Photosynthetic Diode: Electron transport rectification by wetting the quinone cofactor. Electronic Supplementary Information

Daniel R. Martin[1, a)] and Dmitry V. Matyushov[1, b)]
Department of Physics and Department of Chemistry & Biochemistry, Arizona State University, PO Box 871504, Tempe, Arizona 85287

I. SIMULATION PROTOCOL

A. Initial System Setup

The structure of the wild type reaction center (RC) of *Rhodobacter sphaeroides* bacterium was taken from the protein data bank (PDB: 1PCR). Three subunits (H, L, and M), four bacteriochlorophylls (BChl), two bacteriopheophytins (BPh), two quinones (Q_A and Q_B), and the non-heme iron were extracted from the structure file. The protein was parameterized using the CHARMM27 force field with CMAP torsional corrections. The tension-free CHARMM36 force field was used to parameterize the POPC lipid bilayer, while the TIP3P water and counterions were parameterized according to the CHARMM27 force field. All corresponding cofactor (BChl's, BPh's, quinoes, and non-heme iron) parameters were taken from previous work[1,2] in a manner completely consistent with the CHARMM potential energy function.

The simulation cell was created by first orienting the symmetry axis of the protein and corresponding cofactors along the z-axis of the coordinate frame. The complex was then solvated with a thin layer of TIP3P water[3] ($\sim$ 3Å). Following this initial solvation step, the water molecules surrounding the hydrophobic region of the protein were removed and replaced by a POPC bilipid membrane using VMD's membrane plugin.[4] The resulting structure was then completely solvated with TIP3P water and neutralizing sodium ions. The system size was then adjusted by removing outer water and lipid molecules to keep a rectangular shape within the largest to smallest length ratio of 1.0 to 1.5 and less than 1.2×10^5 atoms in the simulation box (size restriction of the Anton supercomputer[5] at the time of the simulations). The resulting system contained 94421 atoms.

The system initialization was done using NAMD 2.7.[4] The system was initialized by first removing any bad contacts introduced by the system setup procedure. This was done by running a very short ($\sim$ 1000 steps) of steepest decent minimization. Once the bad contacts were removed, a short 1 ns NPT equilibration run was performed with the entire RC complex, water, ions, and lipid head groups fixed, allowing the lipid tails to properly relax. Following this lipid initialization step, the protein and corresponding cofactors were harmonically constrained and water was prevented from moving into the lipid using NAMD's tcl forces plugin to force the entering waters out of the hydrophobic region. This part of the system setup was performed for 10 ns. Finally, the system was equilibrated under the NPT protocol for approximately 25 ns with all atoms free. The NPT equilibration simulation was done using the Langevin dynamics in NAMD with a damping coefficient of 1 ps^{-1}, piston period of 100 fs, the piston decay time of 50 fs, the piston target pressure of 1.01325 bar, and constant temperature control set to 300 K. Long-range electrostatic interactions were treated with the particle mesh Ewald technique using a cutoff distance of 12.0 Å. A 2.0 fs timestep was used for all simulations.

B. Conversion from NAMD to Anton and analysis

From the final equilibrated configuration produced by NAMD, the atomic velocities and positions were taken as the initial configuration of the system to be simulated on the Anton supercomputer.[5] All parameters were converted from the CHARMM type force field used in the system initialization procedure to the Desmond type format used by Anton. The conversion of the force field parameters from the CHARMM format to Anton/Desmond format was done using either conversion scripts provided by the Anton and Pittsburgh Super Computer Center (PSC) for the case of standard amino acids, water, ions, and lipid and developed in-house for the corresponding cofactors. From the initial configuration, all patches connecting the non-heme iron to the protein matrix were initially removed so that `viparr` (the Desmond program that applies the force field parameters to the structure) would apply the standard parameter set automatically to the entire system.

