

DNA Deformation-Coupled Recognition of 8-Oxoguanine: Conformational Kinetic Gating in Human DNA Glycosylase

Haoquan Li,[†] Anton V. Endutkin,^{||,⊥} Christina Bergonzo,[†] Lin Fu,[#] Arthur Grollman,[‡] Dmitry O. Zharkov,^{*,||,⊥} and Carlos Simmerling^{*,†,§,Ⓜ}

[†]Department of Chemistry, [‡]Department of Pharmacological Sciences, and [§]Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, New York 11794, United States

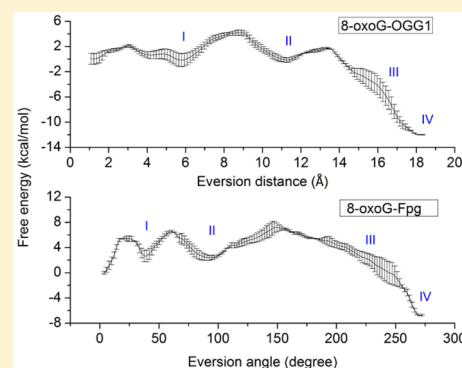
^{||}Novosibirsk State University, 2 Pirogova Street, Novosibirsk 630090, Russia

[⊥]SB RAS Institute of Chemical Biology and Fundamental Medicine, 8 Lavrentieva Avenue, Novosibirsk 630090, Russia

[#]School of Pharmaceutical Sciences, Wenzhou Medical University, Wenzhou 325035, P. R. China

Supporting Information

ABSTRACT: 8-Oxoguanine (8-oxoG), a mutagenic DNA lesion generated under oxidative stress, differs from its precursor guanine by only two substitutions (O⁸ and H⁷). Human 8-oxoguanine glycosylase 1 (OGG1) can locate and remove 8-oxoG through extrusion and excision. To date, it remains unclear how OGG1 efficiently distinguishes 8-oxoG from a large excess of undamaged DNA bases. We recently showed that formamidopyrimidine–DNA glycosylase (Fpg), a bacterial functional analog of OGG1, can selectively facilitate eversion of oxoG by stabilizing several intermediate states, and it is intriguing whether OGG1 also employs a similar mechanism in lesion recognition. Here, we use molecular dynamics simulations to explore the mechanism by which OGG1 discriminates between 8-oxoG and guanine along the base-eversion pathway. The MD results suggest an important role for kinking of the DNA by the glycosylase, which positions DNA phosphates in a way that assists lesion recognition during base eversion. The computational predictions were validated through experimental enzyme assays on phosphorothioate substrate analogs. Our simulations suggest that OGG1 distinguishes between 8-oxoG and G using their chemical dissimilarities not only at the active site but also at earlier stages during base eversion, and this mechanism is at least partially conserved in Fpg despite a lack of structural homology. The similarity also suggests that lesion recognition through multiple gating steps may be a common theme in DNA repair. Our results provide new insight into how enzymes can exploit kinetics and DNA conformational changes to probe the chemical modifications present in DNA lesions.



1. INTRODUCTION

The DNA integrity of living organisms is constantly under attack from reactive oxygen species (ROS). One major product of ROS is 8-oxoguanine (8-oxoG), which is generated by oxidation of guanine. 8-oxoG differs from guanine at only two positions: 8-oxoG has an O⁸ atom at C8 and an H⁷ atom at N7, whereas guanine has an H⁸ at C8 and a lone pair of electrons at N7 (Figure 1A). Like guanine, 8-oxoG can pair with cytosine in the Watson–Crick mode; however, 8-oxoG in the *syn* conformation can also form a stable Hoogsteen base pair with adenine (Figure 1B). Thus, during replication unrepaired 8-oxoG may lead to a G–C to T–A transversion mutation.¹ Increase in the frequency of 8-oxoG lesions has been linked to cancer and aging.^{2,3}

8-oxoG is very stable when paired with cytosine in duplex DNA, and the 8-oxoG-containing DNA is almost indistinguishable from the undamaged duplex.⁴ Nevertheless, 8-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 bases and create apurinic–apyrimidinic

(AP) sites. The human 8-oxoguanine–DNA glycosylase (OGG1) is one of the major glycosylases involved in the human BER pathway and is responsible for removal of 8-oxoG from 8-oxoG:C base pairs in duplex DNA. Formamidopyrimidine–DNA glycosylase (Fpg, also known as MutM) is a functional analog of OGG1 in prokaryotes, but these two enzymes have no sequence or structural homology. Both OGG1 and Fpg search for 8-oxoG by 1D diffusion (sliding) along DNA in a Brownian manner, and the DNA is kinked at the interrogated site.^{5–7} When an 8-oxoG:C base pair is located, the lesion is everted from the DNA helix and inserted deeply into the catalytic pocket, followed by hydrolysis of the *N*-glycosidic bond. Such base eversion (also known as base flipping) is a common feature of diverse glycosylase–DNA complexes. The target nucleotide must rotate out of the double helix so that the glycosylase can accommodate the damaged base at the extrahelical catalytic pocket and hydrolyze the

Received: November 3, 2016

Published: January 18, 2017

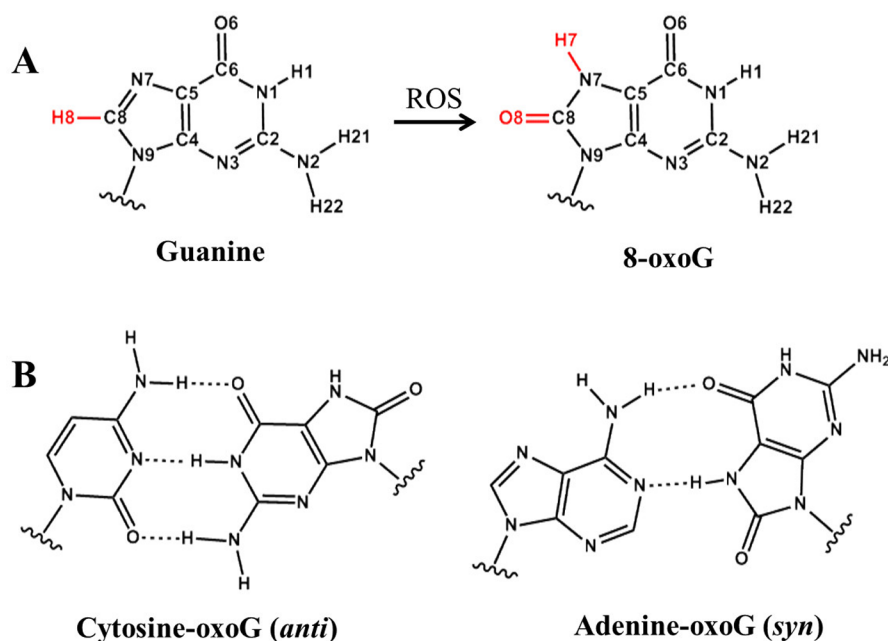


Figure 1. (A) Chemical structure of guanine and 8-oxoG. The structural difference between the two bases is highlighted in red. (B) Watson–Crick pairing of 8-oxoG:C and Hoogsteen pairing of 8-oxoG:A.

glycosidic bond. This process involves a series of conformational changes described as “pinch–push–plug–pull”.⁸ In the case of the OGG1–8-oxoG system, three pre-excision steps have been suggested using stopped-flow kinetics: (i) initial encounter and extrusion of 8-oxoG (“pinch–push”), (ii) insertion of certain OGG1 residues into DNA (“plug”), and (iii) insertion of 8-oxoG into the active site of OGG1, which isomerizes for catalysis (“pull”).⁹

To date, it remains unclear how OGG1 so specifically and rapidly distinguishes the relatively rare 8-oxoG among a tremendous excess of undamaged G. It has been proposed that the Lys249+/Cys253- dipole helps to recognize the N7–O⁸ dipole of 8-oxoG,¹⁰ but this dipole–dipole interaction is not essential because the double mutant Lys249Cys/Cys253Lys retains the ability to excise 8-oxoG.¹¹ Crystallographic studies have suggested that OGG1 can distinguish 8-oxoG from G at the active site by forming a discriminatory hydrogen bond between the carbonyl of Gly42 and the pyrrole N7 atom of 8-oxoG.^{10,12} However, there is no evidence that direct contact between OGG1 and the O⁸ of 8-oxoG plays a role in damage recognition, and the Gly42–H⁷ hydrogen bond is the only direct interaction observed to distinguish 8-oxoG from G. Perplexingly, biochemical studies have revealed several OGG1 substrates besides 8-oxoG that possess O⁸ but not H⁷, such as 2,6-diamino-4-oxo-5N-methylformamidopyrimidine (Me-FaPyG).^{13,14} These findings suggest that both H⁷ and O⁸ may be exploited for lesion recognition, and thus, there are probably some unidentified interactions between OGG1 and the H⁷ and O⁸ atoms other than the single G42–H⁷ hydrogen bond that forms in the active site. In addition, the extremely fast diffusion rate of OGG1 sliding along DNA almost certainly excludes the possibility that OGG1 fully extrudes each interrogated DNA base into its active site.⁵ Our studies have suggested that Fpg can recognize 8-oxoG through multiple gating intermediates during eversion. Therefore, we hypothesize that the substrate recognition by OGG1 may occur not only at the active site but also at one or more gating

intermediates in the base eversion pathway. Unfortunately, early intermediates for the wild type system are likely to be infrequently populated and thus inaccessible to direct experimental characterization, hindering our understanding of the damage recognition mechanism by OGG1.

