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, Carlos Simmerling

Supporting information

Overview of the input structures for PNEB

We used PNEB to generate a path for 8-oxoG eversion and another path for G eversion. For each of the paths, four structures were used as inputs: 1) the structure with the target 8-oxoG or G in the intrahelical state (hererafter in SI referred to as the 8-oxoG-int structure and the G-int structure); 2) the structure with the target 8-oxoG or G slightly everted, which is used to seed the path (referred to as the 8-oxoG-seed1 structure and the G-seed1 structure); 3) another seeding structure with 8-oxoG or G further everted toward the active site pocket (referred to as the 8-oxoG-seed2 structure and the G-seed2 structure); 4) the structure with the 8-oxoG or G in the active site pocket of hOGG1, as the extrahelical endpoint (referred to as the 8-oxoG-ext structure and the G-ext structure).

Generating the input structures from crystallographic coordinates

8-oxoG-seed1 and G-seed1 were generated from coordinates of PDB 2I5W. 8-oxoG-seed2 was generated from coordinates of PDB 1YQK. 8-oxoG-ext and G-ext were generated from PDB 1EBM. All protein mutations present in the crystal structures were reverted back to wild-type residues using Swiss-PdbViewer[1]. Artificial disulfide crosslinks in the crystallization of 1YQK and 2I5W were not included in the simulations. Residues missing electron density at the edge of protein globule ~ 40 Å from the DNA binding site were built using Modeller[2] (G80, D81, K82 in 1EBM, 1YQK and 2I5W, additionally S83 for 2I5W). When building the initial structures, any nucleotides in the crystal structure but not in the 1EBM sequence were deleted, and any in the 1EBM sequence but not in the crystal structure were built in with UCSF Chimera[3] using the coordinates of a standard B-form duplex generated by 3DNA[4]. For base substitutions where G

was changed to 8-oxoG or vice versa, the H^8 or O^8/H^7 atoms were deleted and the desired substitution was then built using the tLEaP module in Amber.[5] Protonation states at pH 7 for all residues in 1EBM were determined using MolProbity[6] and were used as the standard for other structures. The structures were then solvated in ~14,500 TIP3P[7] explicit water molecules in a truncated octahedron box whose boundary was at least 10Å from the any solute atoms.

Each of the solvated structures was minimized and equilibrated in five steps: (i) a 10000-step minimization with positional restraints of 100 $kcal \cdot mol^{-1} \cdot Å^{-2}$ on the heavy atoms; (ii) a 100ps MD simulation in which the system was heated linearly to the target temperature of 310K while the heavy atoms were restrained at 100 $kcal \cdot mol^{-1} \cdot Å^{-2}$; (iii) a 100ps and a 250ps MD simulations with the heavy atoms held by 100 and 10 $kcal \cdot mol^{-1} \cdot Å^{-2}$ restraints, respectively; (iv) a 100ps, a 200ps and a 250ps MD simulations with the heavy atoms of the protein and DNA backbones fixed by 10, 1 and 0.1 $kcal \cdot mol^{-1} \cdot Å^{-2}$ restraints, respectively; (v) a final 2ns unrestrained simulation. During minimization and equilibration SHAKE was employed to constrain bonds involving hydrogen atoms[8], and a 1 fs time step was used. The particle mesh Ewald method was used to approximate long-range Coulombic interactions[9,10] and the nonbonded cutoff was set to 8Å. Through step (iii) to (v) constant temperature of 310K (reflecting the human body temperature) and constant pressure of 1 atm were maintained by the weak-coupling algorithm.[11]

Generating 8-oxoG-int and G-int

For OGG1, the structure containing 8-oxoG in the active site pocket has been solved[12], but the intrahelical interrogation structure is more transient and, unlike for Fpg, has not been reported. Disulfide cross-linking has been used to trap several putative intermediates of base eversion in OGG1; however, as discussed above, formation of the Cys149-Cyt crosslink prevents closing of a transiently opened base pair at the interrogation position, while another crosslink distal from the target base pair (S292C) artificially biases the interrogated base toward the active site pocket[13]. Therefore, to map the full eversion pathway we needed to computationally generate the model intrahelical structures 8-oxoG-int and G-int.

