Supporting Information

Determinants of Photolyase's DNA Repair Mechanism in Mesophiles and Extremophiles

Benjamin J. G. Rousseau,[†] Shoresh Shafei,[†] Agostino Migliore,*[,†] Robert J. Stanley,[‡] David N. Beratan*[,†,§,ǁ]

[†] Department of Chemistry and [ǁ] Department of Physics, Duke University, Durham, North Carolina, 27708, United States

[§] Department of Biochemistry, Duke University, Durham, North Carolina, 27710, United States

[‡] Department of Chemistry, Temple University, Philadelphia, Pennsylvania, 19122, United States

Contents

1. Structures of the photolyase-CPD complexes

The antenna cofactor of the S. tokodaii photolyase (used PDB: 2E0I) is another FADH, for which we already had force field parameterization.[1, 2] However, there were no available parameterizations for the antennae of the *E. coli* (PDB: 1DNP) and *A. nidulans* (PDB: 1QNF) photolyases, which are the 5,10-methenyl-6,7,8-trihydrofolic acid and 8-hydroxy-10-(d-ribo-2,3,4,5-tetrahydroxypentyl)-5-deazaisoalloxazine, respectively. Furthermore, the antenna cofactor of *T. themophilus* was not resolved in the 1IQR crystal structure. In addition, to this lack of information in the previous literature, the absence of the antenna cofactors is expected not to affect the molecular dynamics of the proteins, as the role of these surface-located cofactors is not to assist protein stability, but rather to initiate photoexcitation. Thus, only the antenna of the *S. tokodaii* species was included in our MD study. The photolyase-CPD complexes that we created and used to run the MD simulations are provided as supporting materials, with names: "A_nidulans(1QNF)-CPD_structure", "E_coli(1DNP)-CPD_structure", "T_thermophilus(1IQR)-CPD_structure" and "S_tokodaii(2E0I)-CPD_structure". The

provided coordinates of the protein-CPD complexes were extracted at the end of their respective energy minimizations.

2. MD parameters

The simulation temperatures for the four systems are specified in table 1 of the article. AMBER force fields were employed. The scaling factor for 1-4 electrostatic interactions was set to 0.8333333. The van der Waals interactions were truncated at the cutoff distance of 12 Å, while atomic pairs until 15 Å far apart were included in the list of non-bonded atoms to be considered for periodical interaction energy calculation. Waters and bonds of H atoms to the heavy atoms were made rigid. Full electrostatic energy evaluations were made every 2 steps. The grid spacing used in calculating effective electrostatic potentials with the Particle Mesh Ewald method was 1 Å. The pressure was calculated using hydrogen-group (i.e., the H atom and the bonded heavy atom) based pseudo-molecular virial and kinetic energy.

50000 energy minimization steps were run for all systems. The unit cell vectors were (in Å): (**102**, **82**, **96**) for 1DNP-CPD, (**85**, **84**, **104**) for 1IQR-CPD, (**99**, **96**, **87**) for 1QNF-CPD, and (**78**, **80**, **104**) for 2E0I-CPD.

The minimization was followed by solvent equilibration (fixed protein-CPD complex). Temperature rescaling and the Langevin piston method for pressure control (at the value of 1 atm) were used at this stage. The barostat period was 100 fs, and its characteristic damping time was 50 fs.

After the solvent equilibration, restrained dynamics was carried out. This series of simulations gradually equilibrated the protein-CPD atoms to the solvent bath (which was equilibrated to the desired temperature in the previous simulation). This was done by gradually lifting the restraints imposed on the protein and lesion atoms, so as to avoid undesired large perturbations in the system structure and solvation. As the simulation sequence progressed, the energetic barriers restraining the atomic motion were gradually lowered. In the final restrained simulation, the protein and CPD atoms were almost unrestrained in their movement. Temperature rescaling was again used in the sequence of restrained MD simulations. The exponent for the harmonic constraint energy function was set to 2. The time step was 1 fs and all runs lasted 0.25 ns. In the first of these simulations, the harmonic constraint energy functions were multiplied by a factor of 99. In the next runs, this scaling factor was gradually reduced, using the sequence of values: 50, 10, 5, 1, 0.5 and 0.1

Afterwards, the dynamics was unconstrained. An MD run of 5 ns served as a transition from the partially restrained simulations to the fully released production MD. Langevin thermostat and barostat were used at this stage and in the production run. The damping coefficient for the Langevin dynamics was 5 ps^{-1}. The time step was 1 fs in both simulations. The MD production run lasted 30 ns.

