SI Appendix

Mapping Hole Hopping Escape Routes in Proteins

Ruijie D. Teo,[†,§] Ruobing Wang,[†,‖,§] Elizabeth R. Smithwick,[†] Agostino Migliore,[†*] Michael J. Therien,[†] David N. Beratan[†,‡,¶,*]

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

[‖] Present address: Science Center of Opera Solutions OPCO, LLC, San Diego, California 92130, United States

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

[¶] Department of Physics, Duke University, Durham, North Carolina 27708, United States

[§] These authors contributed equally to this work.

Contents

S1. Redox-active amino acids in cytochrome P450 and Ccp1

S2. Reorganization energies and redox potentials

S3. Electronic coupling calculation

S4. Description of EHPath

S5. Hole-transfer steps in cytochrome P450

S6. Most effective hole hopping pathways in Ccp1

www.pnas.org/cgi/doi/10.1073/pnas.1906394116

S1. Redox-active amino acids in cytochrome P450 and Ccp1

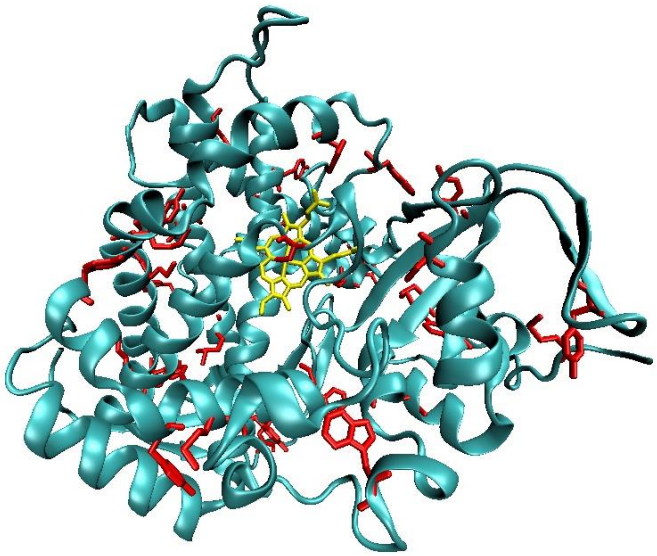

Fig. S1. Heme domain of flavocytochrome P450 (PDB code: 2IJ2 (1)). The redox-active amino acid residues (i.e., Try, Trp, Met, and Cys (2, 3)) are highlighted in red and using the licorice representation. The heme (yellow) is also drawn in licorice style. The figure was generated using VMD (4).

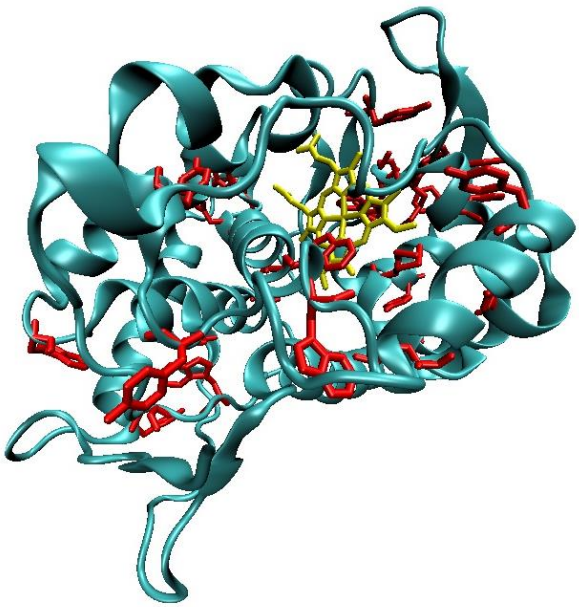

Fig. S2. Ccp1 (cytochrome c peroxidase) from *Saccharomyces cerevisiae* (PDB code 1ZBY (5)). The heme group is highlighted in yellow and the redox active amino acids are shown in red. The figure was generated using VMD (4).