The published crystal structures of OGG1, as well as the intrahelical Fpg structure, provide valuable insights into the conformation of the intrahelical OGG1–DNA complex. In the OGG1–DNA structures the "NNN" motif (containing residues Asn149, Asn150 and Asn151) occupies the void left by the everted base, and the Watson-Crick face of the orphaned cytosine is recognized by direct hydrogen bonds with Arg204[12-17]. On the other hand, while the overall fold does not show significant difference from the bound, the apo structure of OGG1 has the NNN motif and Arg204 retracted[18]. We hypothesize that in the transient state in which OGG1 interrogates an unopened target base pair, OGG1 has not yet adjusted to the bound conformation that is seen in the everted OGG1-DNA structures, but instead remains an apo-like conformation, with NNN motif retracted from the DNA helix and the Arg204 disengaged from the cytosine; otherwise, the target base pair would be disrupted. This assumption is reinforced by the observation that the Fpg–DNA complex shows similar structural differences between its everted state and intrahelical state, both of which have been crystallized. For example, in a structure of Fpg bound to duplex DNA with the interrogated 8-oxoG everted from the duplex, Met77 (using *Geobacillus stearothermophilus* sequence numbering), akin to the NNN motif of OGG1, plugs the DNA void left by 8-oxoG, and Arg112 (the counterpart of Arg204 in OGG1) interacts with the Watson-Crick edge of the orphan cytosine[19]. However, in the structure of Fpg bound to an intact, unopened DNA (stabilized by a disulfide crosslink between Fpg and DNA), the loop containing Met77 is retracted from the DNA, and Arg112 also loses contact with the orphan cytosine[20]. Furthermore, stopped-flow kinetics experiments also suggested that the void-filling of the OGG1 residues into DNA likely occurs after the target base is everted[21]. In addition, the intrahelical Fpg-DNA complex shows significant buckling at the target base pair induced by a wedge Phe114 residue[20], and since OGG1 also has a corresponding wedge residue (Tyr203), target base pair buckling is reasonably expected in the intrahelical OGG1-DNA complex. Stopped-flow kinetics experiments also indicated destabilization of Watson-Crick or stacking interactions in the early stage nonspecific OGG1-DNA complex[21]. Therefore, the intrahelical state model we developed in this work is an apo-like OGG1 with the NNN motif retracted and the Arg204 disengaged from the cytosine of the target pair, interrogating an unopened but

buckled base pair in duplex DNA.

The 8-oxoG-int and the G-int structures were generated from structure 2I5W, whose target base is the least everted among all the OGG1–DNA crystal structures. To achieve an unopened target base pair, the everted target base was forced into the helix to pair with the opposite C; to make space, the protein was simultaneously forced to adopt an apo-like conformation, using crystal structure 1KO9[18] as reference.

For 8-oxoG-int, the initial structure for generating the intrahelical model was taken from the end of step (iv) of the 8-oxoG-seed1 equilibration (see above), then two steps were performed. (a) The 8-oxoG and the opposite C were forced to pair by using distance restraints on the Watson-Crick hydrogen bonds for 1ns, while the rest of DNA was held with 1 $kcal \cdot mol^{-1} \cdot Å^2$ positional restraints. Meanwhile, the plug residues were retracted using 1 $kcal \cdot mol^{-1} \cdot Å^{-2}$ targeted RMSD restraints applied on the protein backbone plus the side chains of N149 and R204, with the coordinates of the 1KO9[18] structure used as a reference for RMSD. (b) The structure was then equilibrated in a 250ps simulation with 0.1 $kcal \cdot mol^{-1} \cdot Å^2$ restraints on the backbone heavy atoms, followed by 2ns unrestrained MD.

For G-int, the initial structure was taken after step (a) of the 8-oxoG-int generation. The 8-oxoG was modified to G, followed by 10000-step minimization with the heavy atoms of the protein-DNA complex restrained, followed by an equilibration corresponding to step (b) of the 8-oxoG-int equilibration.