Once the parameters were confirmed (including all angles, dihedrals, improper dihedrals, bonded, and nonbonded parameters) and checked using in-house developed python scripts, the final set of parameters connecting the non-heme iron to the protein matrix were added. The addition of the parameters required to connect the non-heme iron to the protein matrix were added manually to the Maestro CMS file before producing the final Desmond Molecular System (DMS) file used in the Anton simulation. As a final confirmation, SQL statements were used to confirm that all parameters (including charges, exclusions, dihedrals, improper dihedrals, angles, bonded, and nonbonded parameters) were applied

[a)] Electronic mail: daniel.martin@asu.edu
[b)] Electronic mail: dmitrym@asu.edu

properly within the SQLite DMS file.

C. Simulation and Analysis Protocol

The simulation protocol used on Anton was adapted from the sample `ark` file provided by PSC using the `multigrator` integration algorithm developed by the Desmond group. This integrator allows the user to carry out thermostat and barostat updates less frequently then the outer reversible reference system propagator algorithm (RESPA[6]), thus increasing the overall performance of the simulation. The Martyna-Tobias-Klein (MTK[7]) dynamical scheme recommended by the PSC was used to perform all production simulations on Anton. All production simulations were done at 300 K. The long-ranged electrostatics were calculated using the k-space Gaussian split Ewald (k-GSE) methodology with 64 mesh points along each direction in k-space and a cutoff distance of ∼12 Å. The total simulation time for the forward reaction in the Q_A^-/Q_B state was ∼9 μs and for the state Q_A/Q_B^- was ∼2 μs. All reported simulation were done with a 2.0 fs timestep and a saving frequency of 5–10 ps.

Simulations with Q_A and Q_B harmonically constrained were performed using NAMD 2.8 with a force constant of 1 kcal/mol/Å^2. The constrained positions of the quinones were determined by the averages of the distances between the quinones and the non-heme iron along the two Anton generated trajectories as well as the proximal position of Q_B determined from the PDB entry 1AIG. The Steered Molecular Dynamics (SMD) technique was used to move Q_B into the proximal position by aligning the protein from an extracted Anton frame to that in the 1AIG structure and perform the corresponding rotations/translations to the entire simulation system.

All analysis programs were developed in-house either in Fortran or with VMD's tcl interface. The initial raw Anton trajectories were first converted to the CHARMM dcd format and then re-imaged, making certain that for all configurations the entire RC complex (including all cofactors) was not split between multiple images.

The resulting average distances between the quinones as well as that bewteen each quinone and non-heme iron along each trajectory were calculated. These values are tabulated in Table S1 as well as the simulation lengths, charge states, and number of waters in a 3Å shell surrounding each quinone.

II. ENERGY GAP AND REORGANIZATION ENERGY

A. Coulomb and induction interactions

The Coulomb contributions to the energy gap were calculated using the NAMD pair-interaction plugin with the partial charges, q_i of the active site replaced by the difference charges, $\Delta q_i = q_i^{(2)} - q_i^{(1)}$, where $q_i^{(1)}$ are the charges with electron on the donor site and $q_i^{(2)}$ are the charges with the electron on the acceptor site. The Coulomb contribution to the energy gap from a given part of the thermal bath is

$$X_n^{\mathrm{C}} = \sum_i \Delta q_i \phi_{ni}\,, \tag{S1}$$

where ϕ_{ni} is the total electrostatic potential at site i corresponding to $n = \{a, i, l, p, w\}$ where a="all" atoms, i="ions", l="lipid" atoms, p="protein" atom, and w="water" atoms. The contribution to the energy gap from the induction interaction of the donor-acceptor complex with the thermal bath is given by

$$X^{\mathrm{ind}} = E_{\mathrm{ind}}^{(2)} - E_{\mathrm{ind}}^{(1)} \tag{S2}$$

with

$$E_{\mathrm{ind}}^{(\ell)} = -\frac{1}{2}\sum_j \alpha_j \left[\mathbf{E}^{(\ell)}(\mathbf{r}_j)\right]^2. \tag{S3}$$