S2. Reorganization energies and redox potentials

Table S1 presents values of the reorganization energy λ_s' (namely, half the reorganization energy λ_{SS}' associated with the self-exchange reaction in the *S*-*S* system; see Eq. 3) and of the oxidation potential $E°$ for each indicated redox-active group *S* and specified donor-acceptor (*D*-*A*) center-to-center distance R_{DA}'. The reaction free energy $\Delta G°$ in Eq. 1 is obtained approximately as the difference between the $E°$ values of *A* and *D*. The reorganization energies at the relevant the center-to-center distances R_{DA} are obtained using Eq. 2 and 3 (6).

EHPath can also be used to identify charge-transfer routes and to quantify the electron or hole travel time in protein-DNA complexes. In these cases (as in our tests on primase-nucleic acid) the self-exchange reorganization energy for the *S*-*S* nucleobase dimer (*S* = *G*, *A*) is calculated using the expression

$$\lambda_{SS} = \lambda_{SS}^{i} + \left(\frac{1}{\varepsilon_o} - \frac{1}{\varepsilon_s}\right)(\Delta q)^2 \left(\frac{1}{R_X} - \frac{1}{R_{DA}}\right) \qquad \text{(S1)}$$

where λ_{SS}^{i} is the inner-sphere reorganization energy (Table S1) and R_X is the effective radius of the nucleobase (6). R_{DA} is calculated as described in the main text.

Table S1. Reorganization energies and redox potentials for the indicated redox-active species *S*. Each nucleobase (G or A) is associated with an effective radius R_X (6), and λ_S^i is half the reorganization energy λ_{SS}^i.

S	λ_s' (eV)	λ_S^i (eV)	R_{DA}' (Å)	R_X (Å)	$E°$ (V)
Tyr	1.02[7]	-	12.8[7]	-	0.93[8]
Trp	0.95[7]	-	12.8[7]	-	1.02[8]
Met	1.08[7]	-	12.8[7]	-	1.66[9]
Cys	1.27[7]	-	12.8[7]	-	0.92[10]
Heme	0.85[6]	-	17.45[6]	-	~1[2]
[4Fe4S]	0.75[6]	-	14.44[6]	-	0.08[6, 11]
Gly	~1[a]	-	12.8[a]	-	1.9[12]
G	-	0.373[13]	-	2.10[6]	1.29[14, 15]
A	-	0.2115[13]	-	2.22[6]	1.42[14]

[a] Since λ_s' is not available in the literature, we use the value 1 eV, which is within the range of λ_s' values for the other amino acids listed in Table S1 (0.95-1.27 eV), while R_{DA}' is set to 12.8 eV.

S3. Electronic coupling calculations

The electronic couplings are determined using one of the semiempirical methods described below. Hopfield's equation is used for the couplings between redox-active amino acid residues (16):

$$V_{DA} = \frac{2.7}{\sqrt{N_D N_A}} e^{-0.72 R_{ee}} \qquad \text{(S2)}$$

where N_D and N_A are the numbers of atoms identifying *D* or *A*, respectively, and R_{ee} is the edge-to-edge distance between *D* and *A* (16). The electronic coupling is calculated between the initial and final diabatic electronic states that describe the hole localized on *D* and *A*, respectively. Eq. S1 allows us to calculate approximate V_{IF} values using the heterocyclic groups (Tyr, Trp) or sulfur p-orbitals (Met, Cys) of the redox-active groups involved. For Trp, Tyr, Met, and Cys we used *N* values of 9, 7, 3, and 2 respectively (6) (in particular, the beta carbon and sulfur atom are retained in the cysteine residue (N_{cys} = 2)). Only one atom is used to represent Gly.

