

Predicting the Redox Potentials of Phenazine Derivatives Using DFT-Assisted Machine Learning

Siddharth Ghule,* Soumya Ranjan Dash, Sayan Bagchi, Kavita Joshi,* and Kumar Vanka*

Cite This: *ACS Omega* 2022, 7, 11742–11755

Read Online

ACCESS |



Metrics & More

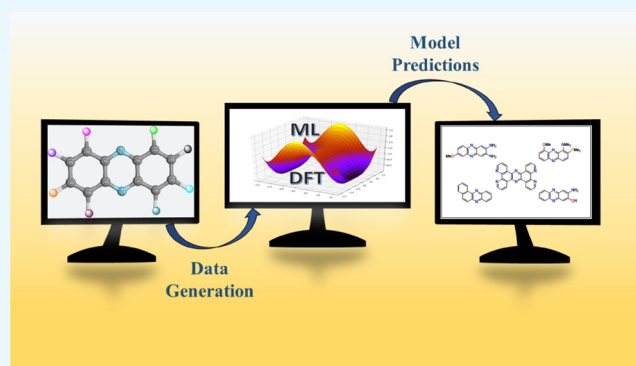


Article Recommendations



Supporting Information

ABSTRACT: This study investigates four machine-learning (ML) models to predict the redox potentials of phenazine derivatives in dimethoxyethane using density functional theory (DFT). A small data set of 151 phenazine derivatives having only one type of functional group per molecule (20 unique groups) was used for the training. Prediction accuracy was improved by a combined strategy of feature selection and hyperparameter optimization, using the external validation set. Models were evaluated on the external test set containing new functional groups and diverse molecular structures. High prediction accuracies of $R^2 > 0.74$ were obtained on the external test set. Despite being trained on the molecules with a single type of functional group, models were able to predict the redox potentials of derivatives containing multiple and different types of functional groups with good accuracies ($R^2 > 0.7$). This type of performance for predicting redox potential from such a small and simple data set of phenazine derivatives has never been reported before. Redox flow batteries (RFBs) are emerging as promising candidates for energy storage systems. However, new green and efficient materials are required for their widespread usage. We believe that the hybrid DFT-ML approach demonstrated in this report would help in accelerating the virtual screening of phenazine derivatives, thus saving computational and experimental costs. Using this approach, we have identified promising phenazine derivatives for green energy storage systems such as RFBs.



1. INTRODUCTION

Today, ~85% of the world's energy demand is being fulfilled by fossil fuels.^{1,2} The limited supply of fossil fuels and the ever-increasing population have raised concerns that we might run out of fossil fuels sooner than expected.^{1,3} Furthermore, electricity production from fossil fuels is one of the major factors responsible for greenhouse gas emissions.⁴ In this age, humanity faces two major challenges of balancing increased energy demand while reducing the environmental impact associated with energy production. In the past decades, investments and research efforts in the green technology have been increased to overcome these challenges.⁵ Significant progress has already been made to access renewable energy sources.^{6,7} Renewable energy sources, being intermittent, require efficient energy storage.⁴ Improvements in the energy storage technology would not only help in the adoption of renewable energy but also help in making efficient use of non-renewable energy sources. Historically, it has been more expensive to store energy than to expand energy generation for handling increased demand.⁸ Thus, grid systems employed today are likely to fail when additional energy cannot be generated during peak demand. The massive Texas Blackout in February 2021 is an example of such a failure.⁹ It suggests that an efficient energy storage technology is urgently required. Unfortunately, only 1.0% of the energy consumed worldwide

can be stored with the energy storage technology accessible today.¹⁰ Furthermore, the contribution of electrochemical batteries to energy storage capacity is less than 2.0%, even though most of the devices we use every day include batteries.^{8,10} Li-ion batteries are widely used today due to their high energy density, high specific energy, long cycle life, and fast charge–discharge cycle.^{4,8,11} Unfortunately, Li-ion batteries suffer from high production costs, safety issues, and high environmental impact.^{2,12} Redox flow batteries (RFBs) have the potential to overcome drawbacks of Li-ion batteries, owing to their high storage capacity, independent control over storage capacity and power, fast responsiveness, ease of scaling, room-temperature operation, cost-effectiveness, high round trip efficiency, safety, and lower environmental impact.^{13–26} RFBs are increasingly being used as energy storage devices in renewable energy systems, thereby helping in the adoption of green energy.^{15,22} A schematic diagram of the typical RFB is