3. MD analysis

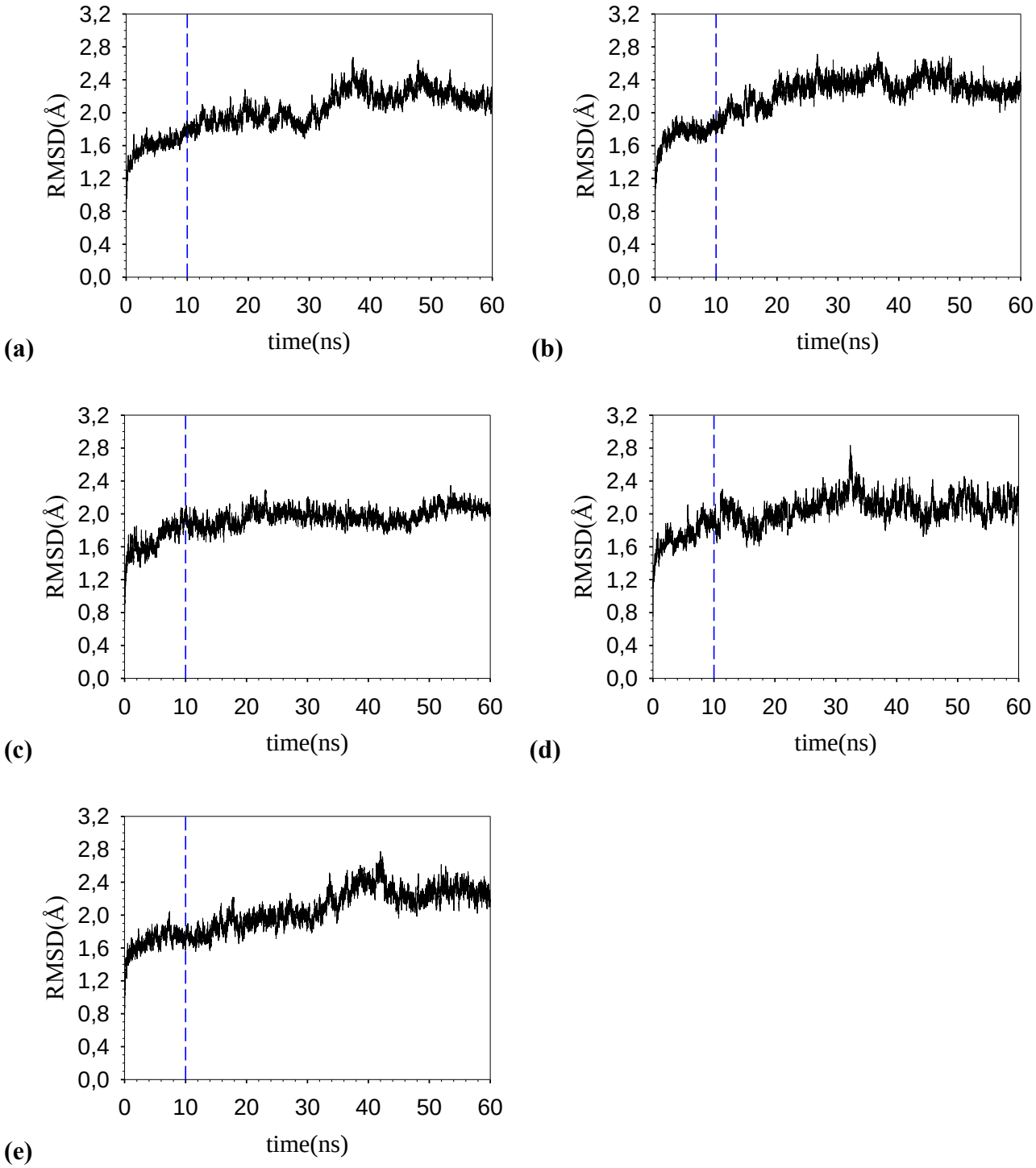

Figure S1. RMSD of [protein-FADH⁻]-CPD, with respect to the initial snapshot, of the 60-ns MD run. The vertical dashed line marks the beginning of the time range used for the tunneling pathway analysis. **(a)** EC at 293 K; **(b)** EC at 310 K; **(c)** AN at 293 K; **(d)** TT at 333 K; and **(e)** ST at 353 K. The size of the RMSD fluctuations during the MD production run is typical of equilibrated protein systems.

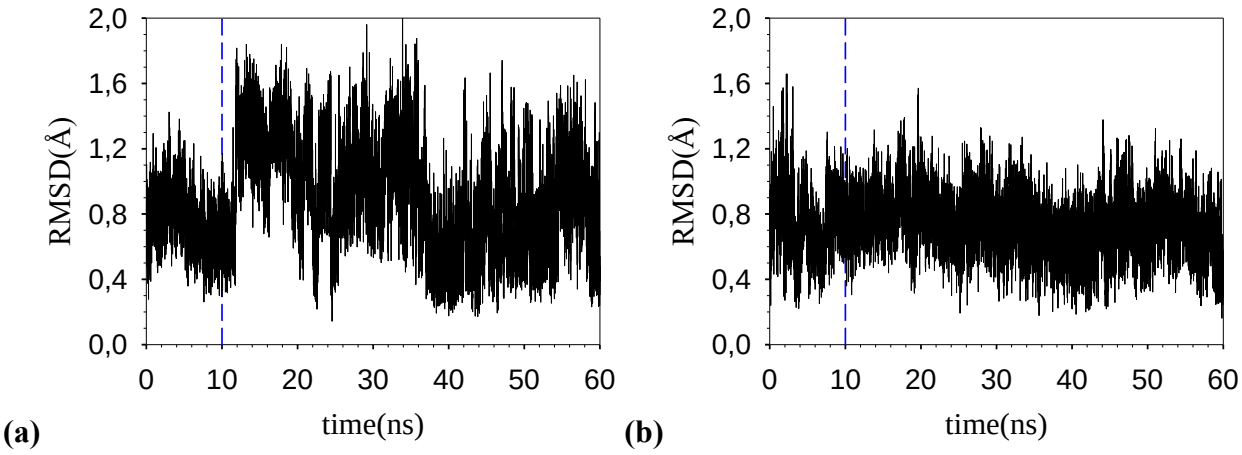

Figure S2. RMSD of adenine (with flavine-CPD frame alignment), with respect to the initial snapshot. **(a)** EC at 310 K; **(b)** TT at 333 K. The relatively large structural fluctuations of adenine in the *E. coli* complex include the flipping of adenine out of an optimal position in the gap between flavin and CPD. For the *T. thermophilus* photolyase-CPD