Crystallography studies have provided important structural clues related to base eversion. The end of the eversion pathway is represented by a crystal structure of fully extrahelical 8-oxoG occupying the active site pocket of an OGG1 sequence mutated to prevent substrate turnover (PDB ID: 1EBM).¹² A putative early intermediate (PDB ID: 2ISW)¹⁵ along the base-eversion pathway showed an opened G:C pair, of which the G is slightly flipped out toward the major groove, stabilized by a hydrogen bond to an 8-oxoG adjacent to the interrogated base. In a putative later intermediate (PDB ID: 1YQK),¹⁰ the target G is extrahelical but rejected from the active site, instead residing at an adjacent “exo-site”. 8-oxoG has also been trapped in an exosite when the active site is sterically obstructed by a Gln319Phe mutation (PDB ID: 2NOF).¹⁶

It should be noted that these intermediate structures employed an engineered cross-link (between the cytosine of the interrogated pair and Cys149, mutated from WT Asn149) to force OGG1 to interrogate a noncognate G:C pair. Cys149 invades the DNA helix from the minor groove and forms a disulfide bond with the cytosine, displacing the target base from its position in the duplex. Furthermore, the details of the cross-link influences the extent of resulting eversion; the Cys149 cross-link shifted the target base to the exosite, while a distal cross-link at position 292 (PDB ID: 3IH7)¹⁷ resulted in complete eversion of undamaged G into the active site. Thus, the crystal structures provide valuable insight but lack information on the energetics of eversion in the biological system.

The major groove eversion mechanism implied by some of these crystal structures is consistent with reports for other glycosylases. For example, the crystal structure of uracil DNA glycosylase (UDG) suggested a major groove eversion

mechanism because UDG, like OGG1, binds DNA from the minor groove, and the minor groove path is sterically hindered.¹⁸ Also, computational studies on the Fpg-DNA complex strongly indicate that the eversion of 8-oxoG through the major groove is energetically more favorable than through the minor groove.¹⁹ Thus, it is plausible that OGG1 also everts 8-oxoG through a major groove pathway.

To understand how OGG1 recognizes 8-oxoG against G and evaluate potential similarities to Fpg, we energetically and structurally characterized the base eversion pathway in OGG1 through the use of molecular dynamics (MD) and enzyme kinetics with modified substrates that cannot form the apparently important transient interactions detected by MD. This combined approach has a unique advantage of connecting energy, structure, and dynamics with high spatial (i.e., atomic) and temporal resolution. Following the protocol of our previous study on the Fpg system,¹⁹ here we compare the free energy and structure of 8-oxoG and G eversion to understand the damage recognition mechanism of OGG1. We found that in addition to the active site discrimination against 8-oxoG and G, the system also provides several earlier stages of potential damage recognition, and the mechanism is similar to that observed for Fpg including a mechanistic role for the crystallographically observed bending of the interrogated DNA.

2. METHODS

2.1. Software, Force Fields, and Parameters. The SANDER module in the Amber 10 and Amber 11 suites of programs^{20,21} was used for molecular dynamics (MD) simulations. For all systems, the ff99SB force field²² was used with the parmbsc0²³ adjustment for DNA α,γ backbone parameters. The parameters for 8-oxoG were obtained from Miller et al.²⁴ The TIP3P²⁵ explicit solvent water model was used to solvate the protein–DNA complexes.

2.2. Calculations of the Base Eversion Paths Using PNEB. As with our previous work for Fpg,¹⁹ we used the PNEB variant²⁶ of the nudged elastic band method.²⁷ This approach finds a minimum energy pathway between two given states (end points). In this work, one end point is defined as OGG1 interrogating an intrahelical, unopened 8-oxoG:C or G:C base (the intrahelical state), whereas the other end point is that with the target base inserted into the active site pocket of OGG1 (the in-pocket state). For the direction of the sampling of the base-eversion path along the major groove, additional intermediate structures were used to seed the initial path, similar to what we did for Fpg.¹⁹ Because using one seeding intermediate failed to generate a major groove path in OGG1, here we used two seeding intermediates (see additional details in the Supporting Information). For both the 8-oxoG-complex and the G-complex, the initial base eversion path was generated by linking 32 replicas: 10 replicas of the equilibrated intrahelical structure model, 10 replicas of the equilibrated in-pocket structure, and 6 replicas of each of the two intermediate guiding seeds. Parameters for running PNEB were same as used in our previous study,¹⁹ except as noted in the Supporting Information.

The 1EBM,¹² 2ISW,¹⁵ and 1YQK¹⁰ crystal structures were considered reasonable candidates for input structures. They represent the everted end state and two putative intermediate states of base eversion, respectively: in the 1EBM structure, the 8-oxoG is everted into the active site of OGG1 (in-pocket state); in the 2ISW structure, the target G is slightly everted toward the major groove, stabilized by a hydrogen bond with the neighboring 8-oxoG; 1YQK is a putative later intermediate in which the target G samples an extrahelical exosite along the major groove path.

For intermediates to seed the major groove pathway for 8-oxoG, we used 2ISW and 1YQK. For G, we elected to use only 2ISW and not 1YQK to seed the path. Crystallographic studies have suggested that 1YQK is a relatively stable intermediate state for an extrahelical G; thus, we used 1YQK as independent validation that the simulated G eversion sampled a pathway consistent with experimental data. A

second guiding seed for G was generated by performing 50 ps MD simulation on the equilibrated 2ISW-based intermediate, forcing the target G to further evert to 9.5 Å of the eversion reaction coordinate (see below) using 10 kcal·mol⁻¹·Å⁻² restraint.

Currently, there is no available crystallographic structure representing a fully intrahelical end point due to the use of cross-links to create artificially specific interrogation structures. Therefore, a model was generated by computationally modifying an existing OGG1 structure. The 2ISW structure was considered the best candidate because the target base is only slightly everted; thus, we expected that it would be relatively straightforward to force the base to be fully intrahelical. The modification was performed after initial equilibration of the 8-oxoG-complex model of 2ISW. Detailed procedures of generating the input structures for NEB can be found in the Supporting Information.

To employ these crystal structures in our eversion model, we reverted all OGG1 sequences back to wild type (active enzyme, with no cross-links) and also made the DNA sequence consistent among all structures, adopted from 1EBM (5'-GGGGTAGACCTGGAC-3', 5'-GTCCA \underline{X} GTCTACCCC-3'; "X" is the interrogation position, using 8-oxoG in the 8-oxoG-complex or G in the G-complex). Note that although the 2ISW crystal structure had an 8-oxoG proximal to the interrogation site, all simulations described here with 8-oxo-G have it located only in the interrogation site.

2.3. Calculation of the Potentials of Mean Force for Base Eversion. Umbrella sampling (US) was used to obtain the potentials of mean force (PMF) as a function of the eversion distance (defined in Figure 2, with further details provided in the Supporting Information).

