CHEMISTRY OF MATERIALS

pubs.acs.org/cm

Article

Quantum Chemistry-Informed Active Learning to Accelerate the Design and Discovery of Sustainable Energy Storage Materials

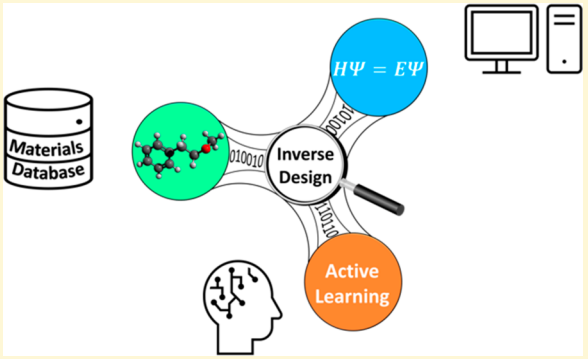
Hieu A. Doan, Garvit Agarwal, Hai Qian, Michael J. Counihan, Joaquín Rodríguez-López, Jeffrey S. Moore, and Rajeev S. Assary*

Cite This: *Chem. Mater.* 2020, 32, 6338–6346

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: We employed density functional theory (DFT) to compute oxidation potentials of 1400 homobenzylic ether molecules to search for the ideal sustainable redoxmer design. The generated data were used to construct an active learning model based on Bayesian optimization (BO) that targets candidates with desired oxidation potentials utilizing only a minimal number of DFT calculations. The active learning model demonstrated not only significant efficiency improvement over the random selection approach but also robust capability in identifying desired candidates in an untested set of 112 000 homobenzylic ether molecules. Our findings highlight the efficacy of quantum chemistry-informed active learning to accelerate the discovery of materials with desired properties from a vast chemical space.



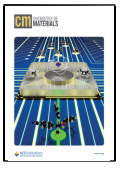
INTRODUCTION

Redox flow batteries (RFBs) are a promising technology for stationary energy storage applications.[1−4] Among the technical challenges associated with nonaqueous redox flow batteries (NRFBs) is their inability to achieve both a long calendar life and a long cycle life. Mechanisms of calendar aging mainly involve the formation of passivating films on positive and negative electrodes.[5,6] These films interfere with charge transfer processes at the electrolyte/electrode interfaces.[7] One likely cause of film formation is the irreversible adsorption of damaged redoxmers. To regenerate electrode interfaces and retain long-lifetime NRFBs, we seek to develop redoxmers that are capable of programmed destruction, thereby providing a means to remove undesirable film deposits. We envisioned a design concept in which redox-active cores are connected to a molecular scaffold via a cleavable tether. By using a redox-triggered cleavage reaction, the recreation of a pristine interface becomes possible simply by altering the electrode's potential to a value that is outside the battery's normal window of operation.

Mesolytic cleavage reactions are thermodynamically favorable bond dissociation processes that fragment a molecule into separated radical and ion species following a single electron oxidation or reduction event.[8] In the context of designing sustainable NRFBs, we envision the fragmentation process to liberate the damaged redox-active cores allowing the deposited redoxmer to return to solution and restore the electrode surface (Scheme 1). In the literature, many classes of molecular scaffolds such as alkoxyamines,[9,10] aromatic disulfides,[11] enol carbonates,[12] and haloacetonitriles[13] have been reported to undergo mesolytic cleavage. Among these, homobenzylic ethers (HBEs) are of practical interest as their mesolytic cleavage properties have been previously utilized for the development and applications of electron-transfer-initiated reactions.[14] Therefore, HBE was chosen as the molecular scaffold for mesolytic cleavage motif in this work.

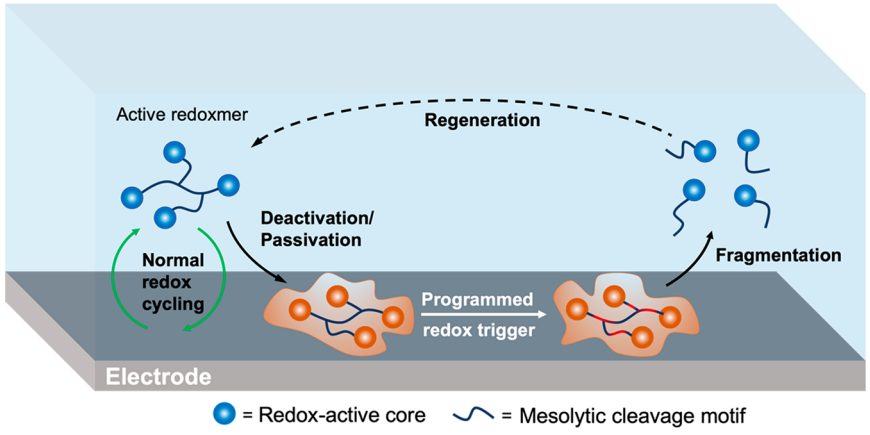
For a molecule to undergo mesolytic cleavage, it must first be oxidized or reduced. Hence, identifying candidate tether motifs with suitable cleavage potentials is the first important step toward the imagined restoration function. As shown in Figure 1a, an HBE scaffold with simple functional group variations generates a set of 1400 molecular candidates through systematic enumeration. The experimental investigation of such a vast structural space is impractical, being both time-consuming and expensive. In contrast, accurate quantum mechanical calculations, e.g., density functional theory (DFT), provide an alternative to accelerate the screening process.[15−17] Indeed, it has been shown that DFT-computed redox potentials of organic molecules are in good agreement with experimental measurements.[18−20] For example, Davis and Fry performed DFT simulations to compute reduction and oxidation potentials for 51 polycyclic aromatic hydrocarbons

Received: February 21, 2020
Revised: May 28, 2020
Published: May 28, 2020



https://dx.doi.org/10.1021/acs.chemmater.0c00768
Chem. Mater. 2020, 32, 6338–6346

Scheme 1. Strategy for Regenerating Active Redoxmer at the Electrolyte/Electrode Interface via Incorporating Programmable Mesolytic Cleavage Motifs



