

FIGURE 5: 8OG and surrounding residues in the Fpg−DNA complex with (a) *syn* 8OG as observed in the crystal structure and (b) *anti* 8OG built by rotation around the glycosidic bond of 8OG. Protein residues have black labels, and atoms of 8OG have maroon labels. Hydrogen bonds are represented by orange dashed lines. Only the base group of 8OG is shown (the remaining atoms linked to C1′ are not shown).

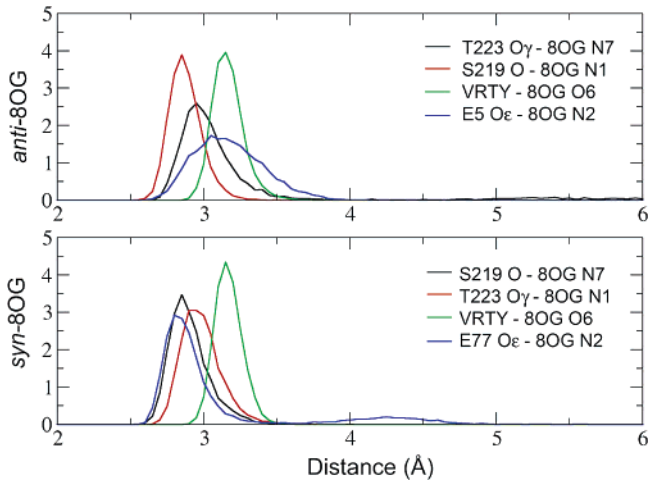

FIGURE 6: Histograms of distances corresponding to hydrogen bonds between Fpg and 8OG. VRTY represents the average distance between O6 of 8OG and the backbone N atoms of residues 221−224.

simulations. Similarly, Perlow-Poehnelt et al. reported no *anti*−*syn* transitions in their simulations of *E. coli* Fpg (*20*). Therefore, we cannot directly obtain the relative stabilities of these conformations by comparing their populations.

To directly compare the stability of the alternate binding modes in the wild-type sequence, we constructed a complex containing 8OG in the *anti* conformation by rotation of the glycosidic angle. Simulations of the wild-type sequence, for both *syn* and *anti* conformations, utilized two independent methods, umbrella sampling and MM-GBSA, to estimate the relative free energy of these binding modes. These methods differ significantly in their approach and thus provide a measure of the reliability of the conclusions. In addition, umbrella sampling has the advantage of being able to estimate the barrier for rotation about the 8OG glycosidic bond in the complex, while MM-GBSA can readily provide estimates for the contribution of different interactions to the relative free energies of the two conformations. Thus, the methods are independent but highly complementary. We first examine the relative stability of the conformations (since the *anti* form has not been observed crystallographically) and then examine the relative free energies.

We also used these MD simulations to investigate the role of the E77 side chain in 8OG binding. Figure 5 demonstrates that E5 and E77 appear to have opposite effects for *syn* and *anti* 8OG conformations. E77 stabilizes *syn* 8OG by forming a hydrogen bond with N2 of 8OG. On the other side of the binding pocket, the E5 side chain has Coulombic repulsion with O8 of 8OG. In *anti* 8OG, these roles are reversed; E5 forms a hydrogen bond with N2 of 8OG, and the side chain carbonyl of E77 repels O8 of 8OG.

Importantly, E5 is strictly conserved while E77 is not. In 85 sequences of Fpg found in the Swiss-Prot database, 48 have serine at the position corresponding to E77, 11 have threonine, and 26 have glutamate. This residue is serine in *E. coli* Fpg. Due to the direct interaction of E77 with 8OG and its different role with *anti* and *syn* 8OG conformations, sequence changes at this position could significantly affect the binding mode. To investigate this possibility, two additional systems were prepared for the E77S Fpg mutant with *syn* and *anti* 8OG.

Using the umbrella sampling procedure described in Materials and Methods, the potential of mean force for rotation of the 8OG glycosidic bond in the Fpg binding pocket was calculated for each of the two sequences (wt and E77S). The results are shown in Figure 7, and detailed values including statistical uncertainties are listed in Table S1. There are two obvious energy minima in the free energy profiles for both systems. One, located at approximately 55°, represents the *syn* conformation of 8OG. The other, at approximately −67°, corresponds to a high-*anti* 8OG.

