Energetic Preference of 8-oxodG Eversion Pathways in a DNA Glycosylase

Supplementary information

Christina Bergonzo[†], Arthur J. Campbell[†], Carlos de los Santos[‡], Arthur P. Grollman[‡], Carlos Simmerling[†,*]

[†]Department of Chemistry, Stony Brook University, Stony Brook, NY 11794-2700

[‡]Department of Pharmacological Sciences, Stony Brook University, Stony Brook, NY 11794-8651

[*]CORRESPONDING AUTHOR carlos.simmerling@stonybrook.edu

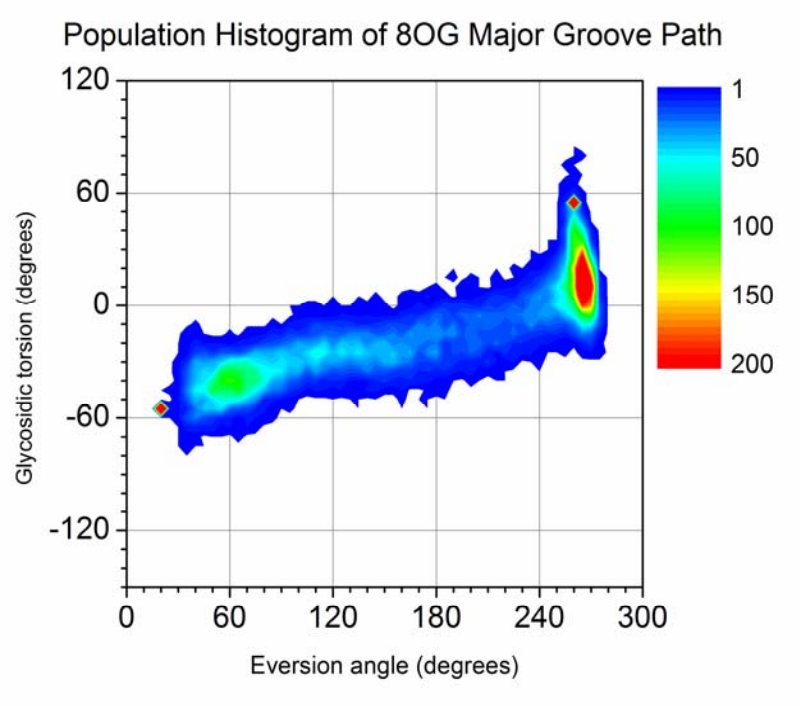

Figure S1 Population histogram of PNEB data along major groove path during final equilibration at 330K over 500ps. Scale is given in number of structures.

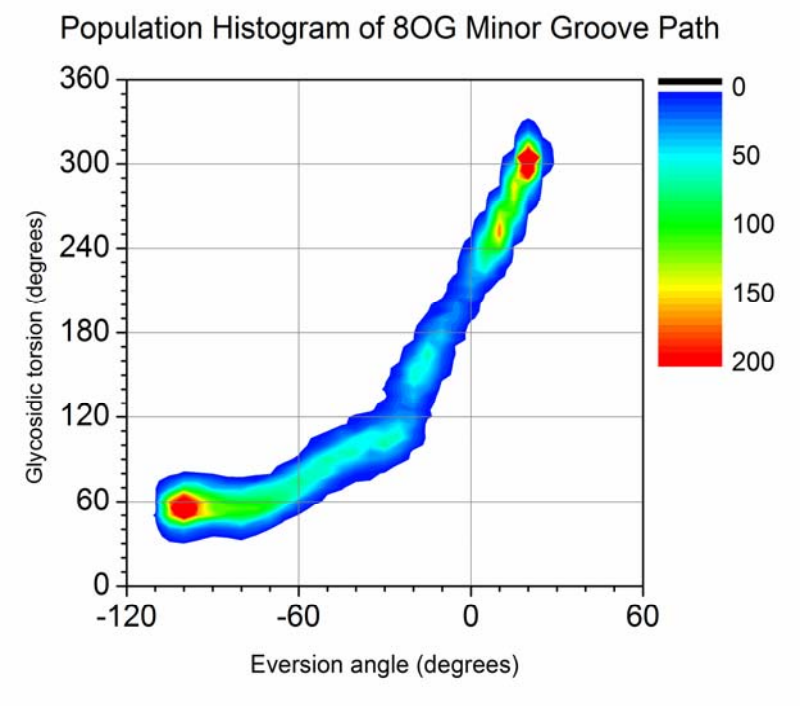

Figure S2 Population histogram of PNEB data along minor groove path during final equilibration at 330K over 500ps. Scale is given in number of structures.