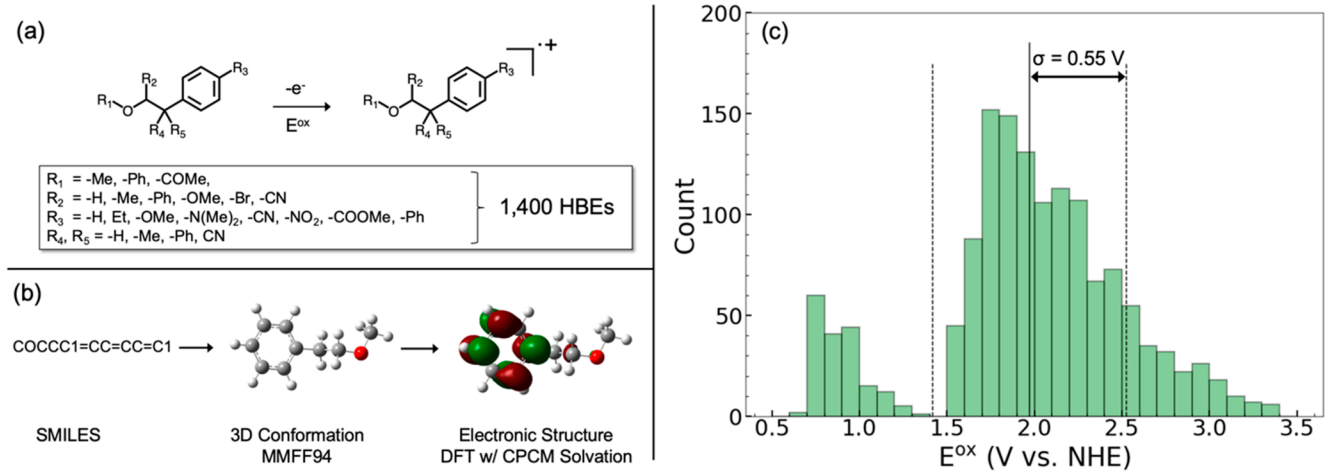


Figure 1. Systematic enumeration of the R-designated residues generates a 1400-molecule homobenzylic ether (HBE) library. (b) Automated scheme for geometry generation and DFT optimization. (c) Distribution of computed oxidation potential values for 1400 HBEs.

and showed a mean absolute deviation of 0.056 V compared to the experimental measurememnts.[18]

As the number of functional groups increases, the number of possible HBEs also grows in a combinatorial fashion. Therefore, even high-throughput computational screening approaches will eventually become intractable. Recently, machine learning (ML) has been successfully applied to predict various materials properties including crystal stability,[21,22] diffusion barrier,[23] band gap,[24,25] melting point,[26] superconducting transition temperature,[27] catalytic activity,[28,29] and others.[30,31] Consequently, ML offers an accelerated approach to materials screening and discovery from extensive molecular libraries.[32−38] Compared to physics-based calculations (e.g., DFT), ML models offer a more efficient method for materials exploration of very large search space.[39] However, a reliable supervised ML model such as regression often requires a substantial amount of training data, which typically come from expensive simulations or a large number of experiments. Active learning, on the other hand, constructs a self-improving cycle that dynamically queries new data for evaluation to maximize its predictive power. Thus, active learning is most useful when data are scarce or difficult to produce, which is often the case in materials research. Indeed, successful demonstrations of active learning applied to materials discovery have been highlighted across multiple disciplines in the recent literature.[40−46]

Here, we employed high-throughput DFT calculations and an active learning model based on Bayesian optimization (BO) to screen HBEs for the ideal design of redoxmers capable of programmed destruction. As the first step toward this goal, we aim to develop a computational method to identify HBE tethers that have redox potentials suitable for a particular electrochemical cell that operates using a given pair of half reactions. Specifically, for potential compatibility with catholytes in NRFBs, it is desirable to select HBEs with an oxidation potential (here noted as E^{ox} implying that mesolytic cleavage is initiated by the oxidation of the molecule) between 1.40 and 1.70 V (vs NHE) for two primary reasons. First, many catholytes including dialkoxybenzenes (DMB),[47] anthraquinone,[48] and 2,2,6,6-tetramethylpiperidin-1-oxyl (TEMPO)[49] have oxidation potentials below 1.40 V vs NHE, so they could be cycled without triggering HBE cleavage until the latter is wanted. Second, applying high oxidative potentials past 1.70 V is known to cause a breakdown in typical NRFB electrolytes and solvents, such as acetonitrile, which is used here for the calculation of E^{ox}. With such criteria, we built a library of HBEs and computed their corresponding oxidation potentials. Subsequently, we assessed a Gaussian process regression model for E^{ox} prediction of HBEs as a potential

https://dx.doi.org/10.1021/acs.chemmater.0c00768
Chem. Mater. 2020, 32, 6338−6346

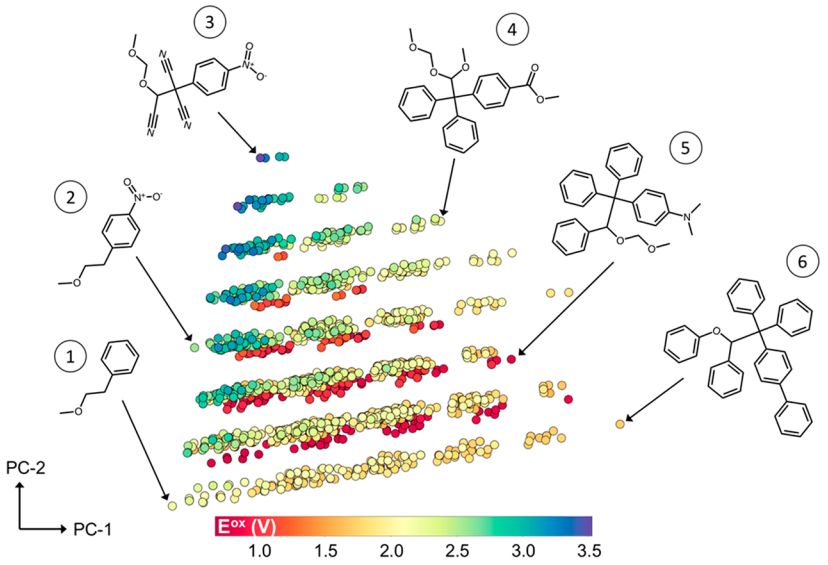


Figure 2. Graphical illustration of the chemical space of 1400 HBEs and their computed oxidation potentials (E^{ox}). PC-1 and PC-2 represent principle components 1 and 2, respectively.

surrogate model for Bayesian optimization. Finally, we developed and evaluated the performance of Bayesian optimization on the DFT-computed data set as well as a previously unseen data set to search for optimal HBE candidates.