Similar to simulations starting from crystallographic intrahelical Fpg in our previous study[22], no tendency of base opening of 8-oxoG:C or G:C was seen during 2ns MD of our intrahelical OGG1 models, with stable hydrogen bonds (**Figure S4**) and buckle angle near -20° (**Figure S5**). As shown in **Figure S6**, the target and 3' pairs are buckled, probably by the wedge residue, in a similar way as that in Fpg; while Arg204 and the NNN motif in OGG1 are retracted from DNA, similar to Arg111 and Met76 in Fpg. Other parameters of the interrogated pair, including shear, stretch, propeller and opening, are also comparable between the OGG1 and Fpg complexes (**Table S1**), likely due to the similarity in local geometry and interactions near the target base

pair in these two enzymes. Overall, we conclude that the intrahelical OGG1 models we built have similar properties and stability as compared to simulations starting from the Fpg intrahelical crystal structure, and are therefore reasonable models for generation of the eversion pathway as discussed below.

Generating G-seed2

G-seed2 was generated by performing 50ps MD simulation on the G-seed1 structure, forcing the target G to further evert to 9.5 Å of the eversion distance using 10 kcal·mol^{-1}·Å^{-2} restraint.

Nudged elastic band calculations and use of two seeds

In our previously reported studies of base eversion in Fpg, a single seed was sufficient to direct sampling along the major groove. In the present study, however, using either one of the two seeds alone was unable to direct sampling away from the minor groove at the initial minimization step of NEB (prior to path sampling and optimization, suggesting that the path length rather than energy was responsible). We believe that this simply reflects how well the seeds direct sampling at the initial minimization when the sampling along the path has large gaps and thus large spring forces. In the Fpg study, no crystal structure of an eversion intermediate was available and we generated a major groove seed by using targeted MD to force significant eversion along the major groove. Here, the availability of two crystallographic intermediates along the major groove prompted us to use them as seeds. However, as shown in **Figure Sup1,** the extent of eversion in these seeds is highly similar to the endpoint structures and thus it is not surprising that they are less effective at directing sampling away from the most geometrically direct minor groove path. Using both crystallographic-based seeds successfully reproduced a major groove path, also consistent with what we showed is the low-energy path for Fpg.

For the first 600ps and last 500ps of the PNEB simulations, the temperature was set to 310K to correspond to the biologically relevant environment for OGG1; whereas in our previous study 330K was used for a thermophilic Fpg variant. The spring force constant at the first 100ps path optimization was set to 1 kcal·mol^{-1}·Å^{2}, whereas in our previous study on Fpg, the same

parameter was set to 2 kcal·mol^{-1}·Å^2. This is because the OGG1 path has two seeding intermediates (vs one for Fpg), allowing smaller spring forces to be effective.

Choice of reaction coordinate for base eversion PMFs

For our prior base eversion calculations in Fpg[22,23], we used a modified COM pseudodihedral angle (CPDb)[24] which describes how far the base is everted from the duplex (**Figure S8**). The free energy profiles we published for Fpg were mapped in two dimensions using CPDb and the glycosidic angle of the everting base. For OGG1, however, we found that the CPDb variable may not accurately map the base eversion progress, since it does not separate the exo-site and in-pocket states, both of which adopt CPDb values angle near 210–220° (**Figure S9**). Thus, we developed a new reaction coordinate, "eversion distance" (**Figure 2**), which better separates the exo-site and the active site pocket positions. In **Figure S10**, we show snapshots from the OGG1 eversion pathways plotted using eversion distance and CPDb. We observe that although structures at the end of the path are not distinguished by CPDb, the eversion distance continues to increase (~15 Å to ~18 Å), providing a better description of the late stages of eversion.

With this new measure, we also found that use of the glycosidic angle as a second dimension in the free energy landscape was unnecessary, because the eversion distance used here is also sensitive to the base rotation along the glycosidic bond (**Figure S11**). As the base enters the active site pocket from the exo-site, where the glycosidic angle decreases from ~0° to ~-120°, the eversion distance accordingly increases from ~15 Å to ~18 Å. Therefore, the eversion distance was used as the sole reaction coordinate in umbrella sampling in this work.