complex, the RMSD indicates a stable position of the adenine between the flavin and CPD. Therefore, in the MD snapshots where water molecules fit between the electron donor and acceptor, and support the strongest ET pathway, adenine is expected to stay nearby in the gap, thus being able to support an ET pathway of similar strength that may be taken into account by a multi-tunneling pathway analysis.

4. Post-MD analysis

Flavin geometry.

The geometry of the anionic flavin moiety was optimized using the B3LYP[5] exchange-correlation functional and the cc-pVTZ basis set, and was also checked using the Grimme's DFT-D3 correction[6] for dispersion interactions (figure S2 and coordinates below). The observed bending agrees with previous literature.[7]

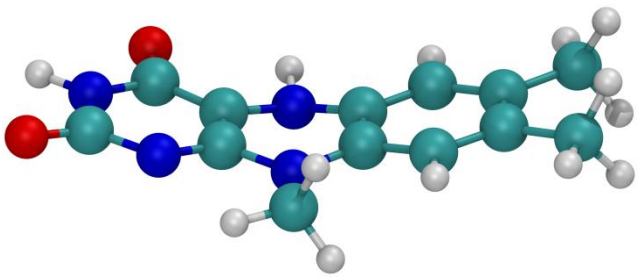

Figure S3. Bended optimal geometry of flavin.