RESULTS AND DISCUSSION

Library Generation and High-Throughput DFT Simulation. The systematic enumeration of HBEs with various functional groups (Rs) begins with a scaffold in the form of a Simplified Molecular-Input Line-Entry System (SMILES) string, namely, [R3]C1 = CC = C(C([R4])([R5])C([R2])-[O][R1]) C = C1. Then, by substituting the R-designated residues within the scaffold with the appropriate functional groups (also in SMILES format), a total of ca. 1400 unique HBEs were generated (Figure 1a). Note that this is not an exhaustive collection of such functional groups. This initial set was chosen primarily for their electron-donating and -withdrawing properties, as well as ease of synthesis. Overall, this HBE data set is diverse as the constituent molecules are composed of 10−39 heavy (i.e., non-hydrogen) atoms and five elements: C, H, O, N, and Br. To validate our computational approach, we compared DFT-derived oxidation potentials with cyclic voltammetry measurements for 11 HBE molecules from the data set. The benchmarking results show a mean absolute error of 0.11 V (Figure S1), indicating sufficient accuracy of our DFT method in computing oxidation potentials. In order to perform DFT calculations on all generated species, we employed a high-throughput workflow that primarily consists of two steps (Figure 1b): (1) Conversion of SMILES inputs to three-dimensional structures, of which initial conformations were created at the force-field level (MMFF94) as implemented in the cheminformatics software RDKit.[50] (2) Generation and mass submission of simulation input scripts to high-performance computing (HPC) systems for quantum mechanical evaluation. The distribution of the computed E^{ox} for 1400 HBEs is shown in Figure 1c, with the mean (μ) and standard deviation (σ) of 1.98 and 0.55 V, respectively. According to Figure 1c, the evaluated HBEs possess a wide E^{ox} window, ranging from 0.50 to 3.50 V. Most importantly, we identified a total of 133 molecules with the desired E^{ox} between 1.40 and 1.70 V (Table S1), which constitutes ca. 9% of the entire HBE candidate library. From Table S1, we noted the majority of the identified HBEs possess either a methoxy (64%) or a phenyl functional group (35%) at the para position of the benzene ring in the scaffold (i.e., R_3 = −OMe or −Ph in Figure 1a). Furthermore, for a randomly selected subset of 10 HBEs, Table S2 shows the exergonic nature of bond fragmentations upon one-electron oxidation and indicates the presence of mesolytic cleavages among the identified molecules. A more rigorous redox-triggered mesolytic cleavage study employing both computations and experiments will be carried out in the future.

Feature Generation and Dimensionality Reduction. In order to create an accurate model to predict E^{ox} for HBEs, we utilized an extensive set of physical and chemical descriptors such as molecular weight, topological surface area, number of valence electrons, and number of aromatic rings. A total of 49 such features were generated automatically from SMILES inputs using RDKit and tabulated in Table S3. Pearson correlation analysis indicates no single feature among those 49 can reliably capture the trend of computed E^{ox} (Figure S2 and Table S3). Therefore, to extract the key components as well as reduce the dimensionality of our feature vector, we performed principle component analysis (PCA). In PCA, original feature components are combined linearly to create new features called principle components (PCs). These PCs are not only linearly independent but also created in a hierarchical order of their contributions such that the first PC captures the most variability in the data, followed by the second PC and so on. When PCA was performed on our data set, we found that 30 PCs are enough to account for 100% of the variance among all 1400 HBEs (Figure S3). Furthermore, since the first 15 PCs already explains 99.7% of the variance, they were chosen as the final feature set that yields a reasonable compromise between accuracy and efficiency. In Figure 2 the distribution of E^{ox} is also plotted against two leading PCs (i.e., PCs that account for the most variance), namely, PC-1 and PC-2. From small HBEs such as (2-methoxyethyl) benzene on the far left (1) to large molecules

https://dx.doi.org/10.1021/acs.chemmater.0c00768
Chem. Mater. 2020, 32, 6338−6346

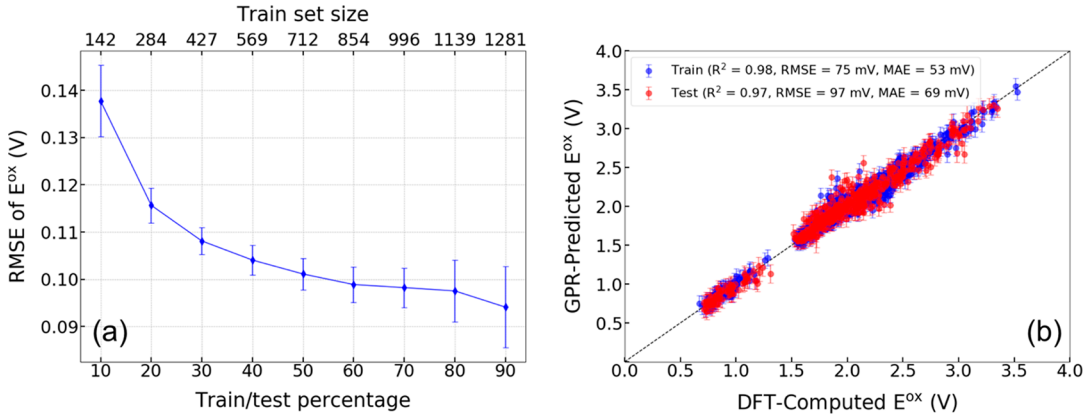


Figure 3. (a) Prediction accuracy of GPR models evaluated at different train set sizes, averaged over 100 runs. (b) Parity plot obtained from the model using a train/test ratio of 60%. In both plots, the error bars indicate one standard deviation.

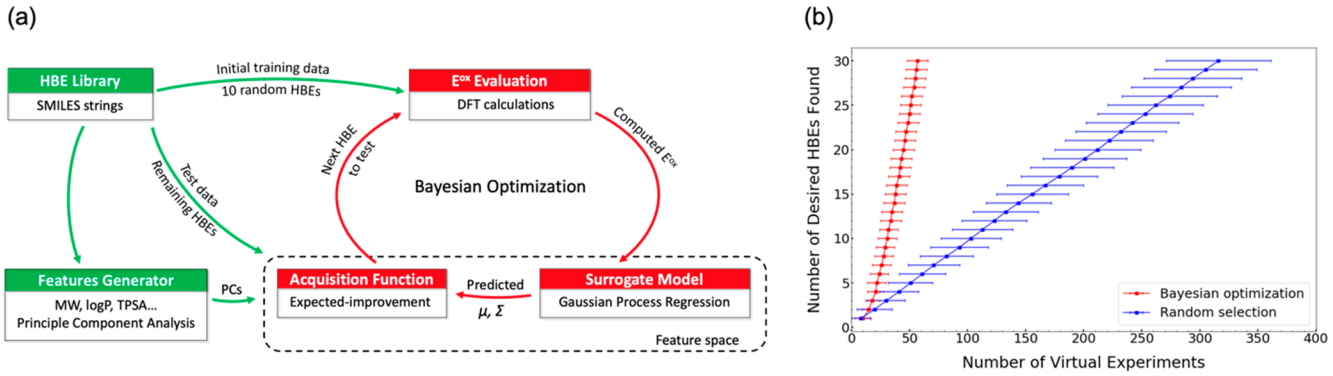


Figure 4. (a) Schematic representation of Bayesian optimization-guided computational workflow for identifying HBEs with desired oxidation potentials (E^{ox}). PCs, μ, and Σ are principle components, mean, and variance, respectively. (b) Comparison of computational efficiency between Bayesian optimization (red) and random selection (blue). The error bars denote one standard deviation from 100 repeated runs.

with multiple aromatic rings such as 4-(2-phenoxy-1,1,2-triphenylethyl)-1,1′-biphenyl on the far right (6), Figure 2 illustrates the structural and electrochemical diversity of our library of 1400 HBEs.