Here, α_j are atomic polarizabilities and $\mathbf{E}(\mathbf{r}_j)$ is the electric field contributions from the entire donor-acceptor complex in state ℓ at the sites corresponding to a set of j atoms, where j can be separated into components $\{a, i, l, p, w\}$ corresponding to all, ions, lipid, protein and water, respectively. Explicitly, the electric field in Eq. (S3) is given by

$$\mathbf{E}^{(\ell)}(\mathbf{r}_j) = \sum_{i\in\{D,A\}} \frac{q_i^{(\ell)}(\mathbf{r}_j - \mathbf{r}_i)}{|\mathbf{r}_j - \mathbf{r}_i|^3}\,. \tag{S4}$$

The reorganization energy is then calculated from the fluctuations of the energy gap along the simulation trajectory using the relation

$$\lambda = (\beta/2)\langle(\delta X)^2\rangle, \tag{S5}$$

where $X = X^{\mathrm{C}} + X^{\mathrm{ind}}$ and $\beta = 1/(k_B T)$. The energy gap and reorganization energy data for each simulation is tabulated in Tables S2–S7.

B. Dynamics of the energy gap

The dynamics of the energy gap was calculated from the Stokes shift autocorrelation function

$$C_X(t) = \langle \delta X(t)\delta X(0)\rangle. \tag{S6}$$

The normalized function was represented by four decaying exponents

$$C_X(t)/C_X(0) = \sum_{i=m}^{4} A_m e^{-t/\tau_m}\,, \tag{S7}$$

resulting in seven parameters given the constraint $\sum_m A_m = 1$.

TABLE S1. Simulation lengths, charges of quinone cofactors, average number of water molecules in the 3 Å shell, and distances between center of masses of the cofactors. Harmonically Constrained (HC) simulations are with both ubiquinones constrained and in one case with water molecules constrained.

Config.	sim. length (μs)	Q_A	Q_B	$\langle N_A \rangle$	$\langle N_B \rangle$	Q_A-Fe (Å)	Q_B-Fe (Å)	Q_A-Q_B (Å)
Anton (1-2)	8.98	-1	0	3.6	0.3	8.0	14.9	21.5
Anton (2-1)	2.17	0	-1	3.0	6.3	9.1	12.5	20.3
NAMD - HC(Q_AQ_B) prox	0.1	-1	0	3.2	4.8	7.5	10.3	17.7
NAMD - HC(Q_AQ_B) prox	0.1	0	-1	1.7	8.5	7.8	9.9	17.5
NAMD - HC(Q_AQ_B)	0.1	0	-1	2.0	4.9	7.7	14.7	21.1
NAMD - HC(Q_AQ_B & water)	0.05	0	-1	1.8	0.0	8.0	14.6	21.2

TABLE S2. Contribution to ΔE_{1-2} and the reorganization energy, λ. This table corresponds to the first row in Table S1. (From ANTON)

ΔE_{1-2} (eV)	all	protein	water	lipid	ions
Coulomb	2.14	1.02	0.63	0.48	0.01
Induction	0.33	0.60	0.31	-0.60	0.00
Total	2.45	1.62	0.94	-0.12	0.01
λ_{1-2} (eV)	all	protein	water	lipid	ions
Coulomb	3.78	4.01	5.12	1.27	0.20
Induction	0.61	0.54	0.12	0.26	0.00
Total	5.06	5.16	5.80	1.56	0.20

TABLE S3. Contribution to ΔE_{2-1} and the reorganization energy, λ. This table corresponds to the second row in Table S1. (From ANTON)