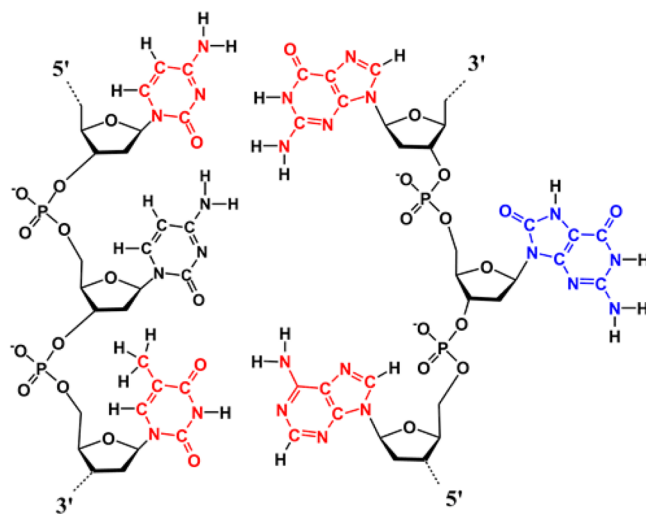


Figure 2. Eversion distance is defined as the distance between the center of mass of the heavy atoms of the base pairs flanking the target base pair (colored in red) and the center of mass of the heavy atoms of the target base (colored in blue). This distance was used as the reaction coordinate for base eversion.

To further enhance conformational sampling, Hamiltonian replica exchange molecular dynamics²⁸ were employed. The US protocol was otherwise similar to those in our previous studies using conventional umbrella sampling.^{29,30} For both the 8-oxoG complex and the G complex, two independent sets of initial structures were taken from the PNEB production trajectories and the US was performed for 2.5 ns per window in an NVT ensemble and repeated for each set of initial structures. US windows were evenly spaced along the eversion distance with 0.3 Å intervals (Figure S1), restrained by 10 kcal·mol⁻¹·Å⁻² umbrella potential. The time step was 2 fs, and exchanges between neighboring windows were attempted every 10 ps. The temperature was maintained at 310 K using a Langevin thermostat³¹ with a 75.0 ps⁻¹ collision frequency. The eversion distance values were recorded every time step and then analyzed using the weighted histogram analysis method (WHAM).³² The difference between the profiles obtained from two independent US runs were used to estimate the

lower bound of uncertainty (precision) in the data. The zero of energy is arbitrary in molecular mechanics, and only relative energies on PMF curves are meaningful. Therefore, absolute energies on the PMFs for different systems cannot be directly compared. Following our previous work on Fpg, the intrahelical end point of each PMF was set to 0 kcal/mol to allow the PMFs to highlight differences in the energetic profiles of 8-oxo-G and G as eversion proceeds.¹⁹

2.4. Enzyme Activity Assay. Oligonucleotides of the sequence 5'-CTCTCCCTpsTpsCXpsCpsTCCTTTCCTCT-3' containing 8-oxoG (X) and a phosphorothioate linkage at one of "ps" positions (or all regular linkages), as well as the complementary strand 5'-AGAGGAAAGGAGCGAAGGGAGAG-3', were synthesized from commercially available monomers (Glen Research, Sterling, VA) using an ASM-700 synthesizer (Biosset, Novosibirsk, Russia). The oxoG-containing strand was ³²P-labeled using T4 polynucleotide kinase (Biosan, Novosibirsk, Russia) and annealed to a 1.5-fold molar excess of the complementary strand. *Escherichia coli* Fpg and human OGG1 were purified from the respective overproducing *E. coli* strains as described.^{33,34} The reaction mixture (20 μ L) included 50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 1 mM EDTA, 1 mM DTT, 100 nM 8-oxoG-containing duplex, and 10 nM OGG1 or Fpg. The reaction was allowed to proceed for 20 min (OGG1) or 10 min (Fpg) at 37 °C. For OGG1, the reaction was terminated by adding NaOH to 50 mM and heating for 1 min at 95 °C, immediately neutralized with HCl, mixed with 0.5 volumes of the loading dye containing 80% formamide, and heated again for 1 min. For Fpg, the reaction was terminated by adding an equal volume of the loading dye and heating for 1 min at 95 °C. The products were separated by electrophoresis in 20% polyacrylamide gel containing 8 M urea and quantified using Molecular Imager FX (Bio-Rad, Hercules, CA).

3. RESULTS AND DISCUSSIONS

3.1. Simulation Strategy. We hypothesize that 8-oxoG recognition by OGG1 occurs in a series of steps, including those prior to complete insertion of target base into the active site. To test this, we structurally and energetically compared the eversion of 8-oxoG and G in OGG1, which are the processes of base eversion from inside DNA (intrahelical state) to the active site of OGG1 (in-pocket state). The base eversion process in OGG1 has been found to occur on the millisecond time scale,⁹ which is unlikely to be adequately sampled by currently affordable conventional MD simulations. We used pathway mapping and umbrella sampling simulations to investigate the free energy profile for 8-oxoG and G eversion in Fpg; the results indicated that 8-oxoG is stabilized by Fpg at several intermediate stages of eversion, whereas the G eversion is energetically unfavored.^{19,35} Here, we used the same methods to characterize the eversion paths for OGG1, permitting comparison of the recognition mechanisms of these two proteins that carry out comparable functions despite a lack of structural similarity.

3.2. Free Energy Profiles of Eversion of 8-OxoG and G in OGG1. Using the protocols described in the Methods section, we generated and compared the free energy profiles along the eversion pathways of 8-oxoG and G (Figure 3). Crystallographic studies have proposed a major groove path for OGG1,^{10,15,16,36} and OGG1's functional analog Fpg also prefers base eversion through the major groove.¹⁹ Thus, in this work we focus only on comparing eversion and recognition through the major groove. The intrahelical position is represented by the local minimum near 1 Å of eversion distance, while complete eversion into the OGG1 active site corresponds to the minimum near 18 Å. The 8-oxoG complex favors the in-pocket everted state by about 12 kcal/mol with respect to the intrahelical state, while for the G complex, the in-pocket state is energetically similar to the intrahelical state. In addition, the

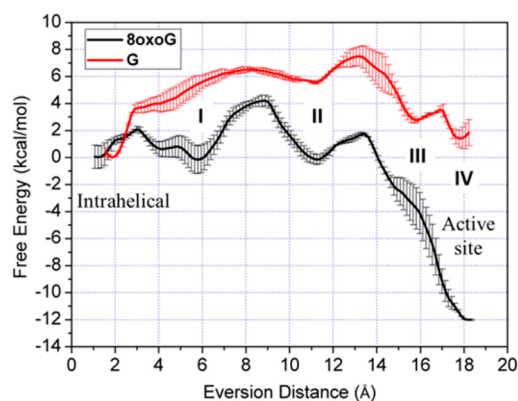


Figure 3. Comparison of computed free-energy profiles for base eversion in the 8-oxoG complex (black) and the G complex (red). A total of four potential 8-oxoG-recognizing steps are labeled as I, II, III, and IV. Error bars reflect the differences between two independent runs.

energetic barrier to eversion of 8-oxoG is only \sim 4 kcal/mol, while G encounters a substantially higher energetic barrier (\sim 7–8 kcal/mol). A higher kinetic barrier to extrusion of G versus 8-oxoG was also observed in Fpg,³⁵ which is consistent with the fact that 8-oxoG is the cognate substrate for OGG1 and Fpg whereas G is not.

The free energy profiles show significant differences between 8-oxoG and G at four regions along the base eversion paths: 3–6, 9–12, 14–17, and 17–18 Å of eversion distance (labeled in Figure 3). These locations suggest multiple intermediate states of energetic and, presumably, structural discrimination between 8-oxoG and G, through which we may gain insight into how damage recognition occurs. The details of these four states are discussed below.

Initial opening of the 8-oxoG:C base pair occurs at an eversion distance of 3 Å (where 8-oxoG loses all hydrogen bonds with the opposite C), requiring only 1–2 kcal/mol (Figure 3). These barriers are significantly lower than those we calculated for 8-oxo-G:C and G:C pairs in the context of duplex DNA,³⁷ probably due to the destabilizing effects of the Y203 wedge as observed in our intrahelical model (see the Supporting Information for details). This is also consistent with our previous work showing that buckling induced by the wedge intercalation in Fpg destabilizes the intrahelical state.³⁸ The energetic cost of breaking the G:C or 8-oxoG:C pairs are comparable, indicating no preferential opening of damaged DNA. Notably, the energy cost for the 8-oxoG:C opening in OGG1 is comparable to that in Fpg¹⁹ and also comparable to the energy barriers of OGG1/Fpg sliding along DNA (0.5 to 2 kcal/mol)⁵ but is significantly lower than G opening in the context of B-DNA, which has a \sim 10 kcal/mol barrier.³⁹ Thus, OGG1 and Fpg probably promote base eversion to facilitate base readout during fast sliding.