Prediction of Oxidation Potentials of Materials Using Gaussian Process Regression. With the 15-PC feature vectors, we built a Gaussian process regression (GPR) model to predict the oxidation potentials[51] of HBEs. GPR is a nonparametric (i.e., not limited to a functional form) approach to regression with the ability to provide uncertainty on the predictions. The statistically meaningful predictions are based on a Bayesian framework, wherein a prior Gaussian process is formulated from training data and a covariance (or kernel) function. In this work, we used the Matérn kernel, a generalization of the radial basis function, to construct the covariance matrices (see Methods). Two metrics were used to evaluate the performance of the GPR models, namely, root mean squared error (RMSE) and coefficient of determination (R^2). Nine models were tested to determine the optimal train/test ratio, and the resulting learning curve is shown in Figure 3a. The construction of each model first involves selecting a percentage (between 10% and 90% in our case) of the entire library of HBEs to serve as the training set. Then, the remaining molecules, which have not been seen by the model, are used as the test data set. In addition, a total of 100 runs (each run starting with a random training/test split) were repeated for each model to build a reasonable statistic of the model performance.[52] For comparison, similar learning curves using different feature sets are plotted in Figure S4, which shows similar accuracy (RMSE < 0.15 V) for models utilizing either 15 PCs, 30 PCs, or 49 original features. Again, this observation confirms the adequacy of our choice of the 15 PCs as descriptors for E^{ox} prediction.

According to Figure 3a, we found that increasing the train/test ratio beyond 60% did not lead to any notable improvement in accuracy. The final regression model, utilizing a train/test ratio of 60%, yields accurate E^{ox} predictions with R^2 = 0.97 and RMSE = 97 mV with respect to DFT-computed values (Figure 3b). Such excellent predictive power of our GPR model provides us confidence in constructing a Bayesian optimization model to identify HBE candidates with desirable E^{ox}, wherein GPR serves as the surrogate model.

Bayesian Optimization (BO) for Prediction of Materials with Desirable Oxidation Potentials. The goal of a BO process applied to computational materials screening is to efficiently identify candidates with desired property while requiring only a minimal number of expensive evaluations such as DFT calculations. At the beginning of a typical BO process, a small, randomly selected number of the data points are used to train a surrogate model. Then, based on the predicted mean values and uncertainties obtained from the surrogate model, the acquisition function optimizes the objective function by suggesting the next data point(s) for evaluation. New labeled data are used to improve the surrogate model and complete one BO iteration.

https://dx.doi.org/10.1021/acs.chemmater.0c00768
Chem. Mater. 2020, 32, 6338−6346

The specific BO process used in the evaluation of 1400 HBEs is shown in Figure 4a. Initially, we employed 10 randomly chosen HBEs and their corresponding DFT-computed E^{ox} to train a GPR surrogate model in a similar fashion as the one detailed in the previous section. This GPR model was used to make predictions of E^{ox} for the remaining HBE molecules in the library. Then, the expected improvement (EI) acquisition function utilized both the prediction and the uncertainty to suggest the next best HBE candidate for DFT evaluation. The EI is a widely used state-of-the-art acquisition function that balances between exploitation (following the trend of the current best estimates) and exploration (diversifying the search to avoid local optima). Once the E^{ox} value of the new HBE was computed, it was added to the current training data set to retrain the GPR model in the next iteration of the BO loop (Figure 4a, red). In order to evaluate the performance of BO on our HBE database, we programmed the BO cycle to repeat until a total of 30 molecules with $E^{ox} \in [1.40\ V, 1.70\ V]$ are found. Based on the statistics observed earlier, there is a 9% chance of finding such an HBE molecule. This low probability is reflected by the slope of the blue curve in Figure 4b, which was constructed by randomly selecting HBEs in the database and recording the number of those with the desired E^{ox}. Indeed, we found that it would take approximately 330 random selections to find 30 desirable HBEs. In comparison, the same number of HBEs were correctly selected within 60 BO evaluations. Since every random selection or BO evaluation requires a DFT calculation, Figure 4b indicates at least 5-fold improvement in efficiency in identifying appropriate candidates obtained from using BO over the random selection. Similar observation can be seen for the comparison between BO and Latin hypercube sampling (Figure S6).

We further examined the generalizability of the BO process by applying it to a significantly larger, previously unseen data set of approximately 112 000 HBEs. This new data set was created by incorporating a wider range of R-designated residues to the HBE scaffold (Table S4). We conducted two independent BO runs, BO1 and BO2, each with an initial training set of 10 randomly chosen HBEs and set the maximum number of HBEs to evaluate to 40. With this setup, both BO runs successfully discovered a number of HBE molecules with the desired E^{ox} (26 in BO1 and 16 in BO2), and the evaluation progress of BO1 is shown in Figure 5. It can be seen that, after an initial exploration within the first 15 steps, BO1's predictive capability starts improving significantly as the gap between the predicted and DFT-computed values decreases rapidly. Table 1 shows several HBE candidates discovered by BO1 and their corresponding R-designated residue identities. Interestingly, we notice that BO1 consistently identifies desired HBEs with either a propoxy or isopropoxy functional group at the R_3 position with respect to the scaffold (i.e., R_3= −OPr or −OiPr) within 40 BO evaluations. Similarly, we identified 16 HBE molecules from BO2. These structural patterns, together with those found previously in the 133 molecules identified from the 1400 data set, suggests the presence of electron donating groups such as alkoxy and phenyl at the para position for designing HBEs with oxidation potential window of [1.40 V, 1.70 V vs NHE]. We note that a complete list of HBE candidates predicted by BO1 and BO2 is provided in Tables S5 and S6, respectively. In summary, we demonstrated that BO coupled with DFT is a robust and efficient method for identifying molecules with desirable properties from large databases.