Received: December 4, 2021

Accepted: February 7, 2022

Published: March 29, 2022



shown in Figure 1. The RFB consists of two storage tanks containing cathode and anode redox-active species dissolved in

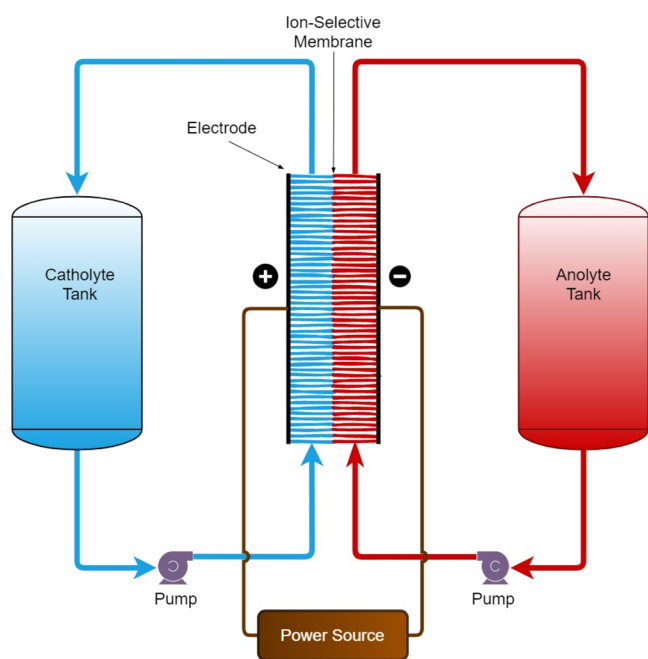


Figure 1. Schematic diagram of a typical RFB.

an electrolyte solution. The electrolyte solution in the positive and negative compartments is termed catholyte and anolyte, respectively. These storage tanks are connected to an electrochemical cell (or current collector) *via* pumps. The electrochemical cell consists of porous electrodes separated by an ion-selective membrane. During operation, electrolytes containing redox-active species are pumped to the electrochemical cell, where they undergo oxidation or reduction depending on the charge/discharge cycle. Then, electrolytes are circulated back to their storage tanks.^{13,24} So far, transition metal-based RFBs (such as vanadium, iron, and chromium) have found some commercial success. However, their widespread adoption has been limited mainly due to high production cost, toxicity, and cell component corrosion associated with the use of transition-metal salts.^{27,28} Therefore, RFBs containing organic redox-active species are being heavily investigated due to their low production cost, access to a massive space of electroactive compounds, and low environmental impact.^{28,29} Many organic compounds such as quinones, viologens, flavins, thiazines, imides, and their derivatives have been investigated for redox-active species in both aqueous and non-aqueous RFBs.^{27,30,31} However, non-aqueous RFBs offer large operating voltage.³⁰ Recently, phenazine derivatives have been shown to be promising redox-active candidates in non-aqueous RFBs. Recent reports have revealed why phenazine derivatives are promising redox-active candidates. Romadina et al. synthesized phenazine derivatives having significantly negative redox potential.³² RFBs require anolytes with high negative redox potential. They showed that the non-aqueous RFB based on the synthesized phenazine derivative is capable of achieving a potential of 2.3 V, high capacities, >95% Coulombic efficiency, and good charge–discharge cycling stability after the initial 20 cycles. Mavrandonakis and co-workers, in their computational investigation, reported the most negative redox-active

candidate based on phenazine for non-aqueous RFBs.²⁷ They showed that tetra-amino-phenazine has 140 mV more negative potential than *N*-methylphthalimide (MePht), which has one of the most negative redox potentials reported so far in RFBs.³³ They also proposed all-phenazine RFB reaching a high potential of 2.83 V. Furthermore, the redox potential of phenazine derivatives could be tuned easily with the addition of appropriate electron-donating or electron-withdrawing functional groups. The synthesis of phenazine derivatives is very economical than mining transition metals. Therefore, phenazine derivatives are currently being investigated as candidates for novel redox-active species.^{27,32}