The free energy landscapes of the two systems start to diverge shortly after the base pair has been opened, indicating an opportunity for damage recognition early in eversion (Figure 3, step I). When 8-oxoG disengages from the paired cytosine and is slightly everted (3–6 Å), it is stabilized by a hydrogen bond between the H⁷ and the second 5' phosphate (the phosphate of the 5' nucleotide, hereafter referred to as p¹) (Figure 4). This damage-specific interaction corresponds to the free energy minimum at 6 Å eversion; whereas for the G-complex, the same phosphate exerts electrostatic repulsion with

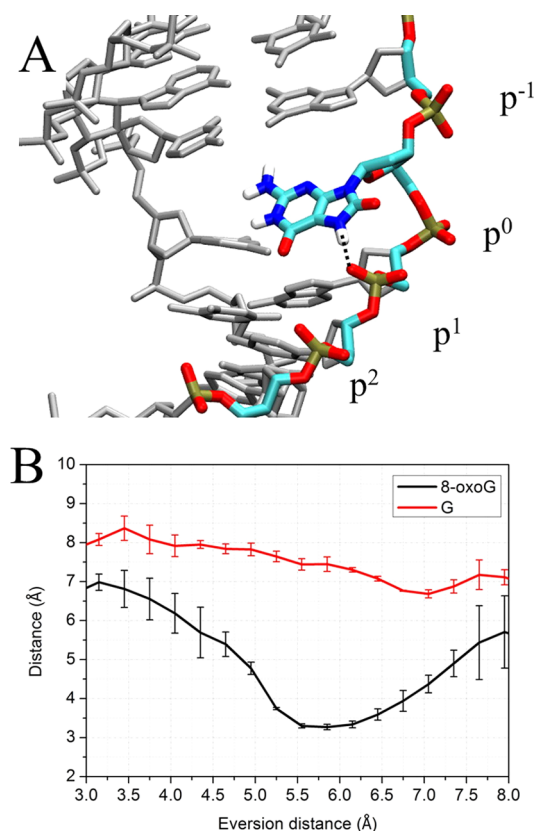


Figure 4. (A) DNA structure in eversion step I of Figure 3 with an eversion distance of 6.1 Å. The black dots indicate the hydrogen bond between the N7 of 8-oxoG and the OP2 of p^1 . Hydrogen atoms are present, but most are not shown for clarity. (B) Comparison of change during eversion of average distance between the OP2 of p^1 and the N7 of 8-oxoG (black) and G (red). The error bars reflect the difference between two independent runs.

the lone pair of electrons on N7, contributing to the energy rise at 4–6 Å eversion (Figure 3). The data suggest that the discriminatory interaction with 8-oxoG and G provided by p^1 constructs an early checkpoint for 8-oxoG damage, which could be more efficient than selection only in the active site. To pass through this unfavorable state, G has to overcome a ~6 kcal/mol energetic barrier (Figure 3), a significant impediment during the fast lesion search by OGG1, and thus G probably rapidly returns to the duplex after minimal eversion followed by continued sliding of OGG1.

It should be noted that the base- p^1 interactions here require a bent DNA; in a standard, unbent B-DNA, such interactions are unlikely due to the larger distance between the base and p^1 as compared to that in OGG1 (Figure 5). This suggests that DNA bending may have a functional role by recruiting the phosphodiester backbone to read the status of N7. This hypothesis is supported by recent studies indicating that a mismatch at the 5'-neighboring position of 8-oxoG strongly decreased the rate of 8-oxoG removal,^{40,41} and here, we suggest that those mismatches may interrupt the favorable 8-oxoG/ p^1 interaction by repositioning the p^1 phosphate and thus hinder 8-oxoG eversion.

When 8-oxoG is everted to a midway point (11 Å eversion, step II), it forms an additional hydrogen bond to p^1 (Figure 6A). G can also form the same hydrogen bonds to p^1 , but the hydrogen-bonding distances between 8-oxoG and p^1 are considerably shorter than those between G and p^1 (Figure

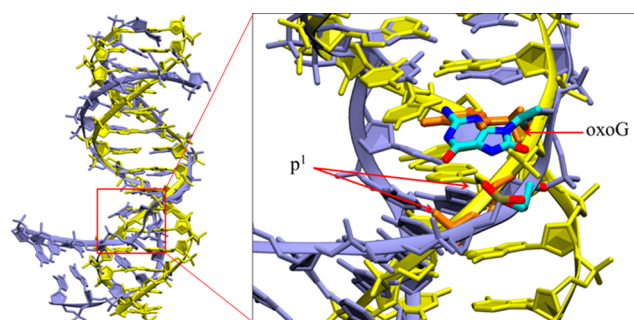


Figure 5. Superimposition of the bent DNA (purple) from Step I and standard B-DNA duplex (yellow, generated by 3DNA).⁴² Hydrogen atoms are present, but most are not shown for clarity. In the magnified region, the B-DNA is in yellow with 8-oxoG and p^1 in orange; the DNA from step I is purple, with 8-oxoG and p^1 colored by atom. The difference in geometries is consistent with the distance data shown in Figure 1B.

6B), suggesting that the p^1 may have stronger interactions with 8-oxoG than with G. This is probably because 8-oxoG at the midway point is also stabilized by the hydrogen bond between O⁸ of 8-oxoG and the backbone amide linking His270 and Val269 (Figure 6A,C). However, His270 does not stabilize G in this state because G lacks O⁸. This hampers ideal positioning of G for hydrogen bonding, and the resulting energy minimum is significantly shallower than that of 8-oxoG complex. This suggests an early potential checkpoint for 8-oxoG, again facilitated through recruitment of the phosphodiester backbone via DNA bending. At the energy barrier between steps I and II (near 9 Å eversion in Figure 3), the structures for 8-oxoG and G are similar, with both adopting a repulsive interaction between p^1 and O⁶ (Figure S2).

3.3. Phosphate p^1 is Critical for the Activity of OGG1.

Because our simulations predicted important transient interactions of the p^1 phosphate with 8-oxoG during the eversion, we analyzed the effect of the replacement of this phosphate with a phosphorothioate linkage. Phosphorothioates keep the negative charge of phosphates but have a bulkier, more-hydrophobic, easier-to-polarize, and less-electronegative sulfur atom replacing one of the nonbridging oxygens; as a result, charge is mostly concentrated on the remaining oxygen, while the sulfur is nearly neutral. All these differences are expected to alter the interactions of the internucleotide linkage with its surroundings, allowing us to probe experimentally the important interactions suggested by the computational results on the biologically relevant DNA systems. Analogous simulations of phosphorothioates would involve significant effort in the development of a modified DNA backbone force field and was, therefore, considered to be beyond the scope of this work.

We substituted p^2 , p^1 , p^{-1} , or p^{-2} phosphates with phosphorothioates (no p^0 substitution was made because 8-oxoG phosphorothioate is presently not available synthetically) and investigated the activity of OGG1 on these substrates. The products were treated with hot alkali to digest all abasic sites and thus follow the base excision reaction only rather than the combined base excision- β -elimination. A substitution at p^1 significantly reduced the activity of OGG1 in both the multiple-turnover (Figure 7) and the single-turnover modes (Figure S3). Notably, in the structure of OGG1 bound to 8-oxoG-containing DNA,¹² p^1 forms no direct contact with the protein, suggesting that this effect is not due to destabilization of the