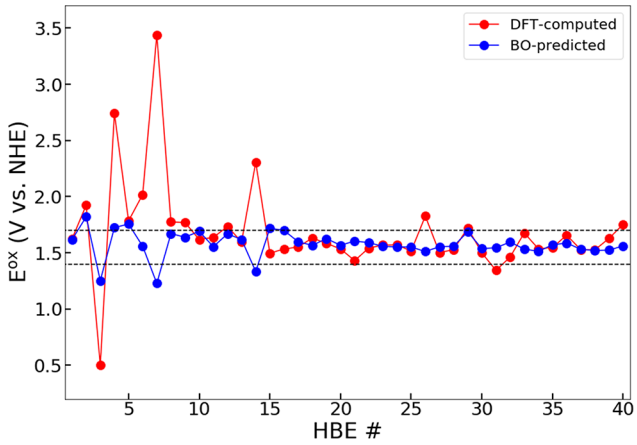


Figure 5. Predicted (blue) and DFT-computed (red) oxidation potentials (E^{ox}) of 40 suggested candidates from a 112 000 HBE data set by BO1. The black dashed lines indicate the desired E^{ox} range (1.40 V−1.70 V). The initial training set consists of 10 randomly selected HBEs.

CONCLUSIONS

We conducted high-throughput density functional theory (DFT) simulations to compute oxidation potentials (E^{ox}) values for 1400 homobenzylic ether (HBE) molecules, to program redoxmer destruction via a mesolytic cleavage reaction. We identified a small percentage, ca. 9%, of these HBEs to possess E^{ox} in the ideal window of [1.40 V, 1.70 V, vs NHE]. Based on the performance of an initial Gaussian process regression (GPR) model for E^{ox} prediction, we implemented an active learning framework based on Bayesian optimization (BO) for efficient identification of desired HBEs. The BO not only determined the ideal HBEs correctly but also demonstrated more than 5-fold improvement in computational efficiency compared to the random selection. When BO was further applied to a previously unseen data set of 112 000 HBEs, it successfully identified 42 optimal HBE candidates after evaluating only 100 molecules. Moreover, BO's ability to uncover potential structural patterns that lead to desired properties may help us recognize important underlying structure−property relations. The use of active learning to efficiently guide high-throughput quantum mechanical simulations provides an accelerated practice to identify candidate components suitable for testing design concepts in the discovery of new materials.

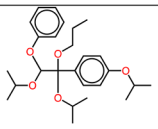
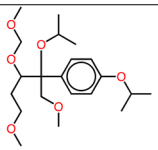
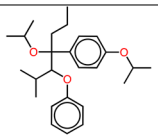
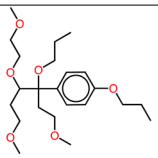
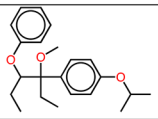
METHODS

Density Functional Theory (DFT). All DFT calculations for neutral and oxidized HBE molecules were performed using the B3LYP/6-31+G(d,p)[53] level of theory as implemented in Gaussian 16, Revision A.03.[54] To account for the free energies of solvation in acetonitrile solvent (ϵ = 37.5), self-consistent reaction-field (SCRF) calculations using the conductor-like polarizable continuum model (CPCM)[55,56] were employed. Numerical integrations were carried using an ultrafine grid. The oxidation potentials, E^{ox}, were computed from the Gibbs free energy change at 298 K (ΔG^{ox}) for the elimination of an electron from the species of interest ($\Delta G^{ox} = G^{oxidized} - G^{neutral}$), using the equation:[57]

$$E^{ox} = \Delta G^{ox}/nF - \text{NHE} \quad (1)$$

https://dx.doi.org/10.1021/acs.chemmater.0c00768
Chem. Mater. 2020, 32, 6338−6346

Table 1. Examples of Desired HBEs Identified Using Bayesian Optimization (Denoted as BO1 above) and DFT-Computed Oxidation Potential (E^{ox} in Volts vs NHE)[a]

HBE #	2-D Structure	R_1	R_2	R_3	R_4	R_5	E^{ox}
13		-Ph	-OPh	-OiPr	-OPr	-OiPr	1.60
15		-COMe	-EtOMe	-OiPr	-OiPr	-COMe	1.50
17		-Ph	-iPr	-OiPr	-OiPr	-Pr	1.55
18		-EtOMe	-EtOMe	-OPr	-OPr	-EtOMe	1.63
20		-Ph	-Et	-OiPr	-OMe	-Et	1.53

[a]The HBE entry numbers (#) are shown in Figure 5. A complete set of molecules identified from BO1 is given in Table S5 of the Supporting Information. The R_i (i = 1 to 5) are substituent groups of the HBE scaffold in Figure 1a. The substituent groups are Ph, phenyl; Et, ethyl; OMe, methoxy; iPr, isopropyl; OPr, propoxy; EtOMe, methoxy ethyl; Pr, propyl.

where n is the number of electrons, F is the Faraday constant, and NHE is the absolute potential of the normal hydrogen electrode, 4.28 V.[58]

Gaussian Process Regression (GPR). All GPR models was built with the Scikit-learn machine learning package.[59] The predicted mean, $\boldsymbol{\mu}_*$, and variance, $\boldsymbol{\Sigma}_*$, are given as

$$\boldsymbol{\mu}_* = \mathbf{K}_*^{\mathrm{T}}\mathbf{K}_{\mathrm{y}}^{-1}\mathbf{y} \tag{2}$$

$$\boldsymbol{\Sigma}_* = \mathbf{K}_{**} - \mathbf{K}_*^{\mathrm{T}}\mathbf{K}_*^{-1}\mathbf{K}_* \tag{3}$$

where $\mathbf{y}$ is the labeled property of the train set and $\mathbf{K} = k(\mathbf{X}, \mathbf{X})$, $\mathbf{K}_* = k(\mathbf{X}, \mathbf{X}_*)$, and $\mathbf{K}_{**} = k(\mathbf{X}_*, \mathbf{X}_*)$. $\mathbf{X}$ and $\mathbf{X}_*$ are the feature matrix of the train and test set, respectively. If $\mathbf{x}$ and $\mathbf{x}'$ represent the feature vectors, then their covariance function $k(\mathbf{x}, \mathbf{x}')$, based on the Matérn kernel (ν = 1.5),[60] is expressed as the followings

$$k(\mathbf{x}, \mathbf{x}') = \left(1 + \frac{\sqrt{3}\,\|\mathbf{x} - \mathbf{x}'\|}{\sigma_l}\right)\exp\left(-\frac{\sqrt{3}\,\|\mathbf{x} - \mathbf{x}'\|}{\sigma_l}\right) + \sigma_n^2 \tag{4}$$

where σ_l and σ_n are length scale and the expected noise level of the data, respectively. These hyperparameters were determined using the maximum likelihood estimation during the training of our models.

