Molecular simulations reveal a common binding mode for glycosylase binding of oxidatively damaged DNA lesions

Kun Song[†], *Catherine Kelso*[$], *Carlos de los Santos*[#], *Arthur P. Grollman*[#], *and Carlos Simmerling*[†$*]

[†]Department of Chemistry, [#]Department of Pharmacological Sciences and [$]Center for Structural Biology

Stony Brook University, Stony Brook, NY 11794-3400

Email: carlos.simmerling@stonybrook.edu

Supporting Information

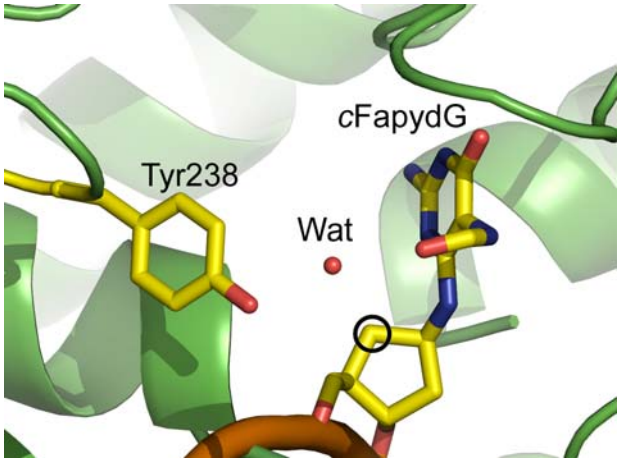

Figure S1. The structure of *c*FapydG in X-ray structure 1XC8. The O4' atom in normal nucleotides was replaced with a carbon(shown in black circle). There is a water bridge between *c*FapydG and the neighboring Tyr238.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14		
5'	G	C	G	A	G	A	A	A	C	A	A	A	G	A		3'
3'		G	C	T	C	T	T	T	FOX	T	T	T	C	T	C	5'
		28	27	26	25	24	23	22	21	20	19	18	17	16	15	

Table S1. The sequence of DNA duplex in X-ray structure 1XC8.

	1	2	3	4	5	6	7	8	9	10	11	12	
5'	G	T	A	G	A	C	C	T	G	G	A	C	3'

3’	C	A	T	C	T	G	8OG	A	C	C	T	G	5’
	24	23	22	21	20	19	18	17	16	15	14	13	

Table S2. The sequence of DNA duplex in X-ray structure 1R2Y. FOX: FapydG. 8OG: 8-oxo-G.

MATERIALS AND METHODS

System preparation.

All initial structures were built using the Leap module of Amber (version 8)(*1*), based on the crystal structure of the *B. stearothermophilus* (*Bacillus stearothermophilus*) Fpg/DNA complex with 8OG in the *syn* conformation (1R2Y.pdb)(*2*). The missing N-terminal methionine residue was not added. The residue index for the rest of the sequence was reduced by one from the PDB structure. The sequence of the DNA duplex was also taken from 1R2Y (Table S2) d[GTAGACCTGGAC]·[GTCCAG*GTCTAC] (where G* is 8OG or FapydG). This sequence was chosen to match the sequence in the X-ray structure 1R2Y. 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 6 Å buffer between the box edge and the nearest protein atom, which was sufficient because there was no large conformation change of the complex system during our simulations. The TIP3P model(*3*) was used to explicitly represent water molecules. Water was added from a pre-equilibrated box and crystallographic water was not retained in building the simulated systems. Following previous studies(*4, 5*), 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 Perlow-Poehnelt et al. (*5*). Force field parameters for 8OG were obtained from Miller et al.(*6*). Force field parameters for FapydG were generated in our previous work(*7*) and the initial conformation of FapydG was obtained from X-ray structure 1XC8. Zinc was modeled using the Stote non-bonded model (q = +2e-, σ = 1.7 Å, ε = 0.67 kcal/mol)(*8*). The remaining protein and nucleic acid parameters employed Amber ff99 (*9, 10*), with modified protein backbone parameters to reduce the alpha-helical bias of those force fields(*11*).