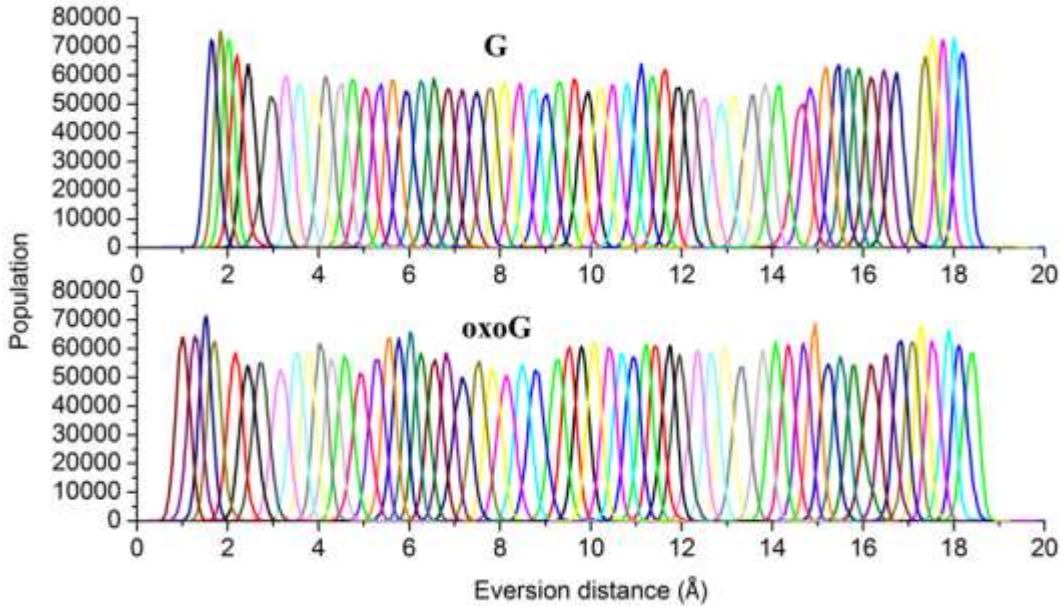

Figure S1. Overlap of the window distributions along the eversion distance in umbrella sampling. Each curve represents the population distribution of a single window.

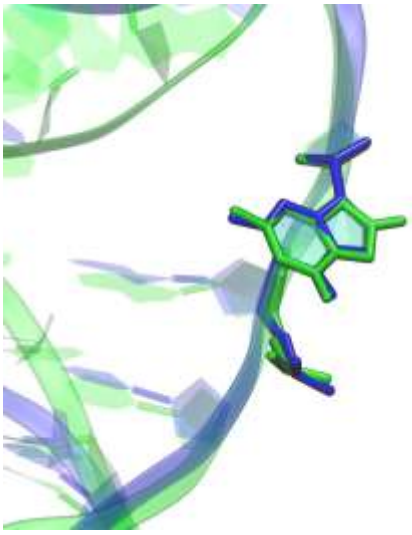

Figure S2. Overlap of structures at barrier region between Steps I and II, with 9.1 Å eversion. Unlike at the neighboring minima where damage-specific H-bonds are formed, no significant difference is observed at the barrier and both show close approach of O6 to p[1].