Flavin coordinates in xyz format:

H	24.90370386	35.93970356	35.36571025
N	28.36564978	33.66353146	42.73450817
C	29.50396745	34.04561345	43.38953804
O	29.91724896	33.53731848	44.43856523
N	30.25578949	35.09159832	42.80525886
C	29.94403046	35.80321627	41.64350073
O	30.68324072	36.70774862	41.20006077
C	28.72498045	35.37514163	41.05970155
N	28.25035389	36.04501733	39.89786378
C	27.35991215	35.38078582	39.07754213
C	27.12344480	35.76550030	37.75528859
C	26.16450557	35.13550885	36.94684001
C	25.93933651	35.60775692	35.52901937
C	25.42980441	34.06975630	37.47619253
C	24.38478451	33.35577988	36.65083275
C	25.67929897	33.66239909	38.80217153
C	26.61886053	34.29128013	39.62055666
N	26.83999980	33.90177877	40.95743446
C	28.02565182	34.31958307	41.62084765
C	26.07222370	32.80732880	41.51125821
H	26.12758121	34.81094720	34.79483306
H	26.59755358	36.44739563	35.28006719
H	31.10766512	35.35118275	43.28463685
H	27.70272734	36.59510575	37.35216934
H	25.11146274	32.82506381	39.19326455
H	26.40464644	32.66281130	42.53830839
H	26.23082645	31.86799453	40.95580157
H	24.99929662	33.04289555	41.48388879
H	23.91740395	32.54447073	37.22000867
H	24.80724391	32.91288936	35.73650017
H	23.58048846	34.03108528	36.32223467
H	28.96431534	36.62281405	39.46859047

Charge distribution in flavin LUMO, LUMO +1, and CPD SOMO used in the pathway analysis.

Table S1 reports the atomic decomposition of the electron charge distributions in the two excited states of the flavin donor (approximately described by the LUMO and LUMO + 1) and in the final diabatic state (i.e., the SOMO of the CPD acceptor). The charges were calculated using DFT at the M06-2x/cc-pVTZ level of computational accuracy.

Table S1. Initial and final electron charge localizations on the *D* and *A* atoms, respectively.

p	$(\mu_p^D)_{\text{LUMO}}$	$(\mu_p^D)_{\text{LUMO + 1}}$	q	$(\mu_q^A)_{\text{SOMO}}$
C7	0.216	0.014	C4	0.256
C9A	0.211	0.030	O4	0.122
C6	0.084	0.195	C4'	0.111
C10	0.084	0.016	C2	0.101
C9	0.070	0.222	C5	0.074
C8	0.063	0.162	O4'	0.064
C5A	0.052	0.222	O2	0.059
C4A	0.029	0.005	C5'	0.053
C2	0.024	0.005	C2'	0.033
N1	0.023	0.001	N3	0.021
H15	0.018	0.000	C6	0.021
H14	0.016	0.007	O2'	0.019
C4	0.014	0.001	H10	0.012
N3	0.014	0.000	N1	0.011
N10	0.013	0.012	N3'	0.009
O2	0.012	0.003	H25	0.008
C7M	0.011	0.005	HA	0.008
C8M	0.011	0.002	N1'	0.006
H19	0.009	0.003	C5M	0.005
H18	0.007	0.019	C5N	-0.003
O4	0.006	0.000	H13	0.002
N5	0.004	0.040	C6'	0.002
H30	0.002	0.006	H11	-0.002
H31	0.002	0.007	HB	0.001
C1*	0.001	-0.003	H9	0.001
H20	0.001	0.003	H21	0.001
H12	0.001	0.000	H23	0.001
H13	0.001	0.007	H12	-0.001
H21	0.000	0.012	H24	0.001
H16	0.000	0.001	H22	0.000
H17	0.000	0.002		
C2*	0.000	0.000		

Tunneling pathway analysis and water effect on the adenine role.