ΔE_{2-1} (eV)	all	protein	water	lipid	ions
Coulomb	-3.53	-1.18	-2.71	0.41	-0.04
Induction	-0.02	0.92	-0.47	-0.40	0.00
Total	-3.55	-0.26	-3.18	0.13	-0.04
λ_{2-1} (eV)	all	protein	water	lipid	ions
Coulomb	8.37	3.37	9.76	1.03	0.67
Induction	0.54	0.97	0.50	0.17	0.00
Total	10.23	4.97	13.17	1.45	0.67

TABLE S4. Results from the simulation of Q_A^-/Q_B with Q_B constrained in the proximal position. This table corresponds to the third row in Table S1. (From a 100 ns NAMD trajectory with constrained Q_A/Q_B)

ΔE_{1-2} (eV)	all	protein	water	lipid	ions
Coulomb	1.55	1.24	0.27	0.00	0.04
Induction	-0.10	0.08	0.10	-0.27	0.00
Total	1.45	1.32	0.37	-0.27	0.04
λ_{1-2} (eV)	all	protein	water	lipid	ions
Coulomb	3.71	4.40	4.53	0.77	0.14
Induction	0.22	0.17	0.07	0.04	0.00
Total	4.45	5.12	4.84	0.81	0.14

TABLE S5. Results from the simulation of Q_A/Q_B^- with Q_B constrained in the proximal position. This table corresponds to the fourth row in Table S1. (From a 100 ns NAMD trajectory with constrained Q_A/Q_B)

ΔE_{2-1} (eV)	all	protein	water	lipid	ions
Coulomb	-5.57	-0.91	-4.39	-0.26	0.00
Induction	-0.70	0.00	-0.54	-0.16	0.00
Total	-6.27	-0.91	-4.93	-0.42	0.00
λ_{2-1} (eV)	all	protein	water	lipid	ions
Coulomb	4.44	2.20	4.66	0.74	0.11
Induction	0.24	0.14	0.18	0.02	0.00
Total	5.44	2.50	5.74	0.72	0.12

TABLE S6. Results from the simulation of Q_A/Q_B^- with Q_A/Q_B constrained in the Q_A^-/Q_B configuration. This table corresponds to the fifth row in Table S1. (From a 100 ns NAMD trajectory with constrained Q_A/Q_B)

ΔE_{2-1} (eV)	all	protein	water	lipid	ions
Coulomb	-4.50	-1.04	-3.47	0.01	0.01
Induction	-0.37	0.51	-0.28	-0.59	0.00
Total	-4.87	-0.54	-3.75	-0.58	0.01
λ_{2-1} (eV)	all	protein	water	lipid	ions
Coulomb	4.86	3.42	7.53	0.93	0.08
Induction	0.29	0.20	0.32	0.07	0.00
Total	5.88	4.41	9.64	0.99	0.08

TABLE S7. Results from the simulation of Q_A/Q_B^- with Q_A/Q_B constrained in the Q_A^-/Q_B configuration and with water constrained. This table corresponds to the sixth row in Table S1. (From a 50 ns NAMD trajectory with constrained Q_A/Q_B and water)

ΔE_{2-1} (eV)	all	protein	water	lipid	ions
Coulomb	-0.99	-0.69	-0.77	0.37	0.10
Induction	0.12	0.58	0.19	-0.64	0.00
Total	-0.87	-0.11	-0.58	-0.27	0.10
λ_{2-1} (eV)	all	protein	water	lipid	ions
Coulomb	1.54	1.22	0.49	0.08	0.00
Induction	0.13	0.10	0.00	0.04	0.00
Total	1.73	1.41	0.50	0.11	0.00

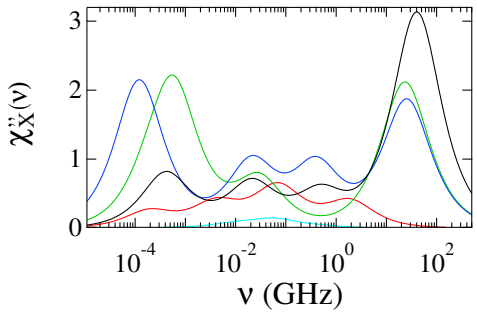

FIG. S1. Loss function of the total energy gap (black). Also shown are loss components from the protein (green), water (blue), lipids (red), and ions (cyan). The individual component contributions are calculated according to Eq. (S10).