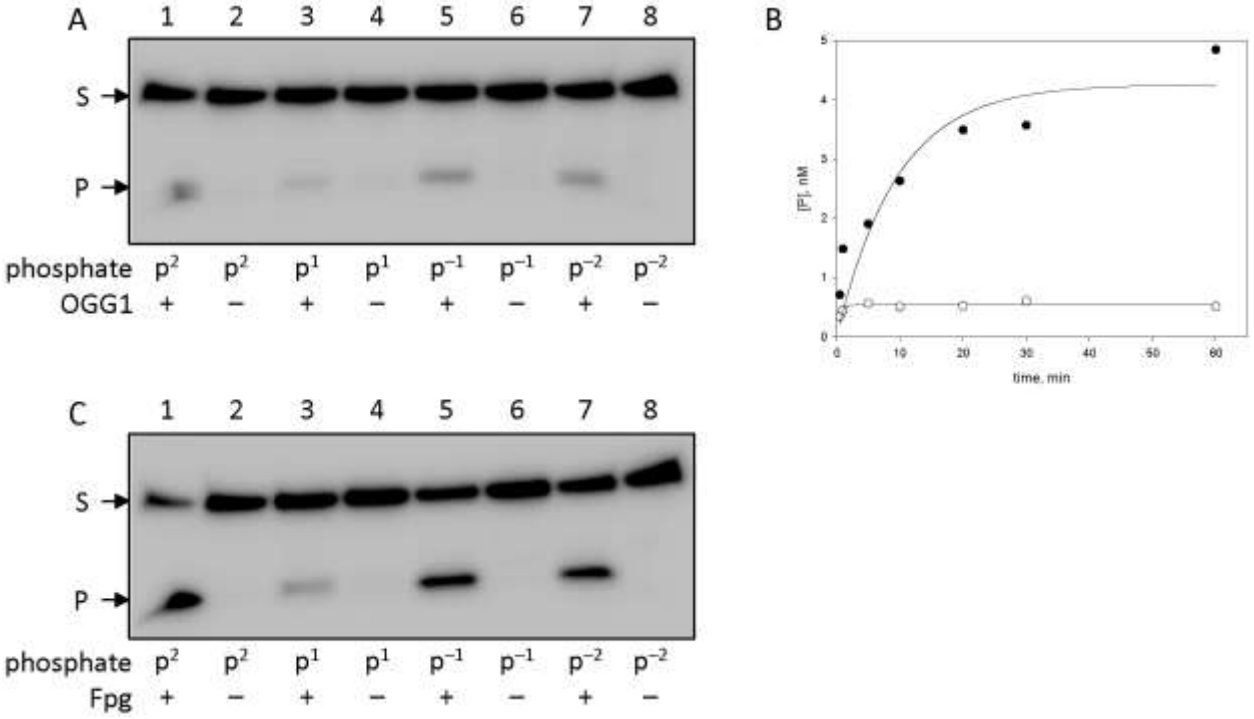

Figure S3. **A**, cleavage of phosphorothioate-substituted 8-oxoG-containing substrate by OGG1 under multiple-turnover conditions (10 nM enzyme, 100 nM substrate, 20 min at 37°C). Odd lanes contain OGG1, even lanes contain no enzyme. S, substrate, P, product. **B**, cleavage of all-phosphate (black symbols) and p^1 phosphorothioate-substituted (white symbols) 8-oxoG-containing substrate by OGG1 under single-turnover conditions (100 nM enzyme, 5 nM substrate, 4°C). **C**, cleavage of phosphorothioate-substituted 8-oxoG-containing substrate by Fpg under multiple-turnover conditions (10 nM enzyme, 100 nM substrate, 20 min at 37°C). Odd lanes contain Fpg, even lanes contain no enzyme. S, substrate, P, product.

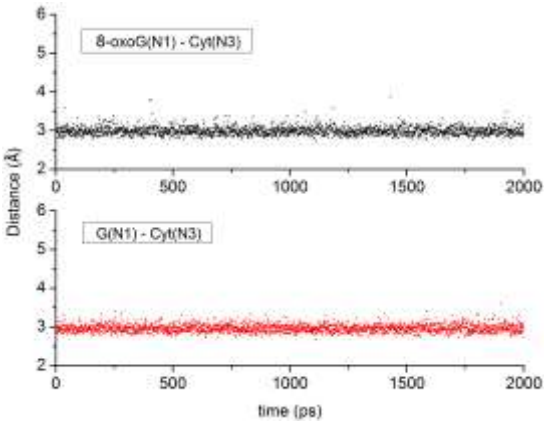

Figure S4. Distance of hydrogen bond 8-oxoG/G(N1) - C(N3) of the target pair during the 2ns unrestrained MD of the equilibration for the OGG1 intrahelical structures, respectively, showing the stability of the target base pair in our intrahelical model.

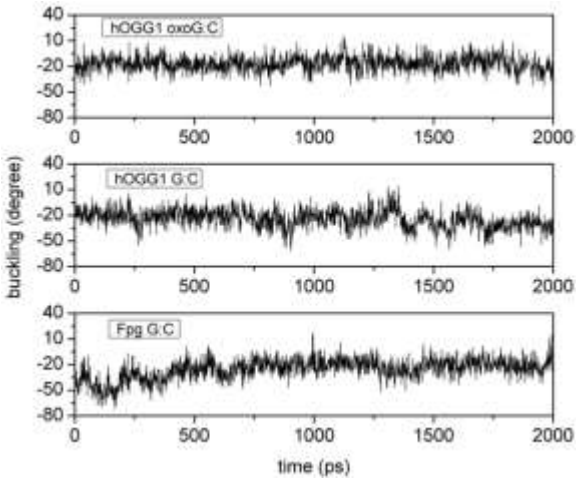

Figure S5. Comparison of buckling of the target pair in the OGG1 intrahelical models and in the Fpg model during the 2ns unrestrained MD of the equilibration.