Although the two energy minima have the same location in wild-type and E77S Fpg, the relative free energy of these two minima is highly sensitive to the effect of this mutation. This is consistent with the interactions shown in Figure 5, in which E77 makes favorable hydrogen bonding interactions with *syn* 8OG but shows unfavorable Coulombic repulsion with *anti* 8OG. With E77, *syn* 8OG is 2.7 kcal/mol more stable than *anti* 8OG, consistent with observation of the *syn* conformation in the crystal structure of this sequence. In the E77S mutation, however, *anti* 8OG becomes 6.1 kcal/mol more stable than the *syn* form. Thus, the E77S mutation has an important impact on the relative stability of the two binding modes.

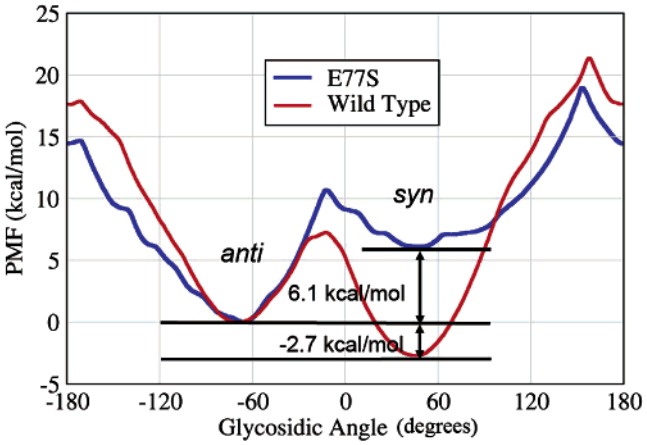

FIGURE 7: Potential of mean force (free energy profile) for rotation around the 8OG glycosidic bond in the Fpg binding pocket. Data are shown for the *B. st.* wild type (red) and the E77S mutant (blue). To aid the comparison, the free energy of the *anti* minimum was assigned a value of zero for both data sets.

Table 1: Relative *Anti* and *Syn* 8OG Binding Free Energies (kilocalories per mole), Calculated by MM-GBSA, in *B. st.* Wild-Type Fpg and the E77S Mutant[a]

	wild type	E77S
$\Delta\Delta E_{VDW}$	2.2 ± 0.1	2.0 ± 0.1
$\Delta\Delta E_{EEL}$	0.4 ± 0.8	-11.5 ± 0.5
$\Delta\Delta G_{pol}$	1.1 ± 0.1	1.0 ± 0.5
$\Delta\Delta G_{SASA}$	-0.1 ± 0.0	0.1 ± 0.0
$\Delta\Delta G_{total}$	3.7 ± 0.6	-8.3 ± 0.1

[a] Positive values indicate that the *anti* conformation is higher in free energy and therefore less favorable. Total free energies are provided as well as individual components of the free energy. Data are discussed in the text: VDW, van der Waals; EEL, electrostatic; POL, electrostatic component of the solvation free energy; SASA, nonpolar solvation energy using the solvent accessible surface area.

As described above, the ionization state of Glu2 affected the local conformations of the active site. Therefore, we investigated whether the Glu2 protonation would affect the free energy profile for rotation of 8OG by repeating the umbrella sampling calculation with protonated Glu2 in the wild-type sequence and comparing it with the profile obtained with deprotonated Glu2. We observe that the ionization state of Glu2 has little effect on the relative free energies of *anti* and *syn* or on the height of the transition state connecting these minima (at glycosidic angles near zero) (Figure S4). Larger differences are observed near the highest energy barrier (~150°) where the statistical uncertainties in the data are large (Table S1).

To further confirm these data from umbrella sampling calculations, MM-GBSA was also used to calculate the relative free energies of *anti* and *syn* binding by Fpg using deprotonated Glu2. The advantage of MM-GBSA is that it is possible to obtain approximate contributions to the total free energy from various types of interactions, including van der Waals, electrostatics, and solvation free energy. Results of this analysis are provided in Table 1.

The results of the MM-GBSA calculations are in good agreement with those obtained by umbrella sampling. For wild-type *B. st.* Fpg, *syn* 8OG is 3.7 kcal/mol more stable than *anti* 8OG (2.7 kcal/mol from umbrella sampling). For the E77S mutant, *anti* 8OG is 8.3 kcal/mol more stable than the *syn* form (6.1 kcal/mol from the umbrella sampling calculations).