The electronic couplings between a heme cofactor and an amino acid residue acceptors ($V_{\text{heme-A}}$) are calculated using Eq. S18 of ref. (6), with the iron-sulfur cluster replaced by the heme group, namely, $V_{heme-A} = \sqrt{V_{heme-heme} V_{A-A}}$. $V_{heme\text{-}heme}$ is estimated using the equation (17)

$$V_{heme-heme} = V_0 e^{-\frac{\beta(R_{cc}-16.5)}{2}} \qquad \text{(S3)}$$

where V_0 = 8.0, β = 2.6 Å^{-1}, and R_{cc} is the heme center-to-center distance. These V_0 and β values correspond to parallel heme dimers with coplanar imidazole ligands (17). While the systems studied do not contain heme dimers, one should take into account the heme orientation with respect

to the nearest redox-active group in a given pathway. Even within a protein, such orientations depend on the pathway under consideration. We made the simple choice of using the above values of V_0 and β for all orientations, since these values are approximately in the middle of the ranges of values obtained for the different *D*-*A* orientations in ref. (17). To implement Eq. S3, we use two heme groups at the same edge-to-edge distance R_{ee} as the actual heme-A redox couple and add twice the effective radius of the heme (~3.63 Å) (6). $V_{A\text{-}A}$ is calculated using Eq. S1.

The electronic coupling between a (high-potential) [4Fe4S] cluster and a protein residue is calculated similarly as $V_{[4\mathrm{Fe4S}]-A} = \sqrt{V_{[\mathrm{Fe4S}]-[\mathrm{Fe4S}]}V_{A-A}}$. $V_{[\mathrm{Fe4S}]-[\mathrm{Fe4S}]}$ is estimated using the equation (18)

$$\log V_{[4Fe4S]-[4Fe4S]} = 1.73 - 0.42R_{cc} \qquad \text{(S4)}$$

where, similarly to the above, R_{cc} is obtained as the sum of R_{ee} (between [4Fe4S] and A) and twice the effective radius of heme (~1.87 Å) (6).

S4. Description of EHPath

The EHPath program uses python libraries (pandas and networkx) for data processing and directed graph analysis. The program first requires the user to provide the input CSV (MS-Dot format) files *donor.csv*, *bridge.csv*, and *acceptor.csv* (sample files can be seen on GitHub). Therefore, one needs to define the charge (electron or hole) donor, bridge(s), and acceptor(s). The charge donor is easily identified as a protein active site or cofactor. The *bridge.csv* file is created using the grep command to extract the redox-active Cys, Met, Tyr and Trp residues from the original PDB file:

```
$ grep -E 'CYS|MET|TRP|TYR' input_PDB.pdb > bridge.pdb
```

The output file will contain all redox-active residues in the protein, but only those in the bridging protein medium (between *D* and *A*) can contribute to the relevant pathways. The

bridge.pdb file is then converted to the CSV format using tools like Excel or Convertio. Note that the redox-active amino acid residues that are part of the protein active site (or bound to the cofactor) working as charge donor or acceptor are neither the start point of a charge hopping pathway nor part of the bridge. For example, if the donor is a [4Fe4S] cluster, Cys residues bonded to the Fe atoms of the cluster should be removed from both donor and bridge lists. For hole hopping pathways as oxidative damage escape routes in biological systems, the potential acceptor residues are solvent-exposed surface residues. For users with no prior information about these surface residues pertaining to their proteins of interest, tools such as PyMOL, which calculate the solvent accessible surface area, help to identify the relevant terminal residues at the protein surface.

Once the CSV files are generated, one can run the EHPath.py program. The user is asked to enter the names of the donor, bridge, and acceptor CSV files. EHPath.py and the CSV files need to be in the same directory. The program asks for the residue number associated with the donor, a cutoff number (where cutoff number + 1 is the maximum number of nodes in any pathway considered by the program), the number of pathways to be printed in the results section, the types of donor and acceptor (type just 'electron' for electron transfer and 'hole' for hole transfer), and the value of a parameter α ($0 \leq \alpha \leq 1$). The cutoff number helps one limit the pathways under consideration to the most direct ones between the terminal charge donor and acceptor. We chose a cutoff number of 4 as a good compromise between computational cost and probability that an efficient pathway contains more than 5 nodes. In fact, our results show that the top-ranked hopping pathway contains less than 5 nodes for each of the systems investigated. The α parameter allows the user to change the reorganization energy λ for each charge-transfer step in a hopping route between the value λ_{DA} calculated using Eq. 2-3 and the value of 0.8 eV used in ref. 3. α is defined as follows:

$$\lambda = 0.8 + \alpha(\lambda_{DA} - 0.8) \quad \text{(in eV)} \qquad \text{(S5)}$$