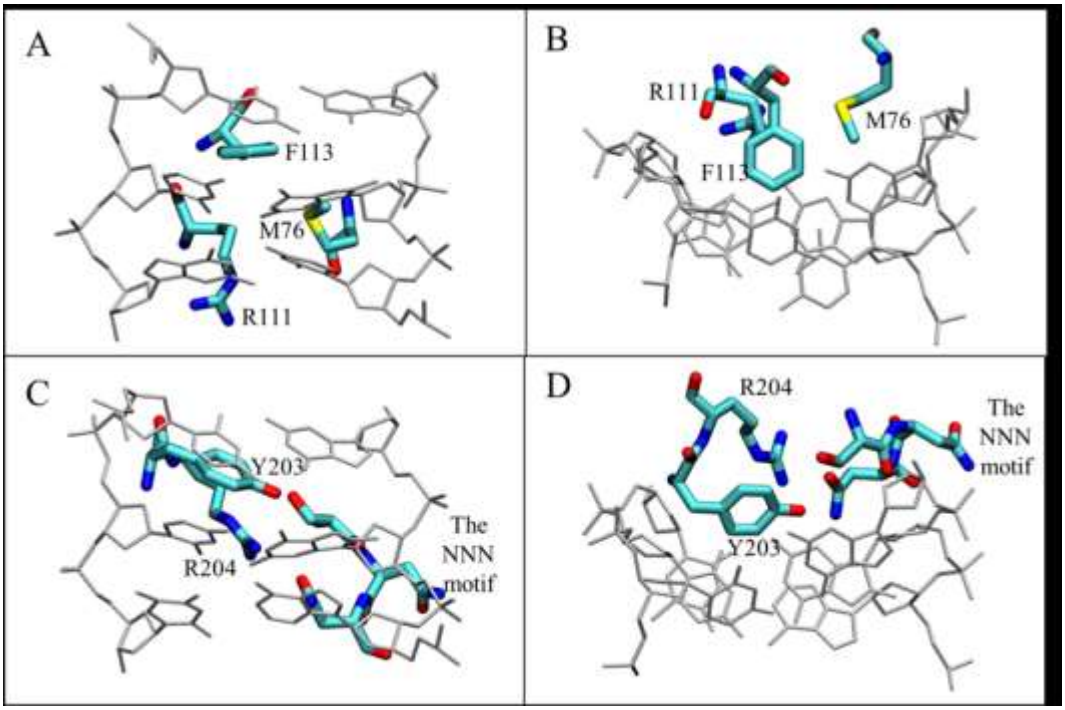

Figure S6. Comparison of intrahelical endpoint structures of Fpg (A,B) and OGG1(C,D). B and D are same structure as A and C, respectively, but viewed from the top of the DNA, highlighting the depth of insertion. The DNA is in grey and the inserting residues are colored by atom.

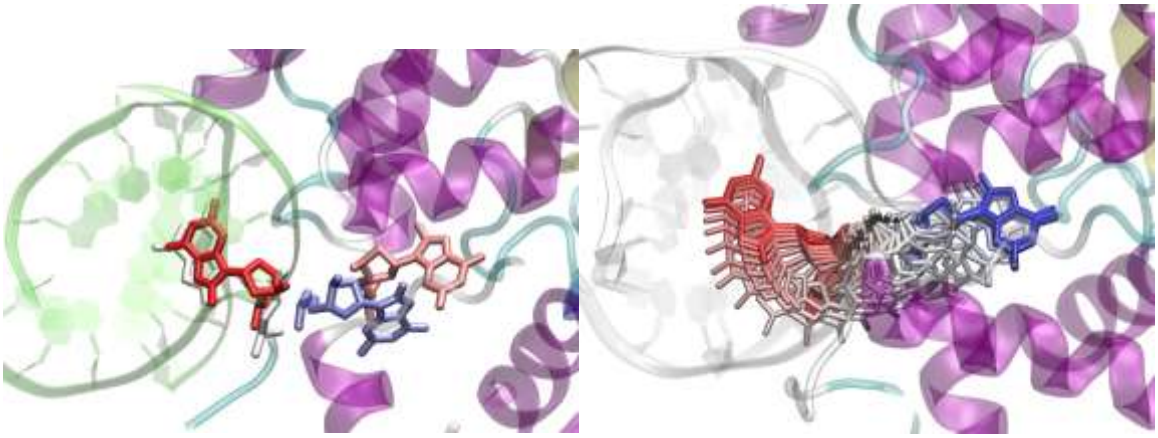

Figure S7. Left: initial structures for the NEB path calculation for 8-oxo-G eversion, with the 8-oxoG-int, 8-oxo-G-seed1, 8-oxo-G-seed1 and 8-oxoG-ext shown in white, red, blue and pink, respectively. Right: Final NEB sampling along the major groove, showing curvature that is not well represented in the initial seeds, giving rationale for the need to use both seeds for effective guiding of the path.