In addition to providing additional support to the umbrella sampling results, MM-GBSA also supplies estimates of the contribution from each energy term. In wild-type *B. st.* Fpg, the total *anti*−*syn* energy difference arises mainly from van der Waals interaction (2.2 out 3.7 kcal/mol). The solvation free energy also favors *syn* 8OG by ~1 kcal/mol. The electrostatic energy does not favor either, which is reasonable since the number of hydrogen bonds in each conformation is the same (Figures 5 and 6).

The MM-GBSA energy decomposition for the E77S mutant is also shown in Table 1. The differences from van der Waals, solvation, and solvent accessible area between the two binding modes are roughly the same as those observed with E77. However, the electrostatic energy difference has changed dramatically, with the *anti* conformation stabilized by electrostatic interactions, ~11 kcal/mol more than the *syn* form. This result is consistent with the discussion above concerning changes in the role of E77 between *anti* and *syn*. The electrostatic energy in E77S favors *anti* 8OG so strongly that it becomes the preferred conformation, despite slightly less favorable van der Waals and solvation free energy.

SUMMARY AND CONCLUSIONS

X-ray crystallography continues to be the main source of information for understanding mechanisms of DNA damage recognition at the molecular level. The need for high-quality, diffracting crystals and the requirement of catalytically inactive damaged DNA−protein complexes are frequently addressed by the introduction of single-amino acid mutations or, even, by the deletion of a short protein segment. The perturbations that such changes may cause to the structure of the complex can be very difficult to evaluate. In contrast, computational methods are ideally suited to evaluation of these putative mutation-derived perturbations via comparison of the structures of complexes with the mutated and wild-type proteins.

We have focused on the role of the inactivating E2Q mutation of Fpg on the specific interactions involved in binding and, perhaps, excision of 8OG. Simulations of a complex of the E2Q mutant and DNA provided results consistent with the crystal structure of the same sequence. Simulations of the complex with the wild-type E2 sequence produced very similar results, with small changes arising from the loss of the group on Q2 that hydrogen bonds to 8OG in both the E2Q crystal structure and simulations. The glycosidic angle of 8OG changed from 108° (E2Q) to 57° (wt), and the N-terminal proline moved closer to 8OG, reducing the distance between the P1 nitrogen and C1′ of 8OG. In both the wild type and E2Q, the *syn* conformation of 8OG and the other key interactions between 8OG and Fpg were maintained as seen in the crystal structure.

We built a model of the wt *B. st.* Fpg−DNA duplex containing *anti* 8OG by using the crystal structure of the complex with *syn* 8OG and rotating the 8OG glycosidic bond. Subsequent unrestrained molecular dynamics simulations resulted in the interesting observation that the binding site (including the βFα10 loop) can readily accommodate both *anti* and *syn* 8OG without significantly changing its conformation. Additionally, hydrogen bonds seen in the crystal structure and simulations of the *syn* 8OG complex

were replaced by a nearly equivalent set for *anti* 8OG. Free energy calculations showed that *syn* 8OG is 2.7 kcal/mol more stable in this binding site than *anti* 8OG.

As suggested by early studies (*30*), the pK_a value of Glu2 can have a large shift toward its protonated state. Therefore, we also simulated the complex with protonated Glu2. The results indicate that the details of the hydrogen bonding between 8OG and Glu2 depend on the ionization state; however, its effect on the relative free energies of 8OG's two binding modes is not significant (Figure S4).

Previous studies also suggested that 8OG could possibly be accommodated by Fpg in either the *anti* or the *syn* conformation but that *anti* is favored (*19*, *20*), on the basis of simulation of *anti* and *syn* 8OG that were modeled into the *E. coli* Fpg crystal structure with a homology-modeled substrate binding loop. In those simulations, the *anti* conformation was shown to make stable contacts that were consistent with existing mutation data, while the *syn* conformation did not. The authors suggested that the E2Q mutation used for crystallization was one possible reason for the disagreement between their results and the crystal structure. In this study, we find that the *syn* 8OG conformation that is preferred with wt *B. st.* Fpg arises from favorable interaction with the nonconserved E77 side chain, which also destabilizes the *anti* conformation. This position in *E. coli* Fpg and many other Fpg sequences is occupied by serine, which has a shorter, neutral side chain. We hypothesized that the residue at this position could alter the preferred conformation of bound 8OG and tested this postulate through free energy calculations using the *B. st.* E77S mutant. These revealed that *anti* 8OG is indeed more stable than *syn* 8OG when E77 is replaced with serine.