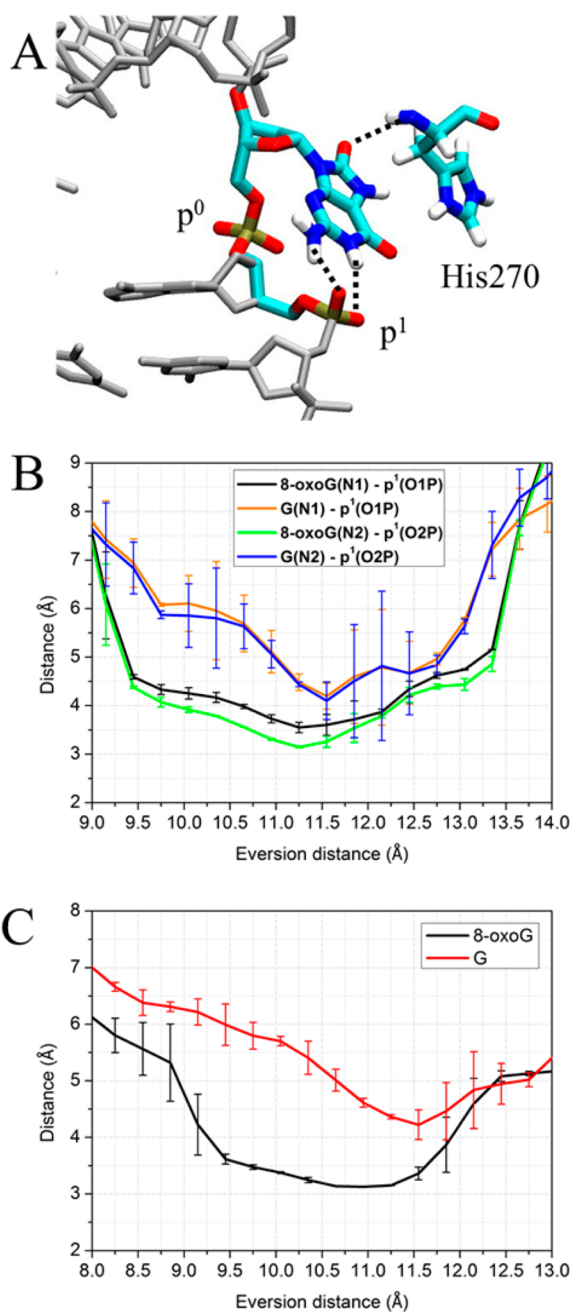


Figure 6. (A) Structure of the 8-oxoG-complex in step II with an eversion distance of 11.6 Å. The hydrogen bonds stabilizing 8-oxoG are depicted as dotted lines. (B) Comparison of distances between p¹ and the nitrogen atoms at the Watson–Crick face of the 8-oxoG (black and green) and G (orange and blue) as a function of eversion. (C) Comparison of distances between the backbone nitrogen of H270 and the O⁸ of 8-oxoG (black) or the H⁸ of G (red). The error bars reflect differences between two independent US runs.

Michaelis complex. The other three substitutions decreased the activity about 2-fold (Figure 7), likely because we used racemic mixtures of *R*- and *S*-phosphorothioates, only one of which conserved the bonds with the protein. Moreover, Fpg tested on this set of substrates showed the same preferential inhibition by the p¹-to-prosphorothioate substitution (Figure S3); in this case, too, p¹ does not form bonds with the protein when the lesion is fully everted.³³ Unlike in the well-known blockage of nuclease hydrolysis by phosphorothioates, p¹ does not

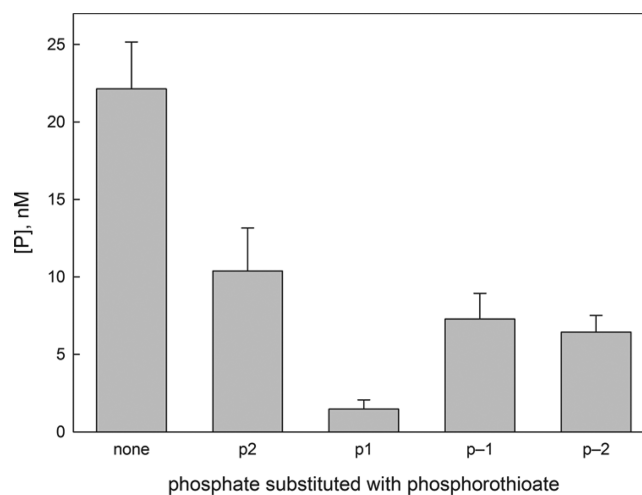


Figure 7. Activity of OGG1 on phosphorothioate-substituted 8-oxoG-containing substrates. See the Methods section for details.

participate in the reaction chemistry of glycosylases, so the inhibition of OGG1 and Fpg is most likely of conformational nature as suggested by the computational model.

3.4. Role of the Exosite. The 1YQK¹⁰ crystal structure represents a putative intermediate with G occupying an exosite, but the biological relevance of this intermediate is unclear due to the artificial nature of the cross-link. As discussed in the Methods section, the exosite structure was not used as input in the PNEB calculation for the G eversion pathway. Nevertheless, we observed that multiple independent G complex PNEB path simulations sampled this intermediate at an eversion distance of ~14.5–16.5 Å (Figure 8), supporting the relevance of the exo

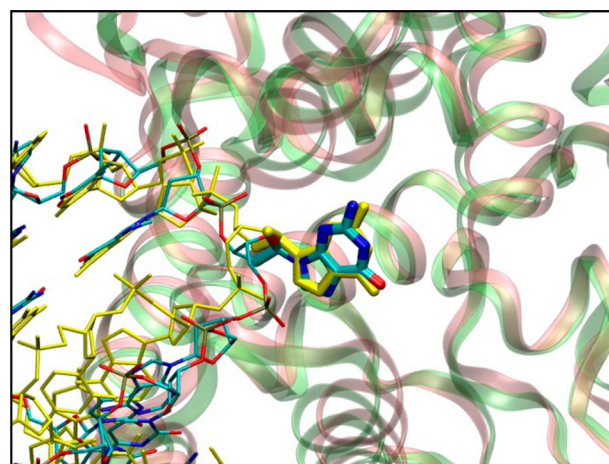


Figure 8. Superposition of the crystal structure 1YQK (DNA in yellow and protein in pink) and the structure sampled during the simulated base eversion (with DNA colored by atom and protein in green). Both structures were superimposed using protein heavy atoms. Hydrogen atoms are present, but most are not shown for clarity.

site (despite the crosslinking that was needed to capture it in the crystal). The simulation data also help rationalize the sensitivity of the conformation observed in the crystal structures to the position of the cross-link because the exosite and in-pocket states are seen to be nearly degenerate in free energy.

In our simulations, the backbone amide of Ile152 and the N7 atom of G form a weak hydrogen bond in the exosite (~3.7 Å

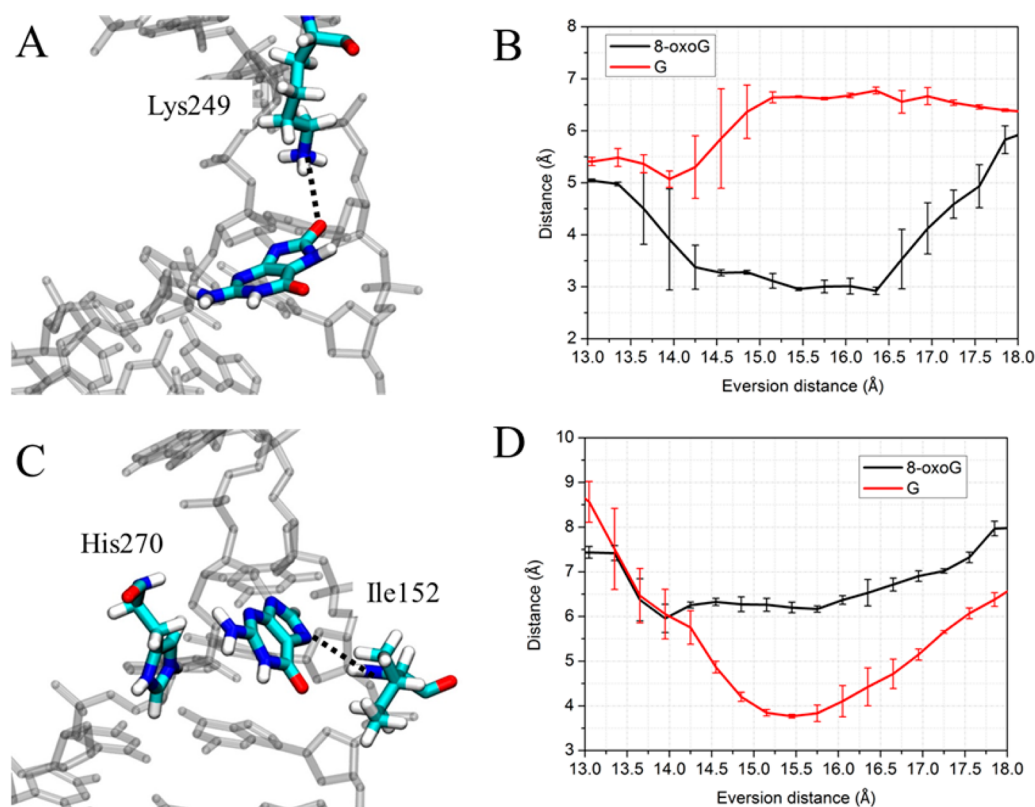


Figure 9. (A) Structure indicating the DNA interaction with Lys249 in step III, with an eversion distance of 14.8 Å. The 8-oxoG-stabilizing hydrogen bond is depicted as a dotted line. (B) Comparison of distances between the Lys249 ϵ -amino and the O⁸ of 8-oxoG (black) or the H⁸ of G (red). (C) Structure of the G complex in step III, with an eversion distance of 13.9 Å. G is stabilized by the stacking with His270 and a hydrogen bond with Ile152. (D) Comparison of distances between the backbone nitrogen of Ile152 and the N7 of 8-oxoG (black) or G (red). Error bars reflect the differences between two independent runs.