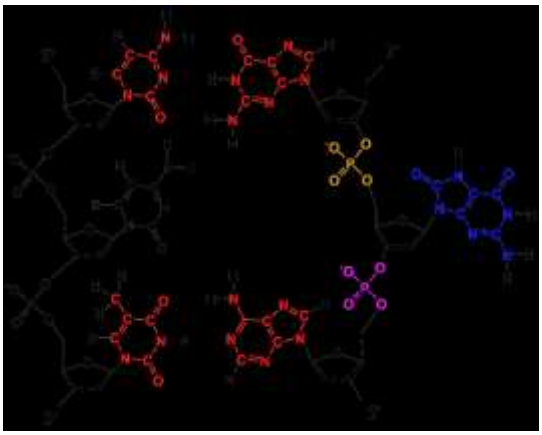

Figure S8. The reaction coordinate in our previous study of Fpg, CPDb, is the pseudodihedral angle calculated using 4 points: (1) the center of mass (COM) of the four flanking bases colored in red; (2) and (3), COM of the 3'- and the 5'- phosphate group of the target base, respectively (colored in orange and purple); (4) COM of the base of the target G/8-oxoG (colored in blue).

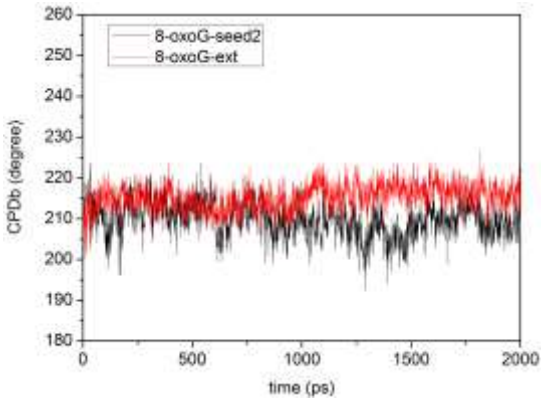

Figure S9. Measurement of CPDb during the 2ns MD of the 8-oxoG-seed2 (8-oxoG in exo site) and 8-oxoG-ext (8-oxoG in active site pocket).

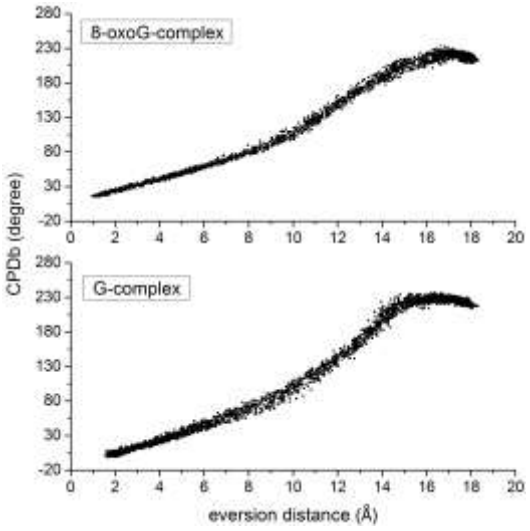

Figure S10. PNEB trajectories of eversion of 8-oxoG and G plotted two-dimensionally in CPDb vs. eversion distance. The eversion angle keeps roughly the same value after the eversion distance reaches 16 Å, indicating that eversion angle is unable to distinguish progress along this portion of the eversion path.