Molecular Dynamics simulations.

All molecular dynamics simulations were carried out with the SANDER module in Amber (*1*). Solvated systems were minimized and equilibrated in three steps: (i) 50 ps MD simulation 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; (iii) Four cycles of 5000 steps 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 (*1*) was used to constrain bonds involving hydrogen atoms. Time step was 0.002 picosecond. The non-bonded cutoff was 8 Å. The particle mesh Ewald method(*12, 13*) was used to calculate long-range electrostatics. The charged grid spacing was 1 $Å^3$. The order of the B-spline interpolation was 4. Constant pressure (1 atm) was maintained with isotropic position scaling and coupling time of 1 picosecond. Constant temperature (300 K) was maintained by the weak coupling algorithm (*14*) with coupling time of 1 picosecond.

Umbrella sampling and potential of mean force calculations.

Umbrella sampling (*15-18*) was used to calculate the potential of mean force (PMF) as a function of 8OG glycosidic angle in the binding site and the dihedral angle of FapydG C4-C5-N7-C8. For calculating the PMF as a function of 8OG glycosidic angle, 36 starting structures were generated using MOIL-VIEW(*19*) by rotating 8OG glycosidic angle in 10° increments from 10° to 360°. These initial structures were energy minimized and one independent 200ps 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}×radian^{-2}. To prevent structural instability at the transition state of the rotation, and to ensure that the PMF represented as closely as possible the energy profile in the structure observed in the crystal, residues farther than 10 Å from 8OG in the initial structure were restrained using Cartesian coordinate positional restraints (force constant 2 kcal×mol^{-1}×Å^{-2}). The other parameters of these simulations were the same as the standard MD simulations. The resulting PMF was obtained by WHAM analysis(*16-18*) of the data using a program provided by Alan Grossfield (freely available at dasher.wustl.edu/alan). Error bars were calculated using the difference between the results calculated from the second half and the entire data set.

For calculating the PMF as a function of the dihedral angle of FapydG C4-C5-N7-C8, 25 starting structures were generated in 10° increments from -150° to 90°. In each window the dihedral angle was restrained to the reference value using a harmonic restraint with a force constant of 100 kcal×mol^{-1}×radian^{-2}. The other parameters of these simulations were the same as the standard MD simulations. The resulting PMF was also obtained by WHAM analysis.

The images in this article were made using VMD(*20*).

Sequence Selection and Alignment

Sequence alignment was used to measure the frequency of the appearance of one key residue E77. In this study the sequence selection is important. There may be more available sequences in certain phyla of bacteria, which may lead to the misinterpretation of the conservation number if all currently available sequences are used. The National Center for Biotechnology Information's Cluster of Orthologous Groups (COG)(*21*) provides a set of sequences evenly distributed among main branches. The cluster of formamidopyrimidine-DNA glycosylase (COG0266) contains 50 proteins from 37 genomes. The sequences were aligned using BLASTP 2.2.4(*22*) and the ones with the PELPEV motif at the N-terminus and the zinc finger motif near the C-terminus were chosen for the conservation analysis.

Table S3. The sequence alignment for the residues at position 77 (shown in red).

mutM	HLGM**S**GSL	ECs4510	HLGM**S**GSL
ZmutM	HLGM**S**GSL	STM3726	HLGM**S**GSL
YPO0052	HLGM**S**GSL	VC0221	HLGM**S**GSL
HI0946	HLGM**S**GSV	PM1145	HLGM**S**GSL
PA0357	HLGM**S**GSL	NMB1295	HLGM**S**GSL
NMA1505	HLGM**S**GSL	XF0071	HLGM**S**GTL