Simulation methods summary

Protein-DNA systems were built using the 2F5O[1] crystal structure of Fpg crosslinked to DNA interrogating an undamaged intrahelical guanine base for the intrahelical endpoint, and the 1R2Y[2] crystal structure of a catalytically inactive mutant trapping 8-oxoguanine (8-oxoG) in the active site for the extrahelical endpoint. Disulfide crosslinks used to trap the complexes were not included in these simulations. The DNA sequence of the 1R2Y structure was standardized to match that of the 2F5O structure by manually deleting the nucleobase from the pdb and building in the standard (2F5O) sequence's nucleobases using the Leap module of AMBER.[3] The standard sequence is as follows: 5'-(AGGTAGATCCGGACGC)-3' and 5'-(TGCGTCC**8OG**GATCTACC)-3'. Additionally, the intrahelical endpoint derived from the 2F5O structure required modification to add coordinates for the

catalytic loop (residues 217 to 237 inclusive, missing in the electron density map) based on the ordered 1R2Y structure. Residue 2 of the 1R2Y structure was mutated back to its wild type sequence of E from Q. The N-terminal proline was modeled as neutral.[4] Parameters for 8-oxoG were taken from Miller et al.[5] Zinc was modeled using the Stote non-bonded model.[6] The 2F5O structure is referred to as "intrahelical" and the 1R2Y structure as "everted."

Molecular dynamics (MD) simulations were carried out with the Sander module of the Amber10 suite of programs.[3] The Amber ff99SB[7] force field with the parmbsc0[8] corrections for nucleic acid backbone torsions was used. During minimization and equilibration constant pressure (1atm) and temperature (330K) were maintained using the weak-coupling algorithm with a coupling time constant of 0.5ps. [9] SHAKE was used to constrain bonds to hydrogen.[10] A time step of 1 fs was employed. The particle mesh Ewald method was used for calculating electrostatic energy with a nonbonded cutoff of 8Å on direct space interactions.[11,12]

The structures were initially minimized using the GB-OBC model and mbondi2 radii set.[13] Coordinates were restrained with 5 kcal mol^{-1} positional restraints on all heavy atoms and minimized using 200 steps of steepest descent minimization, followed by 200 steps of conjugate gradient minimization. The minimized structures were solvated using ~8,000 TIP3P[14] explicit solvent molecules in a truncated octahedron box with 8Å distance between the solute and boundary. They were heated linearly over 100ps to a target temperature of 330K, using positional restraints on all atoms of the protein and DNA, allowing only solvent to move. This was followed by five 1000-step steepest descent minimization rounds in which restraints on sidechains and bases were gradually decreased from 10.0, 5.0, 2.5, 1.0 and 0.0 kcal mol^{-1} A^{-2}. Four 5000-step cycles of MD with decreasing restraints on the protein

and DNA backbone were then performed. A 20ns NVT unrestrained dynamics simulation was carried out using a Langevin thermostat with a collision frequency of 1.0 ps^{-1} and a timestep of 2fs.[15]

Path calculations for the minor groove path were performed by linking 16 copies of the intrahelical structure and 16 copies of the everted structure, to total 32 structures along the path. The major groove path needed to be seeded with an intermediate structure. Targeted MD was performed over 30ps using the intrahelical structure, with a restraint value of 0.0305 kcal/mol deg^2 and a target value of 195° of the eversion dihedral.[16] The restraint mask was applied to heavy atoms of the entire protein and DNA backbone for targeted MD. 30 frames from the targeted MD trajectory were used to seed the PNEB major groove pathway, in addition to the intrahelical and everted endpoint structures, totaling 32 structures along the starting PNEB path.

For PNEB and umbrella sampling simulations, an NVT ensemble was used. NEB forces were calculated for all atoms of the solute and not the solvent. A simulated annealing path optimization procedure was adapted from Mathews and Case 2006.[17] Initial spring forces of 10 kcal/mol/A^2 and an initial collision frequency of 100 ps^{-1} were used. The first stage of path optimization for 80ps at 330K was followed by equilibration of all structures at this temperature for 500ps. Heating of the path to 386K took place over 60ps, with an increased spring constant of 25 kcal/mol/A^2 and decreased collision frequency of 75 ps^{-1}. Equilibration at 386K for 220ps was followed by annealing to 330K over 100ps. Final equilibration at 330K to generate populations for each structure was performed over 500ps.