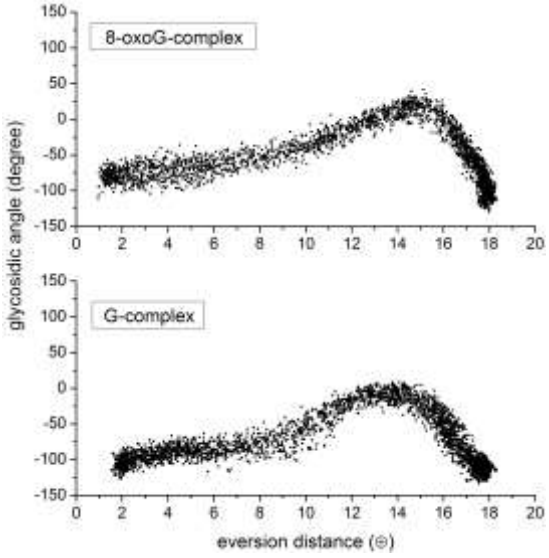

Figure S11. PNEB trajectories of eversion of 8-oxoG and G plotted two-dimensionally in glycosidic angle vs. eversion distance. Along the whole pathways, the change of glycosidic angle is accompanied by a change in eversion distance, showing that the distance is sufficient to map the base rotation part of the eversion process.

Table S1. Comparison between the target base pair parameters of the OGG1 8-oxoG-complex and G-complex of and the Fpg G-complex.

	Shear	Stretch	Stagger	Propeller	Opening
OGG1 8-oxoG-complex	0.0°	0.0°	0.5°	4.3°	-1.7°
OGG1 G-complex	0.0°	-0.1°	0.1°	16.6°	-2.0°
Fpg G-complex	0.2°	0.1°	0.5°	9.4°	2.0°

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

References

(1) Guex, N.; Peitsch, M. C. *Electrophoresis* **1997**, *18*, 2714.

(2) Sali, A.; Blundell, T. L. *Journal of Molecular Biology* **1993**, *234*, 779.

(3) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *Journal of Computational Chemistry* **2004**, *25*, 1605.

(4) Lu, X. J.; Olson, W. K. *Nucleic Acids Research* **2003**, *31*, 5108.

(5) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *Journal of Computational Chemistry* **2005**, *26*, 1668.

(6) Chen, V. B.; Arendall, W. B.; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. *Acta Crystallographica Section D-Biological Crystallography* **2010**, *66*, 12.

(7) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *Journal of Chemical Physics* **1983**, *79*, 926.

(8) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *Journal of Computational Physics* **1977**, *23*, 327.

(9) Cheatham, T. E.; Miller, J. L.; Fox, T.; Darden, T. A.; Kollman, P. A. *Journal of the American Chemical Society* **1995**, *117*, 4193.

(10) Darden, T.; York, D.; Pedersen, L. *Journal of Chemical Physics* **1993**, *98*, 10089.

(11) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. *Journal of Chemical Physics* **1984**, *81*, 3684.

(12) Bruner, S. D.; Norman, D. P.; Verdine, G. L. *Nature* **2000**, *403*, 859.

(13) Crenshaw, C. M.; Nam, K.; Oo, K.; Kutchukian, P. S.; Bowman, B. R.; Karplus, M.; Verdine,

G. L. *J Biol Chem* **2012**, *287, 26916*.

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

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

(16) Banerjee, A.; Yang, W.; Karplus, M.; Verdine, G. L. *Nature* **2005**, *434*, 612.

(17) Norman, D. P.; Chung, S. J.; Verdine, G. L. *Biochemistry* **2003**, *42*, 1564.

(18) Bjoras, M.; Seeberg, E.; Luna, L.; Pearl, L. H.; Barrett, T. E. *J Mol Biol* **2002**, *317*, 171.

(19) Fromme, J. C.; Verdine, G. L. *Journal of Biological Chemistry* **2003**, *278*, 51543.

(20) Banerjee, A.; Santos, W. L.; Verdine, G. L. *Science* **2006**, *311*, 1153.

(21) Kuznetsov, N. A.; Koval, V. V.; Nevinsky, G. A.; Douglas, K. T.; Zharkov, D. O.; Fedorova, O. S. *J Biol Chem* **2007**, *282*, 1029.

(22) Li, H.; Endutkin, A. V.; Bergonzo, C.; Campbell, A. J.; de los Santos, C.; Grollman, A.; Zharkov, D. O.; Simmerling, C. *Nucleic acids research* **2016**, *44*, 683.

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

(24) Song, K.; Campbell, A. J.; Bergonzo, C.; de los Santos, C.; Grollman, A. P.; Simmerling, C. *Journal of Chemical Theory and Computation* **2009**, *5*, 3105.