average distance, Figure 9D), very similar to the distance in the 1YQK crystal structure (~ 4.2 Å). Additional stabilization in the exosite comes from stacking of the π -face of G with the His270 aromatic ring as seen in simulations (Figure 9C), although the high B-factors¹⁰ for the His270 side chain in 1YQK suggest this may also be a weak interaction. Nevertheless, when G loses these weak interactions and enters the active site, it has to overcome a modest energetic barrier. 8-oxoG also samples the exosite, wherein O⁸ forms a hydrogen bond to the ϵ -amino group of Lys249 (Figure 9A,B); the resulting O⁸-Lys249 interaction thus stabilizes 8-oxoG in the exosite. Due to lack of O⁸ on G, the favorable O⁸-Lys249 interaction is not observed in the G complex (Figure 9B), and thus, this interaction helps to further distinguish 8-oxoG and G. Notably, Lys249Gln and Lys249Cys/Cys253Lys mutants retain the ability of extruding 8-oxoG to the active site,^{11,12} suggesting that a lysine in position 249 may not be imperative for 8-oxoG eversion. However, the long, flexible and polar side chains of Gln249 or Lys253 may retain the ability to hydrogen bond to the O⁸ of 8-oxoG, facilitating the eversion of 8-oxoG but presumably to a lesser extent due to reduced charge on the protein side chain. The hypothesis is supported by presteady-state kinetic experiments, which indicated that the Lys249Gln mutant of OGG1 can evert 8-oxoG but does so more slowly than the wild-type protein.⁴³

3.5. Recognition in the Active Site. At an eversion distance near 17–18 Å (Step IV), the system forms the Gly42–N7 hydrogen bond that is the hallmark of the in-pocket state (Figure 10A,B). For 8-oxoG, the free energy of the system

drops dramatically from step III in exosite to step IV in active site (Figure 3), suggesting that 8-oxoG is better-stabilized in the active site than in the intermediate exosite. A rationale for at least part of this energetic difference is suggested by the 8-oxoG complex crystal structure 1EBM, which reveals a discriminatory hydrogen bond between the backbone carbonyl of Gly42 and the protonated N7 of 8-oxoG.¹² For the G complex, the in-pocket state was also captured experimentally (PDB: 3IH7) by using a distal artificial cross-link (Ser292Cys) to force OGG1 to interrogate a G:C pair.¹⁷ In that crystal structure, G occupies almost the same position as does the in-pocket 8-oxoG seen in 1EBM despite the unfavorable interaction between the backbone carbonyl of Gly42 and the lone pair of electrons on N7 of G.¹⁷ This is probably an artificial effect induced by the distal cross-link, as suggested by previous un-cross-linked MD simulations indicating that G was unstable in the 3IH7 crystallographic in-pocket position.¹⁷ In our simulations, the in-pocket end point for G was not generated from 3IH7; instead, it was generated from 1EBM with the in-pocket 8-oxoG changed to G. Nevertheless, we also observed that G becomes retracted from the active site even when placed there, and the backbone of Gly42 rotates to avoid the lone pair electrons on N7 of G (Figure 10C). Similar conformational changes for G were also observed in quantum mechanics–molecular mechanics (QM–MM) simulations previously performed to investigate the active site preference for 8-oxoG or G.¹⁰ These observations are reinforced by failure of OGG1 to catalyze base excision of G even when it is forcibly presented in the active site by cross-linking; this catalytic checkpoint

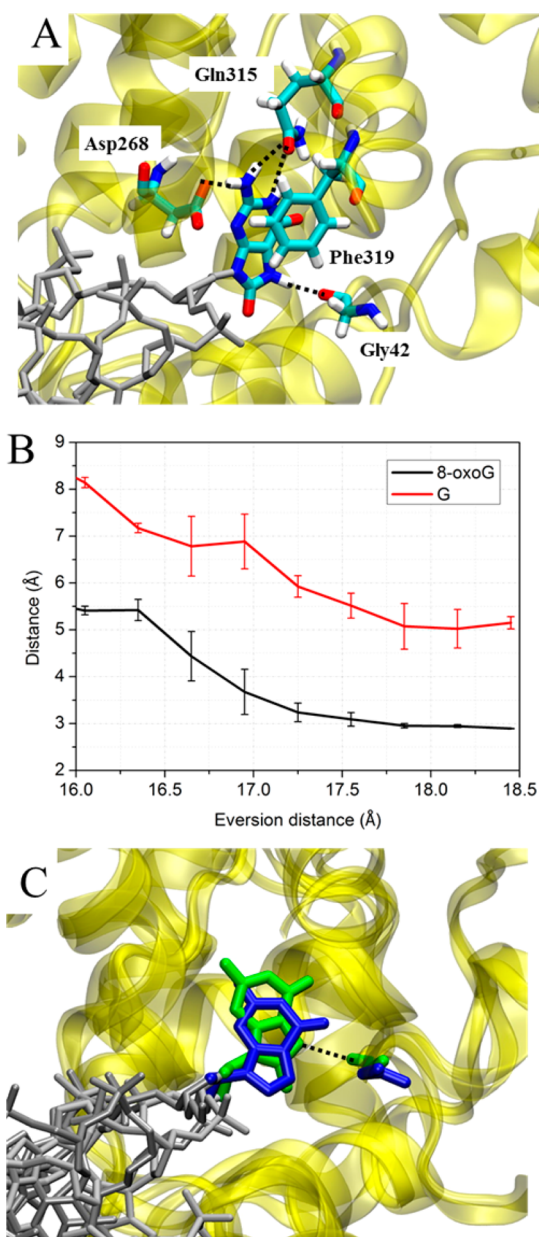


Figure 10. (A) Structure of the 8-oxoG-complex in step IV from Figure 3 with an eversion distance of 16.3 Å. (B) Distance between the carbonyl of G42 and the N7 of either G (red) or 8-oxoG (black). (C) Position of 8-oxoG (green) and G (blue) in the active site, the rest of DNA is shown in gray. The two structures are superimposed on the protein backbone (yellow). The Gly42:8-oxoG H-bond is depicted with a black dotted line. The eversion distances of the 8-oxoG and the G structures are 16.3 and 15.8 Å, respectively.

mechanism is still unclear.¹⁷ Our data suggest that although Gly42 provides the sole interaction that structurally discriminates 8-oxoG and G in the active site, it may be strong enough to keep G from entering the active site as deeply as does 8-oxoG (Figure 10C). Experimental studies have shown that the Gln315Phe, Gln315Trp, Cys253Ile, and Cys253Leu mutations, which perturb the active site disposition of 8-oxoG but do not expel it altogether, can severely reduce the catalytic activity of OGG1,^{16,44} thus suggesting that the catalysis of base excision by OGG1 requires very precise positioning of the reacting moieties. Therefore, the unfavorable interaction of Gly42 may prevent G from achieving the optimal position for catalysis, and

thus, the active-site geometry constitutes the final damage checkpoint of OGG1.

