Electronic Supplementary Information for "Discovery of lead low-potential radical candidates for organic radical polymer batteries with machine-learning-assisted virtual screening"

Cheng-Han Li and Daniel P. Tabor*

Department of Chemistry, Texas A&M University, College Station, TX 77842 USA

E-mail: daniel_tabor@tamu.edu

Contents

S1 Information on Contents of Raw Data Files

The training set data and all calculated properties for the library are given in the files training_set.csv and redox_library.csv, respectively. These files, along with the GP model and the scripts used to generate the plots in the manuscript are provided at the following Github repository: https://github.com/Tabor-Research-Group/redox_mol_screening.

S2 Top-10% Lowest Reduction Potential Molecules

Fig. S1: Top-10% Lowest Reduction Potential Molecules with their GP-calibrated reduction potential (E_{red}) and Boltzmann-averaged electronic coupling (H_{ab}). Raw data provided in csv format.

S3 Further Details on Gaussian Process Regression Model Training

We took 69 molecules with their experimental reduction potential[S1–S8] as the target of the training set for our GP model. The feature variables are constituted of the semi-empirical GFN2-xTB[S9] electron affinity and reduced feature vectors after feature selection by the gradient boosted decision trees for Morgan fingerprints[S10] composed of 4096-bit vectors using a radius of 3 and AUTOCORR 3D descriptors.[S11] We tested different combinations of the threshold value of the lowest feature importance to keep and found using "0.25*median value" for the Morgan fingerprints and "median value" for the AUTOCORR 3D would include enough features while not deteriorating the performance of the GP model with higher dimensionality. GP regression is a non-parametric model that can make predictions by incorporating prior knowledge with kernel functions. We constructed the kernel function as a sum of a radial basis function kernel for GFN2-xTB, molecular fingerprints, and 3D descriptors, respectively. We obtained the GP model for calibrating GFN2-xTB electron affinity by optimizing the hyperparameters through the maximization of the log marginal likelihood with the L-BFGS algorithm.

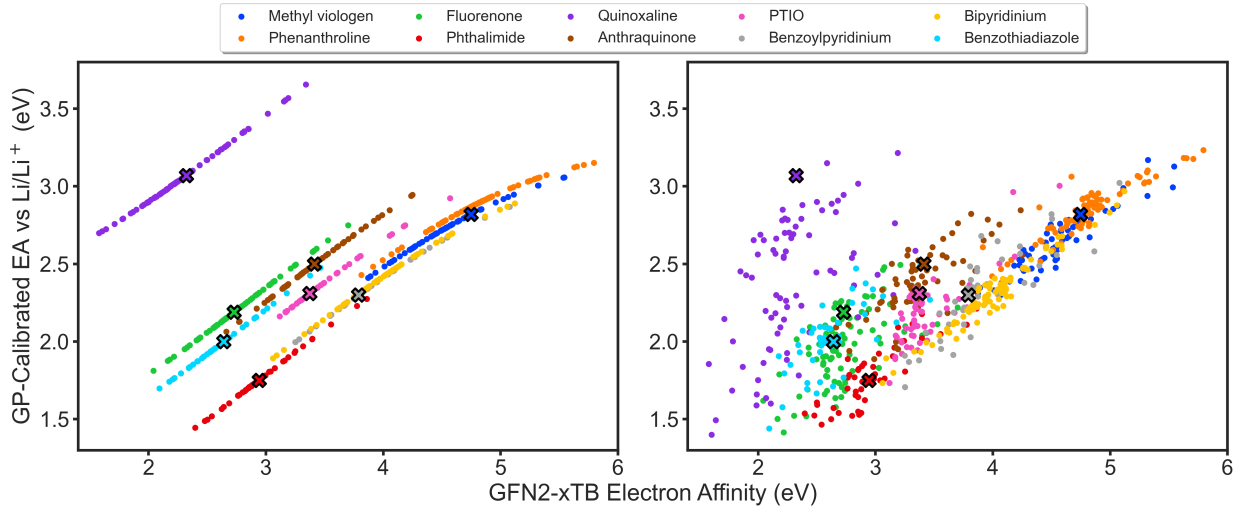

Fig. S2: GFN2-xTB electron affinity versus GP-calibrated electron affinity using core fingerprints (left) and full fingerprints (right) for GP calibrations. Each color represents a type of derivative from each core molecule. The experimental points (which the model will correctly predict if it is included in the training set) are indicated with crosses. The results on the right indicate that the model that includes all structural fingerprint information (rather than just the core fingerprint information) overemphasizes the importance of functional groups and predicts an unrealistically low reduction potential for modestly functionalized template molecules, such as quinoxalines.