Therefore, α = 1 gives $\lambda = \lambda_{DA}$, while α = 0 gives $\lambda = 0.8$ eV.

The program ranks the hopping pathways according to both the approximate and exact mean residence times (based on our kinetic model (6)). An option in the program allows the user to define a hopping pathway, by specifying the residue numbers involved, and to calculate the exact mean residence time for such pathway. Here is an example of a successful run that finds and analyzes the hole hopping pathways between a heme donor with residue number 999 and an amino acid residue acceptor (the maximum number of nodes in any computed pathway is 5, the number of pathways printed in the output is 5, and α = 1).

```
****************************************
Welcome! EHPath is able to analyze and rank electron/hole hopping
pathways between an electron/hole donor and an electron/hole acceptor
according to the overall mean residence time.
****************************************

Please enter the name of donor file (eg: donor.csv):
Donor_Node_2IJ2.csv
Please enter the name of bridge file, (eg: bridge.csv):
Bridging_Nodes_2IJ2.csv
Please enter the name of acceptor file, (eg: acceptor.csv):
Acceptor_Nodes_2IJ2.csv
Please enter residue number of donor: 999
Please enter cutoff_num (cutoff_num + 1 = maximum number of nodes in
the pathway, Warning: A larger cutoff_num will cost more memory and
leads to longer computation time): 4
Please specify the number of pathways to be printed: 5
Please enter the type of donor/acceptor (electron or hole): hole
Please specify the value for α: 1

Start calculation........

Top 5 pathways ranked according to the approximate mean residence
time:
[999, 96, 90, 334] 0.011699829537750969
[999, 115, 305] 0.05850237678612136
[999, 156, 115, 305] 0.08403832975022704
[999, 96, 334] 0.13273387470162293
```

```
[999, 115, 156, 305] 14.437926968672322

Top 5 pathways ranked according to the exact mean residence time:
[999, 96, 90, 334] 0.0367530800137136
[999, 115, 305] 0.060844981316308985
[999, 156, 115, 305] 0.13910192575395705
[999, 96, 334] 0.42140415862323116
[999, 156, 305] 15.103972122225027

End of calculation

Would you like to find the exact mean residence time of a specific
pathway? (Yes/No): Yes

Please specify the pathway with their residue numbers (for example,
999, 156, 115, 112, 305): 999, 156, 115, 112, 305

Exact mean residence time:  3956.3190955084924

****************************************
If you have any questions or suggestions about EHPath, please contact
Mr. Ruijie Teo (rt131@duke.edu) or Dr. Ruobing Wang
(ruobing.wang@operasolutions.com).
Citation: Teo, R. D.; Wang, R.; Smithwick, E.; Migliore, A.; Therien,
M. J.; Beratan, D. N. Proc. Natl. Acad. Sci. U.S.A., 2019.
****************************************

Copyright (C) 2019  Teo, R. D.; Wang, R.; Smithwick, E.; Migliore, A.;
Therien, M. J.; Beratan, D. N.

This program is free software: you can redistribute it and/or modify
it under the terms of the GNU General Public License as published by
the Free Software Foundation, either version 3 of the License, or (at
your option) any later version.

This program is distributed in the hope that it will be useful, but
WITHOUT ANY WARRANTY; without even the implied warranty of
MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU
General Public License for more details.

You should have received a copy of the GNU General Public License
along with this program. If not, see
https://www.gnu.org/licenses/gpl.html.

Press Enter to exit EHPath.
```