Expected Improvement (EI) Acquisition Function. The equation for computing EI is as follows:[46]

$$\mathrm{EI}(\mathbf{x}) = \begin{cases} (\mu(\mathbf{x}) - f(\mathbf{x}^+))\Phi(Z) + \sigma(\mathbf{x})\phi(Z) & \sigma(\mathbf{x}) > 0 \\ 0 & \sigma(\mathbf{x}) = 0 \end{cases} \tag{5}$$

$$Z = \frac{\mu(\mathbf{x}) - f(\mathbf{x}^+) - \epsilon}{\sigma(\mathbf{x})} \tag{6}$$

where $\mu(\mathbf{x})$ and $\sigma(\mathbf{x})$ are the predicted mean and standard deviation from the GPR model, $f(\mathbf{x})$. $\Phi(Z)$ is the cumulative density function (CDF), and $\phi(Z)$ is the probability density function (PDF). $f(\mathbf{x}^+)$ is the property of the best material so far, and $\mathbf{x}^+$ is the feature vector of that material. The parameter ϵ in eq 6 determines the amount of exploration during optimization, and we used a constant value of 0.01 as this yields the optimal trade-off between exploration and exploitation in our data set (Figure S5).

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.chemmater.0c00768.

Homobenzylic ether data sets, principle component analysis, and Gaussian process regression learning curves (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Rajeev S. Assary − *Materials Science Division and Joint Center for Energy Storage Research, Argonne National Laboratory, Lemont, Illinois 60439, United States;* orcid.org/0000-0002-9571-3307; Phone: 630-252-3536; Email: assary@anl.gov

Authors

Hieu A. Doan − *Materials Science Division and Joint Center for Energy Storage Research, Argonne National Laboratory, Lemont, Illinois 60439, United States*

https://dx.doi.org/10.1021/acs.chemmater.0c00768
Chem. Mater. 2020, 32, 6338−6346

Garvit Agarwal − *Materials Science Division and Joint Center for Energy Storage Research, Argonne National Laboratory, Lemont, Illinois 60439, United States*

Hai Qian − *The Beckman Institute for Advanced Science and Technology and Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States; Joint Center for Energy Storage Research, Argonne National Laboratory, Lemont, Illinois 60439, United States;* orcid.org/0000-0002-5304-132X

Michael J. Counihan − *The Beckman Institute for Advanced Science and Technology and Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States; Joint Center for Energy Storage Research, Argonne National Laboratory, Lemont, Illinois 60439, United States*

Joaquín Rodríguez-López − *The Beckman Institute for Advanced Science and Technology and Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States; Joint Center for Energy Storage Research, Argonne National Laboratory, Lemont, Illinois 60439, United States;* orcid.org/0000-0003-4346-4668

Jeffrey S. Moore − *The Beckman Institute for Advanced Science and Technology, Department of Chemistry, and Department of Materials Science and Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States; Joint Center for Energy Storage Research, Argonne National Laboratory, Lemont, Illinois 60439, United States;* orcid.org/0000-0001-5841-6269

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.chemmater.0c00768

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

Gaussian process regression and Bayesian optimization codes used in this work can be found on Github at https://github.com/hieuadoan/QM-ActiveLearning-Paper.

■ ACKNOWLEDGMENTS

This work was supported as part of the Joint Center for Energy Storage Research, an Energy Innovation Hub funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences. We gratefully acknowledge the computing resources provided on "Bebop", a 1024-node computing cluster operated by the Laboratory Computing Resource Center at Argonne National Laboratory. We also thank Dr. Pankaj Rajak for the valuable discussions on Bayesian optimization.

■ ABBREVIATIONS

HBE, homobenzylic ether; DFT, density functional theory; ML, machine learning; PCA, principle component analysis; PCs, principle components; BO, Bayesian optimization

■ REFERENCES

(1) Department of Energy (DOE). *Solving Challenges in Energy Storage*; 2018; p 51.

(2) Kowalski, J. A.; Su, L.; Milshtein, J. D.; Brushett, F. R. Recent Advances in Molecular Engineering of Redox Active Organic Molecules for Nonaqueous Flow Batteries. *Curr. Opin. Chem. Eng.* **2016**, *13*, 45−52.

(3) Wang, W.; Luo, Q.; Li, B.; Wei, X.; Li, L.; Yang, Z. Recent Progress in Redox Flow Battery Research and Development. *Adv. Funct. Mater.* **2013**, *23* (8), 970−986.

(4) Ding, Y.; Zhang, C.; Zhang, L.; Zhou, Y.; Yu, G. Molecular Engineering of Organic Electroactive Materials for Redox Flow Batteries. *Chem. Soc. Rev.* **2018**, *47* (1), 69−103.

(5) Wei, X.; Xu, W.; Huang, J.; Zhang, L.; Walter, E.; Lawrence, C.; Vijayakumar, M.; Henderson, W. A.; Liu, T.; Cosimbescu, L.; et al. Radical Compatibility with Nonaqueous Electrolytes and Its Impact on an All-Organic Redox Flow Battery. *Angew. Chem., Int. Ed.* **2015**, *54* (30), 8684−8687.

(6) Goulet, M. A.; Tong, L.; Pollack, D. A.; Tabor, D. P.; Odom, S. A.; Aspuru-Guzik, A.; Kwan, E. E.; Gordon, R. G.; Aziz, M. J. Extending the Lifetime of Organic Flow Batteries via Redox State Management. *J. Am. Chem. Soc.* **2019**, 8−13.

(7) Huang, J.; Pan, B.; Duan, W.; Wei, X.; Assary, R. S.; Su, L.; Brushett, F. R.; Cheng, L.; Liao, C.; Ferrandon, M. S.; et al. The Lightest Organic Radical Cation for Charge Storage in Redox Flow Batteries. *Sci. Rep.* **2016**, *6* (June), 1−9.

(8) Maslak, P.; Narvaez, J. N. Mesolytic Cleavage of C-C Bonds. Comparison with Homolytic and Heterolytic Processes in the Same Substrate. *Angew. Chem., Int. Ed. Engl.* **1990**, *29* (3), 283−285.

(9) Kirner, H. J.; Schwarzenbach, F.; Van Der Schaaf, P. A.; Hafner, A.; Rast, V.; Frey, M.; Nesvadba, P.; Rist, G. Synthesis of N-Alkoxy Amines via Catalytic Oxidation of Hydrocarbons. *Adv. Synth. Catal.* **2004**, *346* (5), 554−560.

(10) Zhu, Q.; Gentry, E. C.; Knowles, R. R. Catalytic Carbocation Generation Enabled by the Mesolytic Cleavage of Alkoxyamine Radical Cations. *Angew. Chem., Int. Ed.* **2016**, *55* (34), 9969−9973.

(11) Antonello, S.; Daasbjerg, K.; Jensen, H.; Taddei, F.; Maran, F. Formation and Cleavage of Aromatic Disulfide Radical Anions. *J. Am. Chem. Soc.* **2003**, *125* (48), 14905−14916.

(12) Schmittel, M.; Peters, K.; Peters, E. M.; Haeuseler, A.; Trenkle, H. Radical Cation Ester Cleavage in Solution. Mechanism of the Mesolytic O-CO Bond Scission. *J. Org. Chem.* **2001**, *66* (10), 3265−3276.