The process used to generate PMFs from PNEB structures using umbrella sampling was the same as described previously.[18] Each window's starting structure was restrained with a force constant of 0.0609 kcal/mol deg^2 in both the eversion and glycosydic torsion angles and was simulated in an NVT ensemble for 250 ps. The temperature of each window was held constant at 330K using a Langevin[15]

thermostat with a collision frequency of 75.0 ps^{-1}. For the minor groove simulations, a total of 201 and 236 windows were run with 10° spacing for the glycosidic torsion and eversion angles. For the major groove simulations, a total of 576 and 575 windows were run with 5° spacing for the eversion angle and 10° spacing for the glycosidic torsion. 2D WHAM analysis was used to calculate free energies of the combined minor and major groove pathways.[19] To assess convergence, two entirely different sets of starting structures were taken from each PNEB pathway and the umbrella sampling procedure was performed independently for each set (Figure S3).

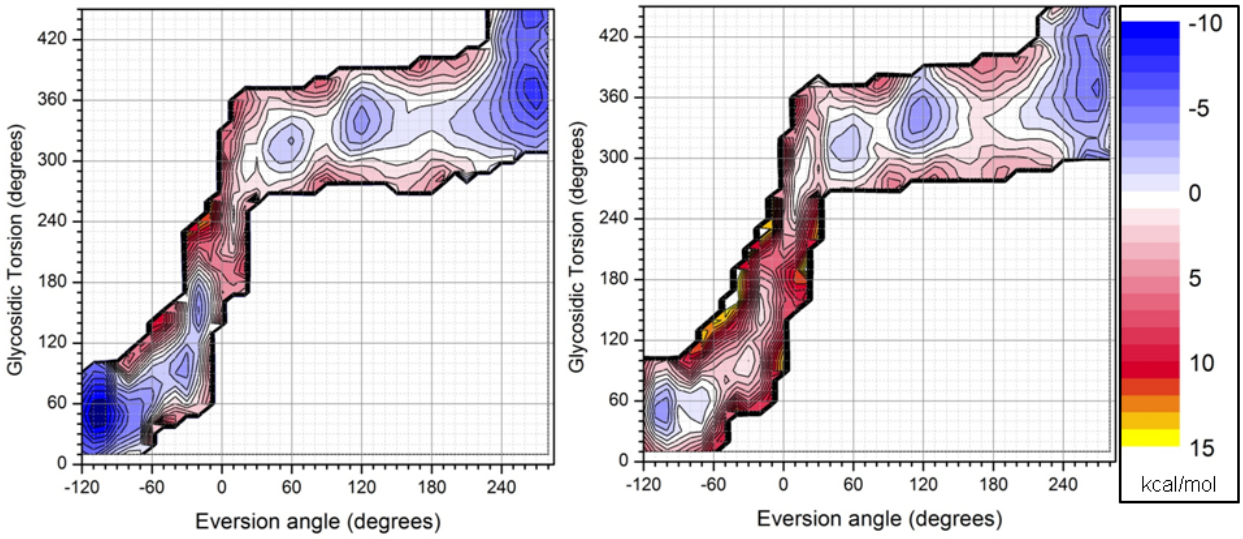

Figure S3. PMF profiles from independent initial structures.

Complete reference 16 in main text is given below:

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

References

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

(2) Fromme, J. C.; Verdine, G. L. *J Biol Chem* 2003, *278*, 51543.

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

(4) Perlow-Poehnelt, R. A.; Zharkov, D. O.; Grollman, A. P.; Broyde, S. *Biochemistry* 2004, *43*, 16092.

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

(6) Stote, R. H.; Karplus, M. *Proteins-Structure Function and Genetics* 1995, *23*, 12.

(7) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* 2006, *65*, 712.

(8) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. *Biophysical Journal* 2007, *92*, 3817.

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

(10) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J Comput Phys* 1977, *23*, 327.

(11) Cheatham, T. E.; Miller, J. L.; Fox, T.; Darden, T. A.; Kollman, P. A. *J Am Chem Soc* 1995, *117*, 4193.

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

(13) Onufriev, A.; Bashford, D.; Case, D. A. *Proteins-Structure Function and Bioinformatics* 2004, *55*, 383.

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

(15) Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. *Biopolymers* 1992, *32*, 523.

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

(17) Mathews, D. H.; Case, D. A. *J Mol Biol* 2006, *357*, 1683.

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

(19) O'Neil, L. L.; Grossfield, A.; Wiest, O. *Journal of Physical Chemistry B* 2007, *111*, 11843.