Steps to define a new charge donor in EHPath (for example, FAD)

Step 1. In the donor.csv file, identify the atoms that are relevant to the charge hopping (that is, the atoms that would be involved in the excess charge distribution, which can be predetermined by means of first-principles computations or based on simple chemical grounds). We consider here the case of FAD, which consists of 53 heavy atoms (atom IDs 5811-5863 in the PDB file with code 5JFC (19)). From the literature (e.g., see ref. (20) and refs. therein), we know that the frontier molecular orbitals (which describe the excess charge distribution in a single-particle picture) are essentially localized on the flavin moiety, which thus defines the donor group. The atom names in the PDB file 5JFC (19) are as follows:

…

```
HETATM 5833  N1  FAD L 503     -65.192  51.052  14.932  1.00 11.62           N
HETATM 5834  C2  FAD L 503     -65.736  49.843  14.641  1.00 14.53           C
HETATM 5835  O2  FAD L 503     -66.679  49.424  15.348  1.00 12.41           O
HETATM 5836  N3  FAD L 503     -65.304  49.076  13.624  1.00 12.20           N
HETATM 5837  C4  FAD L 503     -64.301  49.452  12.822  1.00 12.11           C
HETATM 5838  O4  FAD L 503     -63.921  48.718  11.884  1.00 14.33           O
HETATM 5839  C4X FAD L 503     -63.650  50.762  13.071  1.00 10.38           C
HETATM 5840  N5  FAD L 503     -62.631  51.215  12.308  1.00 12.16           N
HETATM 5841  C5X FAD L 503     -62.057  52.414  12.560  1.00 14.30           C
HETATM 5842  C6  FAD L 503     -61.011  52.859  11.761  1.00 13.12           C
HETATM 5843  C7  FAD L 503     -60.416  54.090  12.014  1.00 11.87           C
HETATM 5844  C7M FAD L 503     -59.282  54.570  11.147  1.00 15.21           C
HETATM 5845  C8  FAD L 503     -60.898  54.934  13.140  1.00 11.74           C
HETATM 5846  C8M FAD L 503     -60.252  56.267  13.412  1.00 14.02           C
HETATM 5847  C9  FAD L 503     -61.944  54.493  13.941  1.00 11.78           C
HETATM 5848  C9A FAD L 503     -62.539  53.262  13.689  1.00 12.61           C
HETATM 5849  N10 FAD L 503     -63.600  52.807  14.496  1.00 10.19           N
HETATM 5850  C10 FAD L 503     -64.173  51.556  14.205  1.00 11.83           C
```

…

Therefore, the following line needs to be added under the ‘Section for truncating redox groups’ of EHPath.py:

donor = donor[~(donor['atomtype'].isin(['N1','C2','O2','N3','C4','O4','C4X','N5','C5X','C6','C7','C7M','C8','C8M','C9','C9A','N10','C10']) & (donor['residuetype'].isin(['FAD'])))]

Step 2. Define the new residue type, starting with the line

if self.Residuetype=='FAD':

and inserting the pertinent reorganization energy and redox potential. The reorganization energy can be obtained from a variety of methods: from simple approaches that use Marcus' expression for the reorganization energy (21) and exploit available theoretical/experimental data in the literature (6) to accurate methods that use classical molecular dynamics (22, 23) or QM/MM (24).

Step 3. Insert the line to calculate the electronic coupling under the sections 'Donor-Bridge' and 'Donor-Acceptor".

S5. Hole-transfer steps in cytochrome P450

Table S2. Hole-transfer **s**teps in Pathways 1 and 2 through cytochrome P450, classified according to the values of the pertinent forward hole-transfer rate constants k_{DA} (see Eq. 1 in the article). The corresponding k'_{SBT} values (Eq. 4) calculated using the free energy parameters of ref. (3) are also reported.