The loss function was calculated using the Fourier-Laplace transform of the fitted correlation function according to the relation

$$\chi''_X(\omega) = 2\lambda \sum_{m=1}^{4} A_m \frac{\omega\tau_m}{1+(\omega\tau_m)^2}\,. \tag{S8}$$

The individual contributions to the loss function can be calculated using the above formalism. For instance, the loss function corresponding to the protein contribution to the energy gap arises from the sum of the Coulomb and induction components,

$$X_{\rm p} = X_p^{\rm C} + X_p^{\rm ind}\,. \tag{S9}$$

The correlation function is then fitted to four exponents and the Fourier-Laplace transform is taken to produce the loss function according to Eq. (S8).

Figure (S1) shows the individual loss functions referring to the total (black), ions (cyan), lipid (red), protein (green), and water (blue). Each function is obtained from a separate fit of the time correlation function and is calculated from the following equation where n stands for an individual component of the thermal bath

$$\chi''_n(\omega) = 2\lambda_n \sum_{m=1}^{4} A_m^n \frac{\omega\tau_m^n}{1+(\omega\tau_m^n)^2}\,. \tag{S10}$$

The fitting parameters are listed in Table S8.

The exponential fit of the energy gap correlation function leads to the non-ergodic reorganization energy given by the following equation[8]

$$\lambda(k) = \lambda \sum_{m=1}^{4} (2A_m/\pi)\mathrm{arccot}(k\tau_j), \tag{S11}$$

where A_m are the fitting coefficients in Eq. (S7). This equation was used to produce Figure 4 in the main text.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation (MCB-1157788 and CHE-1464810) and through XSEDE resources (TG-MCB080116N). Anton computer time was provided by the Pittsburgh Supercomputing Center (PST) and the BTRC for Multiscale Modeling of Biological Systems through Grant P41GM103712-S1 from the National Institutes of Health. The Anton machine at NRBSC/PSC was generously made available by D. E. Shaw Research.

REFERENCES

[1] D. N. LeBard and D. V. Matyushov, J. Phys. Chem. B **113**, 12424 (2009).
[2] D. N. LeBard, D. R. Martin, S. Lin, N. W. Woodbury, and D. V. Matyushov, Chem. Sci. **4**, 4127 (2013).
[3] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).
[4] J. C. Phillips *et al.*, J. Comp. Chem. **26**, 1781 (2005).
[5] J. L. Klepeis, K. Lindorff-Larsen, R. O. Dror, and D. E. Shaw, Curr. Opin. Struct. Biol. **19**, 120 (2009).
[6] M. E. Tuckerman, B. J. Berne, and A. Rossi, J. Chem. Phys. **97**, 1990 (1992).
[7] G. J. Martyna, D. J. Tobias, and M. L. Klein, J. Chem. Phys. **101**, 4177 (1994).
[8] D. V. Matyushov, Acc. Chem. Res. **40**, 294 (2007).

TABLE S8. Fit parameters for the auto-correlation functions and the corresponding loss functions.

	A_1	A_2	A_3	A_4	τ_1 (ns)	τ_2 (ns)	τ_3 (ns)	τ_4 (ns)
total	0.62	0.10	0.13	0.16	0.004	0.33	7.60	388
protein	0.41	0.14	0.38	0.07	0.007	5.38	273.	839
water	0.32	0.15	0.16	0.37	0.006	0.39	7.69	1319
lipid	0.24	0.38	0.22	0.16	0.084	2.29	39.5	846
ions	0.10	0.61	0.27	0.02	0.318	2.75	17.2	372