While our simulations resolved an apparent discrepancy between two proposed models of binding of 8OG by Fpg, the simulations do not reveal any evolutionary advantage for using different binding conformations by *E. coli* and the thermophilic *B. st.* Moreover, the effect that the *syn* or *anti* 8OG binding site conformation, including hydrogen bonds specific to 8OG, has on the mechanism of lesion extrusion and/or damage recognition remains to be established. Future computations will address these points. It is interesting to note, however, that our simulations demonstrated that the *anti*−*syn* conformational change that makes a large difference in the properties of duplex DNA appears to make relatively little difference in the ability of Fpg to specifically bind the lesion.

SUPPORTING INFORMATION AVAILABLE

Figures S1−S4 and Table S1. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Frieberg, E. C., Walker, G. C., Siede, W., Wood, R. D., Shultz, R., and Ellenberger, T. (2006) *DNA Repair and Mutagenesis*, ASM Press, Washington, DC.
2. Halliwell, B., and Gutteridge, J. M. C. (1999) *Free radicals in biology and medicine*, 3rd ed., Oxford University Press, Oxford, U.K.
3. Kasai, H., and Nishimura, S. (1984) Hydroxylation of Deoxyguanosine at the C-8 Position by Ascorbic Acid and Other Reducing Agents, *Nucleic Acids Res. 12*, 2137−2145.
4. Bjelland, S., and Seeberg, E. (2003) Mutagenicity, toxicity and repair of DNA base damage induced by oxidation, *Mutat. Res. 531*, 37−80.
5. Grollman, A. P., and Moriya, M. (1993) Mutagenesis by 8-Oxoguanine: An Enemy Within, *Trends Genet. 9*, 246−249.
6. Tchou, J., and Grollman, A. P. (1993) Repair of DNA Containing the Oxidatively-Damaged Base, 8-Oxoguanine, *Mutat. Res. 299*, 277−287.
7. 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.
8. McCullough, A. K., Dodson, M. L., and Lloyd, R. S. (1999) Initiation of base excision repair: Glycosylase mechanisms and structures, *Annu. Rev. Biochem. 68*, 255−285.
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, *J. Biol. Chem. 272*, 5335−5341.
10. Zharkov, D. O., and Grollman, A. P. (2005) The DNA trackwalkers: Principles of lesion search and recognition by DNA glycosylases, *Mutat. Res. 577*, 24−54.
11. Zharkov, D. O., Shoham, G., and Grollman, A. P. (2003) Structural characterization of the Fpg family of DNA glycosylases, *DNA Repair 2*, 839−862.
12. Sugahara, M., Mikawa, T., Kumasaka, T., Yamamoto, M., Kato, R., Fukuyama, K., Inoue, Y., and Kuramitsu, S. (2000) Crystal structure of a repair enzyme of oxidatively damaged DNA, MutM (Fpg), from an extreme thermophile, *Thermus thermophilus* HB8, *EMBO J. 19*, 3857−3869.
13. 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.
14. Fromme, J. C., and Verdine, G. L. (2003) DNA lesion recognition by the bacterial repair enzyme MutM, *J. Biol. Chem. 278*, 51543−51548.
15. 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, *J. Biol. Chem. 277*, 19811−19816.
16. Serre, L., de Jesus, K. P., Boiteux, S., Zelwer, C., and Castaing, B. (2002) Crystal structure of the *Lactococcus lactis* formamidopyrimidine-DNA glycosylase bound to an abasic site analogue-containing DNA, *EMBO J. 21*, 2854−2865.
17. Francis, A. W., Helquist, S. A., Kool, E. T., and David, S. S. (2003) Probing the requirements for recognition and catalysis in fpg and MutY with nonpolar adenine isosteres, *J. Am. Chem. Soc. 125*, 16235−16242.
18. Lavrukhin, O. V., and Lloyd, R. S. (2000) Involvement of phylogenetically conserved acidic amino acid residues in catalysis by an oxidative DNA damage enzyme formamidopyrimidine glycosylase, *Biochemistry 39*, 15266−15271.
19. Zaika, E. I., Perlow, R. A., Matz, E., Broyde, S., Gilboa, R., Grollman, A. P., and Zharkov, D. O. (2004) Substrate discrimination by formamidopyrimidine-DNA glycosylase: A mutational analysis, *J. Biol. Chem. 279*, 4849−4861.
20. 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 43*, 16092−16105.
21. Cheng, X. L., 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.
22. Case, D. A. e. a. (2004) *AMBER 8*, University of California, San Francisco.
23. Simmerling, C., Elber, R. and Zhang, J. (1995) MOIL-View: A Program for Visualization of Structure and Dynamics of Biomolecules and STO: A Program for Computing Stochastic Paths, in *Modelling of Biomolecular Structure and Mechanisms*, Kluwer, Dordrecht, The Netherlands.
24. 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.
25. 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, *J. Am. Chem. Soc. 125*, 6331−6336.
26. Stote, R. H., and Karplus, M. (1995) Zinc-Binding in Proteins and Solution: A Simple but Accurate Nonbonded Representation, *Proteins: Struct., Funct., Genet. 23*, 12−31.