(13) Cardinale, A.; Isse, A. A.; Gennaro, A.; Robert, M.; Savéant, J. M. Dissociative Electron Transfer to Haloacetonitriles. An Example of the Dependency of in-Cage Ion-Radical Interactions upon the Leaving Group. *J. Am. Chem. Soc.* **2002**, *124* (45), 13533−13539.

(14) Floreancig, P. E. Development and Applications of Electron-Transfer-Initiated Cyclization Reactions. *Synlett* **2007**, *2007* (2), 0191−0203.

(15) Li, S.; Xia, Y.; Amachraa, M.; Hung, N. T.; Wang, Z.; Ong, S. P.; Xie, R. J. Data-Driven Discovery of Full-Visible-Spectrum Phosphor. *Chem. Mater.* **2019**, *31* (16), 6286−6294.

(16) Woods-Robinson, R.; Broberg, D.; Faghaninia, A.; Jain, A.; Dwaraknath, S. S.; Persson, K. A. Assessing High-Throughput Descriptors for Prediction of Transparent Conductors. *Chem. Mater.* **2018**, *30* (22), 8375−8389.

(17) Cai, Y.; Xie, W.; Teng, Y. T.; Harikesh, P. C.; Ghosh, B.; Huck, P.; Persson, K. A.; Mathews, N.; Mhaisalkar, S. G.; Sherburne, M.; et al. High-Throughput Computational Study of Halide Double Perovskite Inorganic Compounds. *Chem. Mater.* **2019**, *31* (15), 5392−5401.

(18) Davis, A. P.; Fry, A. J. Experimental and Computed Absolute Redox Potentials of Polycyclic Aromatic Hydrocarbons Are Highly Linearly Correlated over a Wide Range of Structures and Potentials. *J. Phys. Chem. A* **2010**, *114* (46), 12299−12304.

(19) Han, Y. K.; Jung, J.; Cho, J. J.; Kim, H. J. Determination of the Oxidation Potentials of Organic Benzene Derivatives: Theory and Experiment. *Chem. Phys. Lett.* **2003**, *368* (5−6), 601−608.

(20) Baik, M. H.; Friesner, R. A. Computing Redox Potentials in Solution: Density Functional Theory as a Tool for Rational Design of Redox Agents. *J. Phys. Chem. A* **2002**, *106* (32), 7407−7412.

(21) Bartel, C. J.; Sutton, C.; Goldsmith, B. R.; Ouyang, R.; Musgrave, C. B.; Ghiringhelli, L. M.; Scheffler, M. New Tolerance