XF0170	HLGM**S**GTL	RSc0399	HLGM**T**GTL
RC1038	HLGM**S**GRF	BMEI1946	HLGM**S**GSF
BH3152	HLRM**E**GRY	slr1689	HLRM**T**GQL
mll5585	HLGM**S**GSF	BS_mutM	HLRM**E**GKY
CC3707	HLGM**T**GRF	lin1599	HLRM**E**GKF
AGc561	HLGM**S**GSF	SMc01154	HLGM**S**GSF
SPy0497	HLRM**E**GKY	alr4320	HLRM**T**GQL
L0271	HLRM**E**GKY	Cgl2021	HLGM**S**GQM
SP0970	HLRM**E**GKY	DR0493	HLGM**T**GGF
SA1512	HLGM**S**GSF	UU413	HLRM**E**GKF
MT2994	HLGM**S**G--	Rv2924c	HLGM**S**GQM
ML1658	HLGM**S**GQM	MG262.1	HLRM**E**GKY
MPN380	HLRM**E**GKY	MYPU_3100	HLRM**E**GKY

Out of the 38 selected sequences from Fpg family, there are 23 sequences have serine at position 77, 5 have threonine, and 10 have glutamic acid

References:

1. Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005) The Amber biomolecular simulation programs, *Journal of Computational Chemistry 26*, 1668-1688.
2. Fromme, J. C., and Verdine, G. L. (2003) DNA lesion recognition by the bacterial repair enzyme MutM, *J Biol Chem 278*, 51543-51548.
3. 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, *Journal of Chemical Physics 79*, 926-935.
4. 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.
5. Perlow-Poehnelt, R. A., Zharkov, D. O., Grollman, A. P., and Broyde, S. (2004) Substrate discrimination by formamidopyrimidine-DNA glycosylase: distinguishing interactions within the active site, *Biochemistry-Us 43*, 16092-16105.
6. Miller, J. H., Fan-Chiang, C. C. P., Straatsma, T. P., and Kennedy, M. A. (2003) 8-Oxoguanine enhances bending of DNA that favors binding to glycosylases, *Journal of the American Chemical Society 125*, 6331-6336.
7. Song, K., Hornak, V., Santos, C. D., Grollman, A. P., and Simmerling, C. (In press) Molecular Mechanics Parameters for the FapydG DNA lesion, *Journal of Computational Chemistry*.
8. Stote, R. H., and Karplus, M. (1995) Zinc-Binding in Proteins and Solution - a Simple but Accurate Nonbonded Representation, *Proteins-Structure Function and Genetics 23*, 12-31.
9. 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 2Nd Generation

Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules, *Journal of the American Chemical Society 117*, 5179-5197.

10. 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, *Journal of Computational Chemistry 21*, 1049-1074.
11. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006) Comparison of multiple amber force fields and development of improved protein backbone parameters, *Proteins 65*, 712-725.
12. Darden, T., York, D., and Pedersen, L. (1993) Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems, *Journal of Chemical Physics 98*, 10089-10092.
13. Cheatham, T. E., Miller, J. L., Fox, T., Darden, T. A., and Kollman, P. A. (1995) Molecular-Dynamics Simulations on Solvated Biomolecular Systems - the Particle Mesh Ewald Method Leads to Stable Trajectories of DNA, Rna, and Proteins, *Journal of the American Chemical Society 117*, 4193-4194.
14. 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, *Journal of Chemical Physics 81*, 3684-3690.
15. Kottalam, J., and Case, D. A. (1988) Dynamics of Ligand Escape from the Heme Pocket of Myoglobin, *Journal of the American Chemical Society 110*, 7690-7697.
16. 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, *Journal of Computational Chemistry 13*, 1011-1021.
17. 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, *Journal of Computational Chemistry 16*, 1339-1350.
18. Roux, B. (1995) The Calculation of the Potential of Mean Force Using Computer-Simulations, *Computer Physics Communications 91*, 275-282.
19. 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, Netherlands
20. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics, *Journal of Molecular Graphics 14*, 33-&.
21. Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., Krylov, D. M., Mazumder, R., Mekhedov, S. L., Nikolskaya, A. N., Rao, B. S., Smirnov, S., Sverdlov, A. V., Vasudevan, S., Wolf, Y. I., Yin, J. J., and Natale, D. A. (2003) The COG database: an updated version includes eukaryotes, *Bmc Bioinformatics 4*, -.
22. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J. H., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Research 25*, 3389-3402.