4. DISCUSSION

Previous studies have found considerable similarities in the damage recognition mechanisms of OGG1 and Fpg: both kink the DNA and insert a bulky wedge to disrupt stacking of the target base pair, both extrude 8-oxoG into the active site where the N7–H⁷ of 8-oxoG is recognized by the enzyme, and both insert several plug residues into the DNA to stabilize the extrahelical state, including an arginine residue that recognizes the Watson–Crick interface of the orphaned cytosine.^{12,45}

By comparing the data from this work and our previous work,³⁵ we find additional important evidence for a shared damage recognition mechanism, including the nature of several early checkpoints prior to full eversion. The free energy profiles for 8-oxoG eversion in OGG1 and Fpg show a similar pattern (Figure 11): they have two minima in the first half of eversion

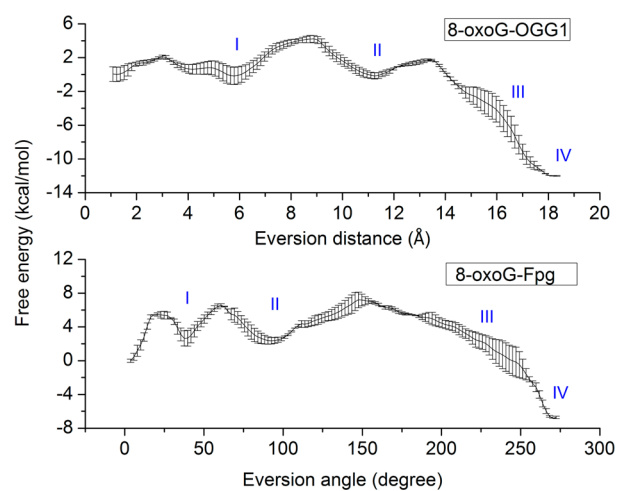


Figure 11. Free-energy profiles for 8-oxoG eversion in OGG1 and Fpg. A total of four distinct steps of eversion are labeled as I, II, III, and IV. Note that the progress variables on the X axes are not the same in the two studies; thus, alignment of the minima between systems is arbitrary. The error bars reflect differences between two independent runs.

(steps I and II) and a global minimum at the extrahelical end point (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 checkpoints for 8-oxoG, because in each step specific interactions are made to the sites modified in the lesion (N7–H⁷ or the O⁸). Notably, such specific interactions are similar in pattern between OGG1 and Fpg (Figure 12): in step I, the slightly everted 8-oxoG is stabilized by the N7–p¹ hydrogen bond; in step II, the Watson–Crick edge of 8-oxoG hydrogen-bonds to p¹, while the O⁸ hydrogen-bonds to a polar side chain (Asn173 in Fpg and H270 in OGG1); in step III, a catalytic residue (Pro1 in Fpg and Lys249 in OGG1) contacts the O⁸ of 8-oxoG; in step IV, 8-oxoG is specifically recognized by a hydrogen bond made to the protonated N7.

Thus, although OGG1 and Fpg are different in overall structure, they employ similar early gates for structural and energetic discrimination that likely facilitate rapid damage recognition on the fast time scale needed for genome-level scanning of DNA. When a glycosylase samples a base pair, it

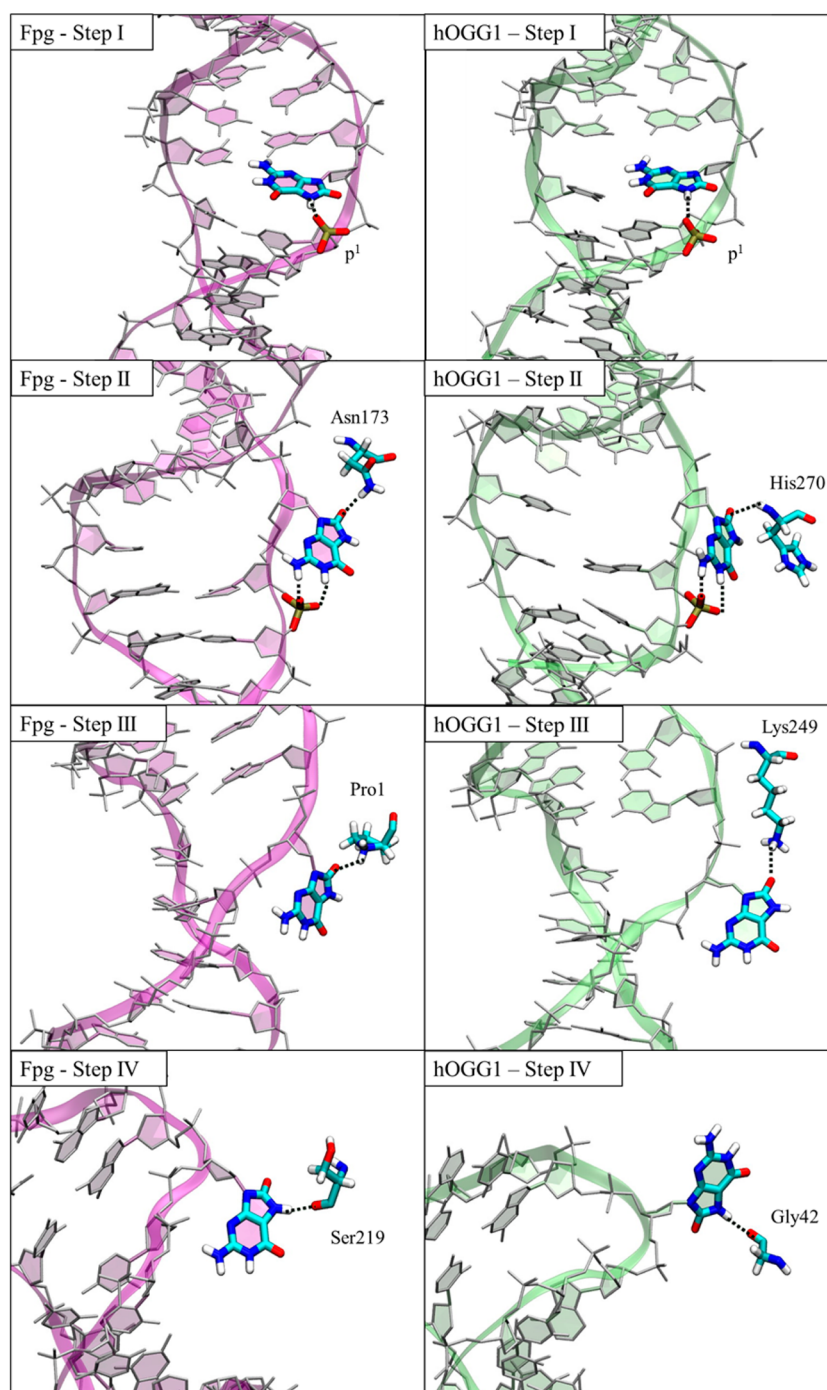


Figure 12. Simulation structures at steps I, II, III, and IV along the 8-oxoG eversion pathway in Fpg and OGG1. H-bonds that contact the O⁸ or the protonated N7 atom of 8-oxoG are indicated by green dotted lines.

enters a kinetic competition between the sampling path with the wedging and plugging residues fully inserted and a 1D diffusion path that requires their withdrawal from the base stack. The energy difference of ~ 5.5 kcal/mol at the earliest eversion intermediate (step I) corresponds to $\sim 10^4$ -fold G versus 8-oxoG discrimination, effectively allowing the enzyme to quickly reject the nondamaged base and resume scanning with minimal time wasted for unproductive encounters. Interestingly, in Fpg, step I contributes little to the selectivity, but step II (~ 10.8 kcal/mol difference) provides $\sim 5 \times 10^7$ -fold discrimination against G.

In summary, we have investigated the damage recognition mechanism of OGG1 by energetically and structurally comparing the base eversion of 8-oxoG and G. On the basis of the free energy profiles as well as the structural analysis that is linked to the energetic differences, we found that OGG1 facilitates 8-oxoG eversion while also hindering G eversion at multiple steps during base eversion, including stages earlier than those suggested by crystal structures of the final everted state. Thus, during a rapid scan of DNA damage, OGG1 could potentially discriminate 8-oxoG from G even when the base is merely slightly everted. Interestingly, in comparable simulations for the Fpg system,³⁵ the potential checkpoints for 8-oxoG

exhibit similar geometries compared to OGG1, suggesting that the damage recognition mechanism may be partly conserved between these two functional analogs, even though they have no sequence or structural similarities.

■ ASSOCIATED CONTENT

📄 Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.6b11433.

Additional computational and experimental method details. Figures showing window distribution and structure overlap, substrate cleavage, hydrogen bond distance, comparison of buckling of target pairs, a comparison of intrahelical endpoint structures, structures for path calculations, reaction coordinates, CPDdb measurement, and PNEB trajectories. A table of DNA properties. (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*carlos.simmerling@stonybrook.edu.