Factor to Predict the Stability of Perovskite Oxides and Halides. *Sci. Adv.* **2019**, *5* (2), eaav0693.
(22) Ye, W.; Chen, C.; Wang, Z.; Chu, I. H.; Ong, S. P. Deep Neural Networks for Accurate Predictions of Crystal Stability. *Nat. Commun.* **2018**, *9* (1), 1−6.
(23) Jalem, R.; Nakayama, M.; Kasuga, T. An Efficient Rule-Based Screening Approach for Discovering Fast Lithium Ion Conductors Using Density Functional Theory and Artificial Neural Networks. *J. Mater. Chem. A* **2014**, *2* (3), 720−734.
(24) Lee, J.; Seko, A.; Shitara, K.; Nakayama, K.; Tanaka, I. Prediction Model of Band Gap for Inorganic Compounds by Combination of Density Functional Theory Calculations and Machine Learning Techniques. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2016**, *93* (11), 1−12.
(25) Pilania, G.; Gubernatis, J. E.; Lookman, T. Multi-Fidelity Machine Learning Models for Accurate Bandgap Predictions of Solids. *Comput. Mater. Sci.* **2017**, *129*, 156−163.
(26) Gu, T.; Lu, W.; Bao, X.; Chen, N. Using Support Vector Regression for the Prediction of the Band Gap and Melting Point of Binary and Ternary Compound Semiconductors. *Solid State Sci.* **2006**, *8* (2), 129−136.
(27) Owolabi, T. O.; Akande, K. O.; Olatunji, S. O. Prediction of Superconducting Transition Temperatures for Fe-Based Superconductors Using Support Vector Machine. *Adv. Phys. Theor. Appl.* **2014**, *35*, 12−26.
(28) Jinnouchi, R.; Hirata, H.; Asahi, R. Extrapolating Energetics on Clusters and Single-Crystal Surfaces to Nanoparticles by Machine-Learning Scheme. *J. Phys. Chem. C* **2017**, *121* (47), 26397−26405.
(29) Chowdhury, A. J.; Yang, W.; Walker, E.; Mamun, O.; Heyden, A.; Terejanu, G. A.; et al. Prediction of Adsorption Energies for Chemical Species on Metal Catalyst Surfaces Using Machine Learning. *J. Phys. Chem. C* **2018**, *122* (49), 28142−28150.
(30) Deringer, V. L.; Caro, M. A.; Jana, R.; Aarva, A.; Elliott, S. R.; Laurila, T.; Csányi, G.; Pastewka, L. Computational Surface Chemistry of Tetrahedral Amorphous Carbon by Combining Machine Learning and Density Functional Theory. *Chem. Mater.* **2018**, *30* (21), 7438−7445.
(31) Pilania, G.; Whittle, K. R.; Jiang, C.; Grimes, R. W.; Stanek, C. R.; Sickafus, K. E.; Uberuaga, B. P. Using Machine Learning to Identify Factors That Govern Amorphization of Irradiated Pyrochlores. *Chem. Mater.* **2017**, *29* (6), 2574−2583.
(32) Kim, C.; Pilania, G.; Ramprasad, R. From Organized High-Throughput Data to Phenomenological Theory Using Machine Learning: The Example of Dielectric Breakdown. *Chem. Mater.* **2016**, *28* (5), 1304−1311.
(33) Gómez-Bombarelli, R.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Duvenaud, D.; Maclaurin, D.; Blood-Forsythe, M. A.; Chae, H. S.; Einzinger, M.; Ha, D. G.; Wu, T.; et al. Design of Efficient Molecular Organic Light-Emitting Diodes by a High-Throughput Virtual Screening and Experimental Approach. *Nat. Mater.* **2016**, *15* (10), 1120−1127.
(34) Ren, F.; Ward, L.; Williams, T.; Laws, K. J.; Wolverton, C.; Hattrick-Simpers, J.; Mehta, A. Accelerated Discovery of Metallic Glasses through Iteration of Machine Learning and High-Throughput Experiments. *Sci. Adv.* **2018**, *4* (4), eaaq1566.
(35) Sendek, A. D.; Cubuk, E. D.; Antoniuk, E. R.; Cheon, G.; Cui, Y.; Reed, E. J. Machine Learning-Assisted Discovery of Solid Li-Ion Conducting Materials. *Chem. Mater.* **2019**, *31* (2), 342−352.
(36) Cubuk, E. D.; Sendek, A. D.; Reed, E. J. Screening Billions of Candidates for Solid Lithium-Ion Conductors: A Transfer Learning Approach for Small Data. *J. Chem. Phys.* **2019**, *150* (21), 214701.
(37) Ahmad, Z.; Xie, T.; Maheshwari, C.; Grossman, J. C.; Viswanathan, V. Machine Learning Enabled Computational Screening of Inorganic Solid Electrolytes for Suppression of Dendrite Formation in Lithium Metal Anodes. *ACS Cent. Sci.* **2018**, *4* (8), 996−1006.
(38) Hautier, G.; Fischer, C. C.; Jain, A.; Mueller, T.; Ceder, G. Finding Natures Missing Ternary Oxide Compounds Using Machine Learning and Density Functional Theory. *Chem. Mater.* **2010**, *22* (12), 3762−3767.
(39) Chen, C.; Zuo, Y.; Ye, W.; Li, X.; Deng, Z.; Ong, S. P. A Critical Review of Machine Learning of Energy Materials. *Adv. Energy Mater.* **2020**, *10*, 1903242.
(40) Gopakumar, A. M.; Balachandran, P. V.; Xue, D.; Gubernatis, J. E.; Lookman, T. Multi-Objective Optimization for Materials Discovery via Adaptive Design. *Sci. Rep.* **2018**, *8* (1), 1−12.
(41) Mannodi-Kanakkithodi, A.; Pilania, G.; Huan, T. D.; Lookman, T.; Ramprasad, R. Machine Learning Strategy for Accelerated Design of Polymer Dielectrics. *Sci. Rep.* **2016**, *6*, 20952.
(42) Lookman, T.; Balachandran, P. V.; Xue, D.; Yuan, R. Active Learning in Materials Science with Emphasis on Adaptive Sampling Using Uncertainties for Targeted Design. *npj Comput. Mater.* **2019**, *5* (1), 21 DOI: 10.1038/s41524-019-0153-8.
(43) Xue, D.; Balachandran, P. V.; Hogden, J.; Theiler, J.; Xue, D.; Lookman, T. Accelerated Search for Materials with Targeted Properties by Adaptive Design. *Nat. Commun.* **2016**, *7*, 1−9.
(44) Yuan, R.; Liu, Z.; Balachandran, P. V.; Xue, D.; Zhou, Y.; Ding, X.; Sun, J.; Xue, D.; Lookman, T. Accelerated Discovery of Large Electrostrains in BaTiO3-Based Piezoelectrics Using Active Learning. *Adv. Mater.* **2018**, *30* (7), 1702884.
(45) Kim, C.; Chandrasekaran, A.; Jha, A.; Ramprasad, R. Active-Learning and Materials Design: The Example of High Glass Transition Temperature Polymers. *MRS Commun.* **2019**, *9*, 860.
(46) Bassman, L.; Rajak, P.; Kalia, R. K.; Nakano, A.; Sha, F.; Sun, J.; Singh, D. J.; Aykol, M.; Huck, P.; Persson, K.; et al. Active Learning for Accelerated Design of Layered Materials. *npj Comput. Mater.* **2018**, *4* (1), 74.
(47) Silcox, B.; Zhang, J.; Shkrob, I. A.; Thompson, L.; Zhang, L. On Transferability of Performance Metrics for Redox-Active Molecules. *J. Phys. Chem. C* **2019**, *123* (27), 16516−16524.
(48) Wang, W.; Xu, W.; Cosimbescu, L.; Choi, D.; Li, L.; Yang, Z. Anthraquinone with Tailored Structure for a Nonaqueous Metal-Organic Redox Flow Battery. *Chem. Commun.* **2012**, *48* (53), 6669−6671.
(49) Wei, X.; Xu, W.; Vijayakumar, M.; Cosimbescu, L.; Liu, T.; Sprenkle, V.; Wang, W. TEMPO-Based Catholyte for High-Energy Density Nonaqueous Redox Flow Batteries. *Adv. Mater.* **2014**, *26* (45), 7649−7653.
(50) Landrum, G. *RDKit: Open-Source Cheminformatics Software.*
(51) Jinich, A.; Sanchez-Lengeling, B.; Ren, H.; Harman, R.; Aspuru-Guzik, A. A Mixed Quantum Chemistry/Machine Learning Approach for the Fast and Accurate Prediction of Biochemical Redox Potentials and Its Large-Scale Application to 315000 Redox Reactions. *ACS Cent. Sci.* **2019**, *5* (7), 1199−1210.
(52) Chen, L.; Tran, H.; Batra, R.; Kim, C.; Ramprasad, R. Machine Learning Models for the Lattice Thermal Conductivity Prediction of Inorganic Materials. *Comput. Mater. Sci.* **2019**, *170* (July), 109155.
(53) Rassolov, V. A.; Ratner, M. A.; Pople, J. A.; Redfern, P. C.; Curtiss, L. A. 6-31G* Basis Set for Third-Row Atoms. *J. Comput. Chem.* **2001**, *22* (9), 976−984.
(54) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; et al. *Gaussian 16*; Gaussian, Inc.: Wallingford, CT, 2016.
(55) Barone, V.; Cossi, M. Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model. *J. Phys. Chem. A* **1998**, *102* (11), 1995−2001.
(56) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. Energies, Structures, and Electronic Properties of Molecules in Solution with the C-PCM Solvation Model. *J. Comput. Chem.* **2003**, *24* (6), 669−681.
(57) Guerard, J. J.; Arey, J. S. Critical Evaluation of Implicit Solvent Models for Predicting Aqueous Oxidation Potentials of Neutral Organic Compounds. *J. Chem. Theory Comput.* **2013**, *9* (11), 5046−5058.
(58) Truhlar, D. G.; Cramer, C. J.; Lewis, A.; Bumpus, J. A. Molecular Modeling of Environmentally Important Processes: Reduction Potentials. *J. Chem. Educ.* **2004**, *81* (4), 596.

(59) Varoquaux, G.; Buitinck, L.; Louppe, G.; Grisel, O.; Pedregosa, F.; Mueller, A. Scikit-Learn. Get Mobile Mob. *Comput. Commun.* **2015**, *19* (1), 29−33.
(60) Rasmussen, C. E.; Williams, C. K. I. *Gaussian Processes for Machine Learning*; The MIT Press: 2006; Vol. *11*, DOI: 10.1142/S0129065704001899.

https://dx.doi.org/10.1021/acs.chemmater.0c00768
Chem. Mater. 2020, 32, 6338−6346