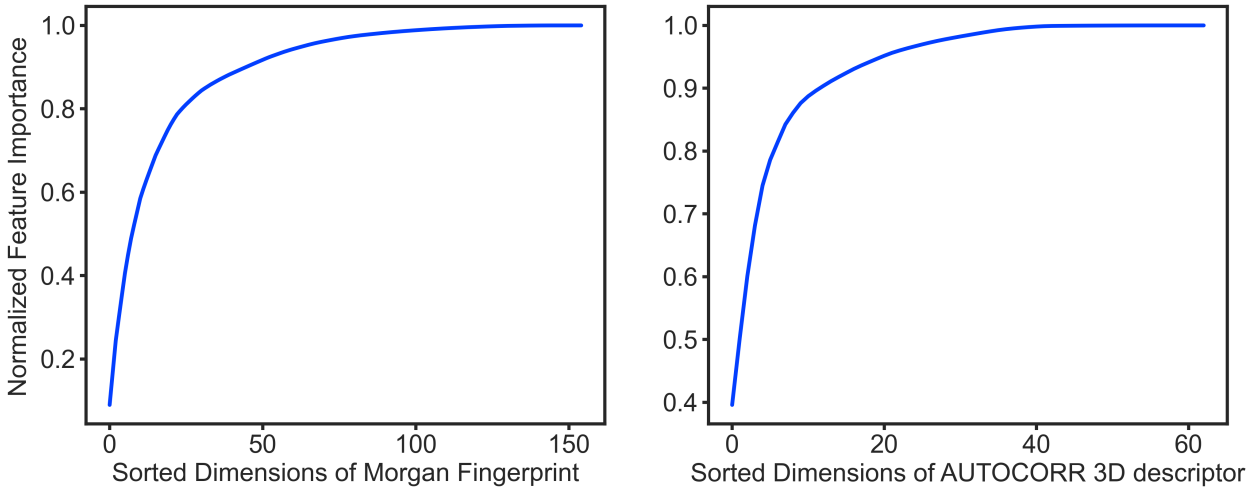

Fig. S3: Cumulative Feature Importance, which starts (on the left) from the feature dimension with the highest importance for Morgan fingerprint and AUTOCORR 3D descriptor.