*dzharkov@niboch.nsc.ru.

ORCID

Carlos Simmerling: 0000-0002-7252-4730

Funding

This work was supported by the National Institutes of Health grant no. R01CA017395. Biochemical experiments were supported by the Russian Science Foundation grant no. 14-24-00093 to D.O.Z. C.S. gratefully acknowledges support from Henry and Marsha Laufer.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank the Laufer Center for Physical and Quantitative Biology at Stony Brook University for access to computational resources and support and an NSF Petascale Computational Resource (PRAC) Award from the NSF (OCI-1036208) to C.S.

■ REFERENCES

- Grollman, A. P.; Moriya, M. *Trends Genet.* **1993**, *9* (7), 246–249.
- Lindahl, T.; Wood, R. D. *Science* **1999**, *286* (5446), 1897–1905.
- Lindahl, T. *Nature* **1993**, *362* (6422), 709–715.
- Crenshaw, C. M.; Wade, J. E.; Arthanari, H.; Frueh, D.; Lane, B. F.; Nunez, M. E. *Biochemistry* **2011**, *50* (39), 8463–8477.
- Blainey, P. C.; van Oijen, A. M.; Banerjee, A.; Verdine, G. L.; Xie, X. S. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (15), 5752–7.
- Banerjee, A.; Santos, W. L.; Verdine, G. L. *Science* **2006**, *311* (5764), 1153–7.
- Chen, L.; Haushalter, K. A.; Lieber, C. M.; Verdine, G. L. *Chem. Biol.* **2002**, *9* (3), 345–50.
- Stivers, J. T. *Prog. Nucleic Acid Res. Mol. Biol.* **2004**, *77*, 37–65.
- Kuznetsov, N. A.; Koval, V. V.; Nevinsky, G. A.; Douglas, K. T.; Zharkov, D. O.; Fedorova, O. S. *J. Biol. Chem.* **2007**, *282* (2), 1029–38.
- Banerjee, A.; Yang, W.; Karplus, M.; Verdine, G. L. *Nature* **2005**, *434* (7033), 612–8.
- Dalhus, B.; Forsbring, M.; Helle, I. H.; Vik, E. S.; Forstrom, R. J.; Backe, P. H.; Alseth, I.; Bjoras, M. *Structure* **2011**, *19* (1), 117–27.
- Bruner, S. D.; Norman, D. P.; Verdine, G. L. *Nature* **2000**, *403* (6772), 859–866.

(13) Hamm, M. L.; Gill, T. J.; Nicolson, S. C.; Summers, M. R. *J. Am. Chem. Soc.* **2007**, *129* (25), 7724–5.

(14) Asagoshi, K.; Yamada, T.; Terato, H.; Ohyama, Y.; Monden, Y.; Arai, T.; Nishimura, S.; Aburatani, H.; Lindahl, T.; Ide, H. *J. Biol. Chem.* **2000**, *275* (7), 4956–4964.

(15) Banerjee, A.; Verdine, G. L. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (41), 15020–5.

(16) Radom, C. T.; Banerjee, A.; Verdine, G. L. *J. Biol. Chem.* **2007**, *282* (12), 9182–94.

(17) Crenshaw, C. M.; Nam, K.; Oo, K.; Kutchukian, P. S.; Bowman, B. R.; Karplus, M.; Verdine, G. L. *J. Biol. Chem.* **2012**, *287* (30), 24916–24928.

(18) Slupphaug, G.; Mol, C. D.; Kavli, B.; Arvai, A. S.; Krokan, H. E.; Tainer, J. A. *Nature* **1996**, *384* (6604), 87–92.

(19) Bergonzo, C.; Campbell, A. J.; de Los Santos, C.; Grollman, A. P.; Simmerling, C. *J. Am. Chem. Soc.* **2011**, *133* (37), 14504–6.

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

(21) 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. AMBER 10; University of California: San Francisco, CA, 2008.

(22) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Struct., Funct., Genet.* **2006**, *65* (3), 712–725.

(23) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., 3rd; Loughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92* (11), 3817–29.

(24) Miller, J. H.; Fan-Chiang, C. C. P.; Straatsma, T. P.; Kennedy, M. A. *J. Am. Chem. Soc.* **2003**, *125* (20), 6331–6336.

(25) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79* (2), 926–935.

(26) Bergonzo, C.; Campbell, A. J.; Walker, R. C.; Simmerling, C. *Int. J. Quantum Chem.* **2009**, *109* (15), 3781–3790.

(27) Jónsson, H.; Mills, G.; Jacobsen, K. W., Nudged elastic band method for finding minimum energy paths of transitions. In *Classical and Quantum Dynamics in Condensed Phase Simulations*; Berne, B. J.; Ciccotti, G.; Coker, D. F., Eds.; World Scientific: Hackensack, NJ, 1998; p 385.

(28) Dashti, D. S.; Roitberg, A. E. *J. Chem. Theory Comput.* **2013**, *9* (11), 4692–4699.

(29) Song, K.; Campbell, A. J.; Bergonzo, C.; de los Santos, C.; Grollman, A. P.; Simmerling, C. *J. Chem. Theory Comput.* **2009**, *5* (11), 3105–3113.

(30) MacKerell, A. D.; Banavali, N. K. *J. Mol. Biol.* **2002**, *319* (1), 141–160.

(31) Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. *Biopolymers* **1992**, *32* (5), 523–535.

(32) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13* (8), 1011–1021.

(33) Gilboa, R.; Zharkov, D. O.; Golan, G.; Fernandes, A. S.; Gerchman, S. E.; Matz, E.; Kycia, J. H.; Grollman, A. P.; Shoham, G. *J. Biol. Chem.* **2002**, *277* (22), 19811–19816.

(34) Kuznetsov, N. A.; Koval, V. V.; Zharkov, D. O.; Nevinsky, G. A.; Douglas, K. T.; Fedorova, O. S. *Nucleic Acids Res.* **2005**, *33* (12), 3919–3931.

(35) Li, H.; Endutkin, A. V.; Bergonzo, C.; Campbell, A. J.; de los Santos, C.; Grollman, A.; Zharkov, D. O.; Simmerling, C. *Nucleic Acids Res.* **2016**, *44* (2), 683–694.

(36) Lee, S.; Radom, C. T.; Verdine, G. L. *J. Am. Chem. Soc.* **2008**, *130* (25), 7784–7785.

- (37) Cheng, X.; Kelso, C.; Hornak, V.; de los Santos, C.; Grollman, A. P.; Simmerling, C. *J. Am. Chem. Soc.* **2005**, *127* (40), 13906–18.
- (38) Kuznetsov, N. A.; Bergonzo, C.; Campbell, A. J.; Li, H.; Mechetin, G. V.; de los Santos, C.; Grollman, A. P.; Fedorova, O. S.; Zharkov, D. O.; Simmerling, C. *Nucleic Acids Res.* **2015**, *43* (1), 272–281.
- (39) Giudice, E.; Varnai, P.; Lavery, R. *Nucleic Acids Res.* **2003**, *31* (10), 2703–2703.
- (40) Kirpota, O. O.; Endutkin, A. V.; Ponomarenko, M. P.; Ponomarenko, P. M.; Zharkov, D. O.; Nevinsky, G. A. *Nucleic Acids Res.* **2011**, *39* (11), 4836–4850.
- (41) Sassa, A.; Beard, W. A.; Prasad, R.; Wilson, S. H. *J. Biol. Chem.* **2012**, *287* (44), 36702–10.
- (42) Lu, X. J.; Olson, W. K. *Nucleic Acids Res.* **2003**, *31* (17), 5108–5121.
- (43) Kuznetsova, A. A.; Kuznetsov, N. A.; Ishchenko, A. A.; Saparbaev, M. K.; Fedorova, O. S. *Biochim. Biophys. Acta, Gen. Subj.* **2014**, *1840* (1), 387–95.
- (44) Lukina, M. V.; Popov, A. V.; Koval, V. V.; Vorobjev, Y. N.; Fedorova, O. S.; Zharkov, D. O. *J. Biol. Chem.* **2013**, *288* (40), 28936–28947.
- (45) Fromme, J. C.; Verdine, G. L. *J. Biol. Chem.* **2003**, *278* (51), 51543–8.