Supporting Information for "eMap: A Web Application for Identifying and Visualizing Electron or Hole Hopping Pathways in Proteins"

Ruslan N. Tazhigulov, James R. Gayvert, Melissa Wei, and Ksenia B. Bravaya* Department of Chemistry, Boston University, Boston, MA 02215, United States

Third-Party Applications, Components and License Information

- Biopython [1, 2] (Biopython permissive license and BSD-3-Clause)
- Bootstrap (MIT)
- DSSP [3, 4] (Boost Software License 1.0)
- Flask (BSD-3-Clause)
- Flask Table (BSD-3-Clause)
- jQuery (MIT)
- MSMS [5] (MSLIB License)
- NetworkX [6] (BSD-3-Clause)
- NGL Viewer [7, 8] (MIT)
- NumPy [9] (BSD-3-Clause)
- PyGraphviz [10] (BSD-3-Clause)
- RDKit (BSD-3-Clause)
- SciPy [11] (BSD-3-Clause)
- Select2 (MIT)
- ViewerJS (MIT)

^{*}Electronic address: kbravaya@gmail.com

Dependence on Penalty Function Parameters

eMap allows one to modify the default penalty function ($\epsilon = \alpha \exp \left[-\beta (R - R_{\text{offset}})\right]$) parameters under the "Advanced" set of options. We tested three different cases, keeping the "Graph Parameters" (Edges per Vertex, Distance Cutoff, SD Cutoff) at their default values:

- (A) $\alpha = 1.0, \beta = 2.3 \text{ Å}^{-1}, R_{\text{offset}} = 0.0 \text{ Å}, \text{distance-dependent model, eMap default}$
- (B) $\alpha = 0.6, \beta = 1.7 \text{ Å}^{-1}, R_{\text{offset}} = 1.4 \text{ Å}$, through-space penalty in the *Pathways* Model
- (C) $\alpha = 0.6, \beta = 0.5 \text{ Å}^{-1}, R_{\text{offset}} = 1.4 \text{ Å}$, with the softer decay parameter β

For Cryptochrome (PDB: 1U3D), we specified the source (FAD510-2, lumiflavin part of FAD) and the target (W324), and compared eMap reports corresponding to the different cases with each other (Fig. S1).

Α	#	Source	1	2	3	4	5	Score
	1a	FAD510-2	W400	W377	W324			24.15
	1b	FAD510-2	W385	Y53	W377	W324		35.25
	1c	FAD510-2	W400	W334	W379	W324		36.37
	1d	FAD510-2	W400	W377	W492	W324		49.20
	1e	FAD510-2	W385	Y53	Y309	W377	W324	50.67
В	#	Source	1	2	3	4	5	Score
	1a	FAD510-2	W400	W377	W324			15.42
	1b	FAD510-2	W385	Y53	W377	W324		22.81
	1c	FAD510-2	W400	W334	W379	W324		23.64
	1d	FAD510-2	W400	W377	W492	W324		33.12
	1e	FAD510-2	W385	Y53	Y309	W377	W324	33.39
С	#	Source	1	2	3	4	5	Score
	1a	FAD510-2	W400	W377	W324			5.00
	1b	FAD510-2	W385	Y53	W377	W324		7.33
	1c	FAD510-2	W400	W334	W379	W324		7.58
	1d	FAD510-2	W400	W377	W492	W324		10.37
	1e	FAD510-2	W385	Y53	Y309	W377	W324	10.60

FIG. S1: eMap reports for Cryptochrome (PDB: 1U3D), corresponding to the different sets of the penalty function parameters.

One may notice that despite the apparent dependence of the "Score" on the penalty function parameters, the results do not change qualitatively: the top 5 shortest pathways, and even their order, do not depend on the chosen penalty function parameters. While this is the case for the cryptochrome, we believe that changing parameters of the penalty function might potentially alter the predicted pathways in general and can be found useful by advanced users.