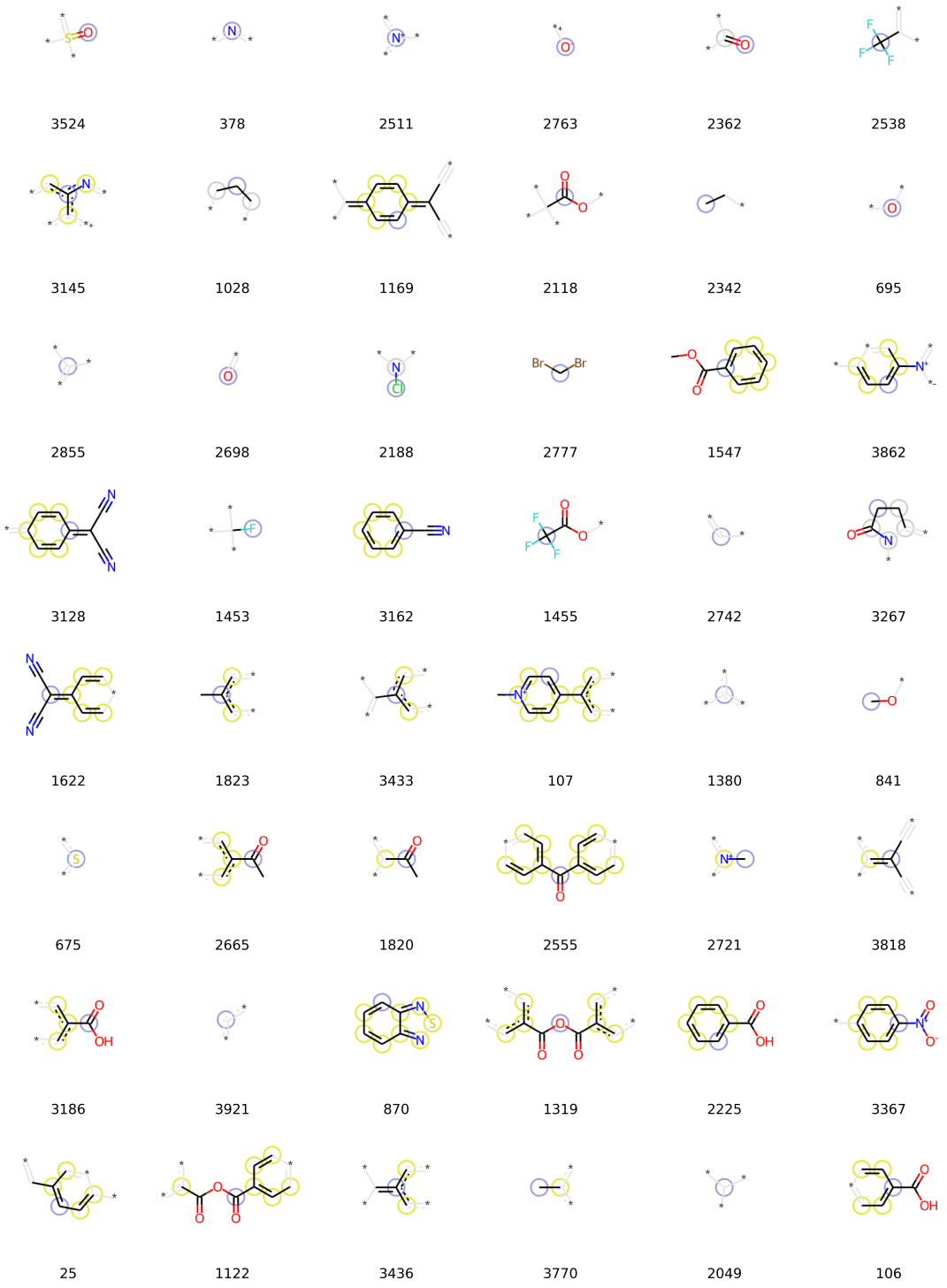

Fig. S4: Visualization of atoms and bonds in a molecules that correspond to the bits with top importance (Fig. S3) along with their indices. Highlighted colors in blue indicate the central atom in the environment, yellow highlights indicate aromatic atoms, and gray indicates aliphatic ring atoms. Star signs and bonds drawn in light gray represent wildcard atoms and bonds that are not directly part of the given bit.

S4 Electronic Coupling Sampling

S4.1 Random Dimer Structure Generation

First, we place two monomers optimized at the GFN2-xTB level of theory completed overlapped in the origin (Note that the inclusion of the implicit solvent model during optimization also avoids the situation where two charged monomers repel each other). From this initial configuration, a random rotation followed by a random translation was performed for one of the monomers. The maximum translation for a monomer is the diameter of a molecule defined by the longest distance between a pair of heavy atoms within the monomer. This distance was constrained to the dimers having a closest heavy atom distances between two monomers within 2.5 to 3.5 Å. An end-to-end example of the dimer generation procedure is also provided in the Github repository.

S4.2 Sampling Size

It is impossible to sample the continuous space of dimer configurations–taking 10 points in each dimension for all the six degrees of freedom between the monomers (under a rigid body approximation) would require a sampling of one million configurations. The following procedure was implemented for sampling the dimer configurations. During the first step of the sampling of dimer structures, each structure was sampled randomly. We make an assumption that the electronic coupling from these sampled geometries follows the normal distribution. Though the statistics may not exactly reflect this distribution, we found it too computationally expensive to perform a Boltzmann-weighted sampling. Thus, we made the random samplings for the electronic coupling of each molecule with a margin of error = 0.04 eV (taken to be approximately chemical accuracy), under the 95% confidence interval. Through statistical sampling arguments, we can approximate the sample size to reach the required margin of error under a certain confidence interval by

$$N \geq \left(\frac{z * \sigma}{MOE}\right)^2$$

where N is the sample size, z is the z-score under the desired confidence interval, σ is the population standard deviation, and MOE is the margin of error. A pilot study was conducted to to estimate

σ by calculating electronic coupling on 30 random dimer structures generated from each conformer of all top candidates. In this case, we found that the standard deviation of the coupling for each molecule is between 0.006 to 0.2 eV with a mean of 0.07 eV. We made a conservative decision by assuming the population standard deviation is 0.2 eV. Within this model, we need $N = 100$ to reach our required criteria.