27. 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. (1995) A Second Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules, *J. Am. Chem. Soc. 117*, 5179−5197.
28. Wang, J. M., Cieplak, P., and Kollman, P. A. (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem. 21*, 1049−1074.
29. 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: Struct., Funct., Genet.* (in press).
30. Fuxreiter, M., Warshel, A., and Osman, R. (1999) Role of active site residues in the glycosylase step of T4 endonuclease V. Computer simulation studies on ionization states, *Biochemistry 38*, 9577−9589.
31. Forsyth, W. R., Antosiewiez, J. M., and Robertson, A. D. (2002) Empirical relationships between protein structure and carboxyl pK_a values in proteins, *Proteins: Struct., Funct., Genet. 48*, 388−403.
32. Li, H., Robertson, A. D., and Jensen, J. H. (2005) Very fast empirical prediction and rationalization of protein pK_a values, *Proteins: Struct., Funct., Bioinf. 61*, 704−721.
33. Gordon, J. C., Myers, J. B., Folta, T., Shoja, V., Heath, L. S., and Onufriev, A. (2005) H++: A server for estimating pK_as and adding missing hydrogens to macromolecules, *Nucleic Acids Res. 33*, W368−W371.
34. 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.
35. 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.
36. 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, *J. Am. Chem. Soc. 117*, 4193−4194.
37. 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.
38. Kottalam, J., and Case, D. A. (1988) Dynamics of Ligand Escape from the Heme Pocket of Myoglobin, *J. Am. Chem. Soc. 110*, 7690−7697.
39. 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.
40. 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.
41. Roux, B. (1995) The Calculation of the Potential of Mean Force Using Computer-Simulations, *Comput. Phys. Commun. 91*, 275−282.
42. Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S. H., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D. A., and Cheatham, T. E. (2000) Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models, *Acc. Chem. Res. 33*, 889−897.
43. Massova, I., and Kollman, P. A. (2000) Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding, *Perspect. Drug Discovery Des. 18*, 113−135.
44. Feig, M., Onufriev, A., Lee, M. S., Im, W., Case, D. A., and Brooks, C. L. (2004) Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures, *J. Comput. Chem. 25*, 265−284.
45. Onufriev, A., Bashford, D., and Case, D. A. (2004) Exploring protein native states and large-scale conformational changes with a modified generalized born model, *Proteins: Struct., Funct., Bioinf. 55*, 383−394.
46. Sitkoff, D., Sharp, K. A., and Honig, B. (1994) Accurate Calculation of Hydration Free-Energies Using Macroscopic Solvent Models, *J. Phys. Chem. 98*, 1978−1988.
47. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics, *J. Mol. Graphics 14*, 33.
48. DeLano, W. L. (2002) *Pymol*, DeLano Scientific, San Carlos, CA.
49. Amara, P., Serre, L., Castaing, B., and Thomas, A. (2004) Insights into the DNA repair process by the formamidopyrimidine-DNA glycosylase investigated by molecular dynamics, *Protein Sci. 13*, 2009−2021.
50. 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, *J. Biol. Chem. 270*, 11671−11677.

BI060380M