Pathways Model Analysis for Cryptochrome (PDB: 1U3D)

The *Pathways* model developed by Beratan and Onuchic [12, 13] has been designed to predict the electron transfer (ET) pathways for the ET reactions dominated by tunneling. To illustrate robustness and simplicity of eMap, we compared performance of eMap and the *Pathways* VMD plugin [14] (which become formally identical with the proper choice of flags for the *Pathways* model) by considering cryptochrome protein (1U3D PDB ID) as an example. The *Pathways* model with the default parameters does not yield the experimentally observed hopping pathway through tryptophan triad as the most efficient channel. However, if the bridge atoms are carefully and properly selected enforcing through-space tunneling only (effectively turning off through-covalent bond and through-hydrogen bond tunneling) the model yields the correct hopping pathways. Turning off through-covalent bond and through-hydrogen bond tunneling can be achieved by using the flag "-b" in the *Pathways* plugin [14] for VMD [15] and specifying atoms belonging to the side chains of the aromatic amino acids only. For the 1U3D example:

pathways -d "resname FAD and name N1 C2 O2 N3 C4 O4 C4X N5 C5X C6 C7 C7M C8 C8M C9 C9A N10 C10" -a "resid 324 and name CG CD1 NE1 CE2 CD2 CE3 CZ3 CH2 CZ2 CG CD1 CE1 CZ CE2 CD2" -b "resname TRP TYR PHE HIS and name CG CD1 NE1 CE2 CD2 CE3 CZ3 CH2 CZ2 CG CD1 CE1 CZ CE2 CD2"

The capabilities and limitations of the *Pathways* model are briefly outlined on the manual page at https://people.chem.duke.edu/~ilya/Software/Pathways/docs/pathways.html. Note that the cross-platform eMap (as a web application) automatically determines potential electron transfer active (ETA) moieties, intuitively simplifying analysis for the end user, yet providing the flexibility for specifying own ETA moieties and varying the penalty function/graph parameters. At the same time, eMap, in contrast to the *Pathways* model, is limited to predicting hopping through aromatic sites or other specified ETA moieties only. Fig. S2 shows the pathway predicted with the *Pathways* model used with the default parameters, which differs from the experimentally observed FAD-W400-W377-W324 channel. eMap with the Distance Option "Closest Atom" becomes isomorphic to the *Pathways* model, if the aromatic amino acid side chains are specified as the bridge atoms, and the through-covalent bond and through-hydrogen bond tunneling are turned off in the *Pathways* model. The pathways model used with the proper

choice of the flags and parameters as discussed above (Sec. III, main text).



FIG. S2: The *Pathways* model predicts the incorrect pathway (red) between FAD and W324, involving I403 and T404 instead of W400 and W377, if no restrictions on the bridge atoms are imposed.

- [1] Hamelryck, T.; Manderick, B. *Bioinformatics*, **2003**, *19*, 2308–2310.
- [2] Cock, P. J. A.; Antao, T.; Chang, J. T.; Chapman, B. A.; Cox, C. J.; Dalke, A.; Friedberg, I.; Hamelryck, T.; Kauff, F.; Wilczynski, B. et al. *Bioinformatics*, 2009, 25, 1422–1423.
- [3] Kabsch, W.; Sander, C. *Biopolymers*, **1983**, 22, 2577–2637.
- [4] Joosten, R. P.; te Beek, T. A. H.; Krieger, E.; Hekkelman, M. L.; Hooft, R. W. W.; Schneider, R.; Sander, C.; Vriend, G. Nucleic Acids Res., 2011, 39, D411–D419.
- [5] Sanner, M. F.; Olson, A. J.; Spehner, J. C. Biopolymers, 1996, 38, 305–320.
- [6] Hagberg, A.; Swart, P.; S Chult, D. Exploring network structure, dynamics, and function using NetworkX. Technical report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008.
- [7] Rose, A.; Hildebrand, P. Nucleic Acids Res., 2015, 43, W576–W579.
- [8] Prlić, A.; Bradley, A. R.; Duarte, J. M.; Rose, P. W.; Rose, A. S.; Valasatava, Y. *Bioinformatics*, 2018, 34, 3755–3758.
- [9] van der Walt, S.; Colbert, S. C.; Varoquaux, G. Computing in Science & Engineering, 2011, 13, 22–30.
- [10] PyGraphviz License, **2013**, https://pygraphviz.github.io/documentation/pygraphviz-1.3/reference/legal.html.
- [11] Jones, E.; Oliphant, T.; Peterson, P. et al. SciPy: Open source scientific tools for Python, 2001–.
- [12] Onuchic, J. N.; Beratan, D. N.; Winkler, J. R.; Gray, H. B. Annu. Rev. Biophys. Biomol. Struct., 1992, 21, 349–377.
- [13] Beratan, D.; Onuchic, J.; Winkler, J.; Gray, H. Science, 1992, 258, 1740–1741.
- [14] Balabin, I. A.; Hu, X.; Beratan, D. N. J. Comput. Chem., 2012, 33, 906–910.
- [15] Humphrey, W.; Dalke, A.; Schulten, K. J. Mol. Graph., **1996**, 14, 33–38.