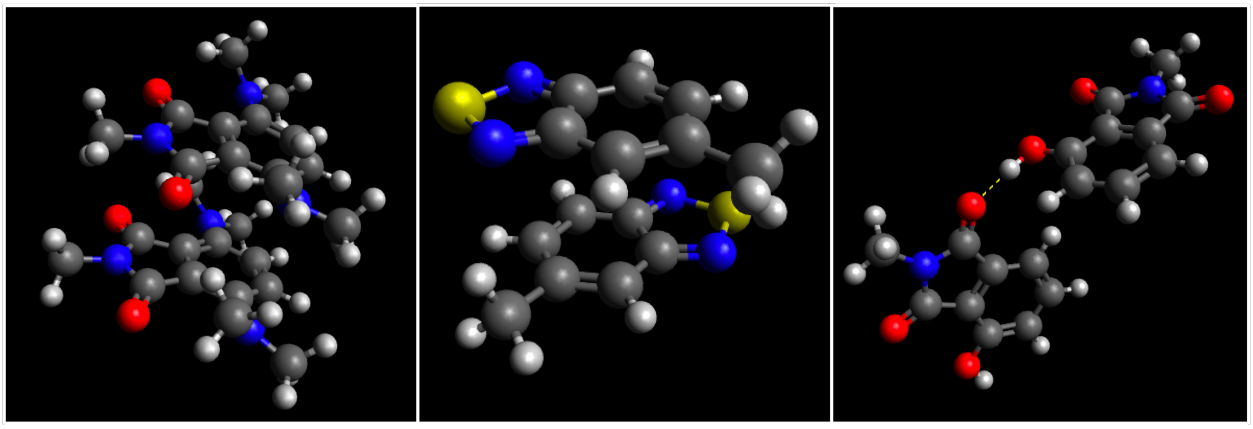

Fig. S5: Three cases of dimer structures with lower energies interaction energies than the parent unfunctionalized template molecules. Here, we highlight cases that have higher or lower Boltzmann-averaged electronic coupling than the template molecules. Overlapped phthalimide derivative dimer with higher electronic coupling (left). Staggered benzothiadiazole derivative dimer with lower electronic coupling (middle). Phthalimide derivative dimer stabilized by hydrogen bonding (yellow dashed line) with lower electronic coupling (right).

References

[S1] Cook, S. K.; Horrocks, B. R. Heterogeneous Electron-Transfer Rates for the Reduction of Viologen Derivatives at Platinum and Bismuth Electrodes in Acetonitrile. *ChemElectroChem* **2017**, *4*, 320–331.

[S2] Kwon, G.; Lee, S.; Hwang, J.; Shim, H.-S.; Lee, B.; Lee, M. H.; Ko, Y.; Jung, S.-K.; Ku, K.; Hong, J.; Kang, K. Multi-redox Molecule for High-Energy Redox Flow Batteries. *Joule* **2018**, *2*, 1771–1782.

[S3] Zhang, C.; Niu, Z.; Ding, Y.; Zhang, L.; Zhou, Y.; Guo, X.; Zhang, X.; Zhao, Y.; Yu, G.

Highly Concentrated Phthalimide-Based Anolytes for Organic Redox Flow Batteries with Enhanced Reversibility. *Chem* **2018**, *4*, 2814–2825.

[S4] Duan, W. et al. "Wine-Dark Sea" in an Organic Flow Battery: Storing Negative Charge in 2,1,3-Benzothiadiazole Radicals Leads to Improved Cyclability. **2017**, *2*, 1156–1161.

[S5] Brushett, F. R.; Vaughey, J. T.; Jansen, A. N. An All-Organic Non-aqueous Lithium-Ion Redox Flow Battery. **2012**, *2*, 1390–1396.

[S6] Ding, Y.; Li, Y.; Yu, G. Exploring Bio-inspired Quinone-Based Organic Redox Flow Batteries: A Combined Experimental and Computational Study. *Chem* **2016**, *1*, 790–801.

[S7] Duan, W. et al. A symmetric organic-based nonaqueous redox flow battery and its state of charge diagnostics by FTIR. *J. Mater. Chem. A* **2016**, *4*, 5448–5456.

[S8] Hendriks, K. H.; Sevov, C. S.; Cook, M. E.; Sanford, M. S. Multielectron Cycling of a Low-Potential Anolyte in Alkali Metal Electrolytes for Nonaqueous Redox Flow Batteries. *ACS Energy Lett.* **2017**, *2*, 2430–2435.

[S9] Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.* **2019**, *15*, 1652–1671, PMID: 30741547.

[S10] Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742–754.

[S11] Todeschini, R.; Consonni, V. *Handbook of Chemoinformatics*; John Wiley & Sons, Ltd, 2003; Chapter VIII.2, pp 1004–1033.