Hole donor-acceptor	k_{DA} (s^{-1})	k_{SBT} (s^{-1})
HEM-W96	5.0×10^{3}	4.5×10^{6}
W96-W90	8.7×10^{1}	1.2×10^{7}
W90-Y334	3.9×10^{5}	5.2×10^{9}
HEM-C156	2.1×10^{1}	-
C156-Y115	8.7×10^{4}	-
Y115-M112	2.1×10^{-2}	-
M112-Y305	1.4×10^{9}	-

S6. Most effective hole hopping pathways in Ccp1

Table S3. The 20 fastest hole hopping pathways in Ccp1, with terminal hole acceptors identified in ref. (25) (PDB 1ZBY), are ranked based on their computed τ_M values (Eq. 5). The corresponding $\tau_{M,approx}$ values (Eq. 6) are also reported.

Hole Hopping Pathways	τ_M (s)	$\tau_{M,approx}$ (s)
HEM-W191-Y229	2.5×10^{-3}	7.9×10^{-4}
HEM-W191-W211	2.5×10^{-3}	7.9×10^{-4}
HEM-Y36	1.2×10^{-2}	1.2×10^{-2}
HEM-Y187-W191-Y229	6.1×10^{-2}	3.1×10^{-2}
HEM-Y187-W191-W211	6.1×10^{-2}	3.1×10^{-2}
HEM-W191-Y236-W223	9.4×10^{-2}	3.0×10^{-2}
HEM-W191-Y236-W211	9.4×10^{-2}	3.0×10^{-2}
HEM-Y51-W191-Y229	1.0×10^{-1}	3.2×10^{-2}
HEM-Y51-W191-W211	1.0×10^{-1}	3.2×10^{-4}
HEM-Y187-W223	1.2×10^{-1}	1.1×10^{-1}
HEM-W191-Y187-W223	1.2×10^{-1}	1.1×10^{-1}
HEM-Y51-W57	2.2×10^{-1}	6.9×10^{-2}
HEM-Y51-W191-Y236-W223	2.2×10^{-1}	6.1×10^{-2}
HEM-Y51-W191-Y236-W211	2.2×10^{-1}	6.1×10^{-2}
HEM-Y51-W191-Y187-W223	2.3×10^{-1}	1.4×10^{-1}
HEM-W191-W223	2.6×10^{-1}	8.3×10^{-2}
HEM-W191-Y51-W57	3.9×10^{-1}	1.0×10^{-1}
HEM-Y51-W191-W223	4.5×10^{-1}	1.1×10^{-1}
HEM-Y153-Y187-W191-Y229	6.1×10^{-1}	5.0×10^{-1}
HEM-Y153-Y187-W191-W211	6.1×10^{-1}	5.0×10^{-1}

References

1. Girvan HM*, et al.* (2007) Structural and spectroscopic characterization of p450 bm3 mutants with unprecedented p450 heme iron ligand sets. New heme ligation states influence conformational equilibria in p450 bm3. *J. Biol. Chem.* 282(1):564-572.

2. Winkler JR & Gray HB (2015) Electron flow through biological molecules: does hole hopping protect proteins from oxidative damage? *Q. Rev. Biophys.* 48(4):411-420.

3. Gray HB & Winkler JR (2015) Hole hopping through tyrosine/tryptophan chains protects proteins from oxidative damage. *Proc. Natl. Acad. Sci. U. S. A.* 112(35):10920-10925.

4. Humphrey W, Dalke A, & Schulten K (1996) VMD: visual molecular dynamics. *J. Mol. Graph.* 14(1):33-38.

5. Bonagura CA*, et al.* (2003) High-resolution crystal structures and spectroscopy of native and compound I cytochrome *c* peroxidase. *Biochemistry* 42(19):5600-5608.

6. Teo RD*, et al.* (2019) Charge transfer between [4Fe4S] proteins and DNA is unidirectional. Implications for biomolecular signaling. *Chem* 5:122-137.

7. Heck A*, et al.* (2012) Charge transfer in model peptides: obtaining Marcus parameters from molecular simulation. *J. Phys. Chem. B* 116(7):2284-2293.