The trajectory generated from the MD production run on each system was trimmed as described

in the article. After that, the pathways plugin[3] for the VMD software[4] was used to analyze all the ET pathways connecting all donor and acceptor atoms in each MD snapshot, and thus to find and visualize the best ET pathways between the atoms in the donor and acceptor. In the analysis, it was necessary to treat the hydrogen atoms explicitly to prevent the plug-in from crashing.

Table S2 reports a large fluctuation in the water and adenine contributions to bridge the electron superexchange between the flavin donor and the CPD acceptor. f_{water} is calculated considering the strongest ET pathways that include at least water but no adenine atoms. Passing from the first to the second time interval indicated, the increase in ET mediation by water correlates with the decrease in ET mediation by adenine, but the former is much larger than the latter. This fact implies that, in the second time range, water molecules also mediate several ET pathways that were characterized by direct tunneling between flavine and CPD during the first time range.

Table S2. f values in percentage form (in the indicated time windows during the MD production run) for adenine and water in the *T. thermophiles* photolyase-CPD complex.

time range	S1 excited state		S2 excited state	
	$f_{adenine}$	f_{water}	$f_{adenine}$	f_{water}
25.0–27.5 ns	29%	9%	27%	10%
27.5–30.0 ns	17%	55%	15%	60%

Figure S4 shows a MD snapshot of the FADH-CPD complex where a water molecule mediates the strongest electron tunneling pathway.

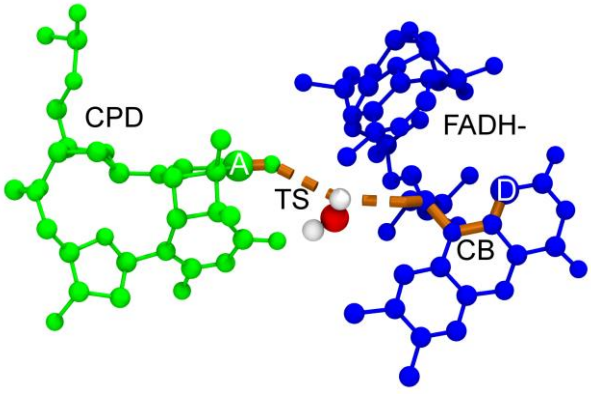

Figure S4. Strongest electron tunneling pathway (in orange) between an acceptor atom (A) in CPD (green) and a donor atom (D) in FADH^{-} (blue) in a MD snapshot of the *T. thermophiles* photolyase-CPD complex. A water molecule is part of this pathway.

Tunneling pathway analysis without water-mediated tunneling.

Table S3. *f* values, in percentage form, for *A. nidulans* (AN), *E. coli* (EC), T. *thermophilus* (TT), and *S. tokodaii* (ST) photolyases at the listed temperatures (in K), with the flavin initially in the S1 or S3 state. The pathway analysis used to produce the data in this table covers 25 ns of MD production run, and counts the ET pathways for direct flavin-to-CPD tunneling and for adenine-mediated superexchange, while the tunneling pathways through water are not considered.

species	AN	EC	EC	TT	ST
temperature	293	293	310	333	353
f (S1)	32%	29%	3.0%	39%	44%
f (S2)	30%	24%	1.9%	38%	41%

1. Antony, J.; Medvedev, D. M.; Stuchebrukhov, A. A., *J. Am. Chem. Soc.* **2000,** *122*, 1057.
2. Sato, R.; Kitoh-Nishioka, H.; Ando, K.; Yamato, T., *Chem. Phys. Lett.* **2015,** *633*, 247.
3. Balabin, I. A.; Hu, X. Q.; Beratan, D. N., *J. Comput. Chem.* **2012,** *33*, 906.
4. Humphrey, W.; Dalke, A.; Schulten, K., *J. Mol. Graph.* **1996,** *14*, 33.
5. Becke, A. D., *J. Chem. Phys.* **1993,** *98*, 5648.
6. Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H., *J. Chem. Phys.* **2010,** *132*, 154104.
7. Walsh, J. D.; Miller, A. F., *Theochem-J. Mol. Struct.* **2003,** *623*, 185.