8. DeFelippis MR*, et al.* (1991) Electrochemical properties of tyrosine phenoxy and tryptophan indolyl radicals in peptides and amino acid analogs. *J. Chem. Phys.* 95(8):3416-3419.

9. Merényi G, Lind J, & Engman L (1996) The dimethylhydroxysulfuranyl radical. *J. Phys. Chem.* 100(21):8875-8881.

10. Surdhar P & Armstrong DA (1986) Redox potentials of some sulfur-containing radicals. *J. Phys. Chem.* 90(22):5915-5917.

11. Funk MA, Judd ET, Marsh ENG, Elliott SJ, & Drennan CL (2014) Structures of benzylsuccinate synthase elucidate roles of accessory subunits in glycyl radical enzyme activation and activity. *Proc. Natl. Acad. Sci. U.S.A* 111(28):10161-10166.

12. Armstrong DA, Rauk A, & Yu D (1995) Solution thermochemistry of the radicals of glycine. *J. Chem. Soc., Perkin Trans. 2* (3):553-560.

13. Khan A (2013) Reorganization, activation and ionization energies for hole transfer reactions through inosine–cytosine, 2-aminopurine – thymine, adenine–thymine, and guanine–cytosine base pairs: a computational study. *Comput. Theor. Chem.* 1013:136-139.

14. Steenken S & Jovanovic SV (1997) How easily oxidizable is DNA? One-electron reduction potentials of adenosine and guanosine radicals in aqueous solution. *J. Am. Chem. Soc.* 119(3):617-618.

15. Steenken S, Jovanovic SV, Bietti M, & Bernhard K (2000) The trap depth (in DNA) of 8-oxo-7,8-dihydro-2‘deoxyguanosine as derived from electron-transfer equilibria in aqueous solution. *J. Am. Chem. Soc.* 122(10):2373-2374.

16. Hopfield JJ (1974) Electron transfer between biological molecules by thermally activated tunneling. *Proc. Natl. Acad. Sci. U.S.A* 71(9):3640-3644.

17. Smith DMA, Rosso KM, Dupuis M, Valiev M, & Straatsma TP (2006) Electronic coupling between heme electron-transfer centers and its decay with distance depends strongly on relative orientation. *J. Phys. Chem. B* 110(31):15582-15588.

18. Voityuk AA (2010) Electron transfer between [4Fe–4S] clusters. *Chem. Phys. Lett.* 495(1):131-134.

19. Lubner CE*, et al.* (2017) Mechanistic insights into energy conservation by flavin-based electron bifurcation. *Nat. Chem. Biol.* 13:655.

20. Rousseau BJG, Shafei S, Migliore A, Stanley RJ, & Beratan DN (2018) Determinants of photolyase's DNA repair mechanism in mesophiles and extremophiles. *J. Am. Chem. Soc.* 140(8):2853-2861.

21. Marcus RA & Sutin N (1985) Electron transfers in chemistry and biology. *Biochim. Biophys. Acta* 811(3):265-322.

22. Blumberger J & Klein ML (2006) Reorganization free energies for long-range electron transfer in a porphyrin-binding four-helix bundle protein. *J. Am. Chem. Soc.* 128(42):13854-13867.

23. Tipmanee V, Oberhofer H, Park M, Kim KS, & Blumberger J (2010) Prediction of reorganization free energies for biological electron transfer: a comparative study of Ru-modified cytochromes and a 4-helix bundle protein. *J. Am. Chem. Soc.* 132(47):17032-17040.

24. Hu LH*, et al.* (2011) Reorganization energy for internal electron transfer in multicopper oxidases. *J. Phys. Chem. B* 115(45):13111-13126.

25. Kathiresan M & English AM (2017) LC-MS/MS suggests that hole hopping in cytochrome c peroxidase protects its heme from oxidative modification by excess H_2O_2. *Chem. Sci.* 8(2):1152-1162.