These investigations remain primarily experimental. Unfortunately, the vast chemical space offered by organic compounds cannot be explored using experimental procedures. Quantum mechanical density functional theory (DFT) computations have been used heavily in materials science research due to high accuracy but are very slow and cannot screen millions of molecules in a reasonable amount of time. Therefore, a fast and reliable method to screen millions of compounds without compromising accuracy is required. In this regard, machine-learning (ML) algorithms have shown excellent predictive accuracies along with short development and prediction times.^{34–38} Therefore, ML models have been used extensively to screen millions of molecules in materials science and drug discovery.^{39–43} ML models generally require a large amount of data for accurate predictions. When the quantity of data is limited, feature engineering is employed to generate the most informative features. These features are expected to capture the appropriate molecular information necessary to predict the target quantity. Feature engineering requires domain knowledge, relying on having access to experts.^{44–46} In small data sets, DFT-based or experimentally determined features have been used due to their high accuracy. However, some reports also explore simple features based on the molecular structure.^{47–52}

In this work, we investigated four ML models to predict the redox potentials of phenazine derivatives in the dimethoxyethane (DME) solvent. The training-set containing 151 phenazine derivatives was obtained from the previously reported DFT study having 189 phenazine derivatives with only one type of functional group per molecule (20 unique functional groups).²⁷ Molecular features were computed from the optimized neutral structures using the RDKit python library.⁵³ Model accuracy was improved through feature selection and hyperparameter optimization using the external validation set. Then, the model performance was assessed on the external test-set compiled from the literature consisting of new functional groups, multiple functional groups, and diverse structures. Their redox potential was computed using the DFT. The trained models were employed to predict the redox potentials of randomly generated phenazine derivatives with multiple functional groups. We also carried out feature importance analysis and discussed the structure–functional relationship of phenazine derivatives. Finally, promising candidates were identified for the anolyte from the external test-set and multiple functional group test-sets.

2. MATERIALS AND METHODS

2.1. Computational Details. The redox potentials of phenazine derivatives were computed using the DFT workflow described in the paper by Mavrandonakis et al.²⁷ All the DFT calculations were performed with Gaussian 09 software.⁵⁴

Table 1. Representative Structures from Training-Set/Internal Test-Set^a

Mol ID: 1	Mol ID: 3	Mol ID: 5
Mol ID: 48	Mol ID: 52	Mol ID: 172

^aMol IDs were assigned to identify derivatives from the corresponding data set.

Table 2. Representative Structures from the External Test-Set^a

Mol ID: 1	Mol ID: 3	Mol ID: 5
Mol ID: 15	Mol ID: 17	Mol ID: 28

^aMol IDs were assigned to identify derivatives from the corresponding data set.

Geometry optimization of neutral and reduced forms of phenazine and its derivatives were carried out in the gas phase by employing B3LYP/6-31+G(d,p) level of theory.^{55–58} Harmonic frequency analysis was performed for all the structures to confirm them as minima. Solvation effects of DME were incorporated during the single-point calculations using the M06-2X functional,⁵⁹ by employing the SMD solvation model (details in the Supporting Information).^{60,61} The term “Redox Potential” in this report corresponds to the “Reduction Potential” with respect to unsubstituted phenazine molecule (i.e., the parent phenazine). The redox potentials of phenazine derivatives were computed using the following equations:



$$E_1^0 = -\frac{\Delta G_{(\text{rxn},\text{sol})}}{nF} + E_{1(\text{ref})}^0 \quad (2)$$

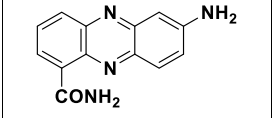
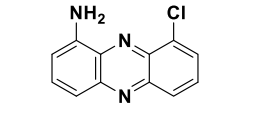
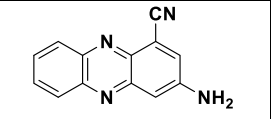
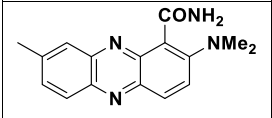
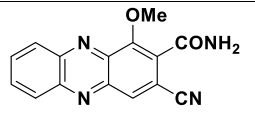
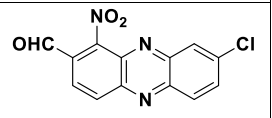
$$\Delta G_{(\text{rxn},\text{sol})} = G_{([\text{XPZ}]^-,\text{sol})}^0 + G_{([\text{PZ}],\text{sol})}^0 - G_{([\text{PZ}]^-,\text{sol})}^0 - G_{([\text{XPZ}],\text{sol})}^0 \quad (3)$$

$$G_{(\text{sol})}^0 = G_{(\text{therm},\text{gas})}^{(\text{B3LYP})} + E_{(\text{sol})}^{\text{M06-2X}} \quad (4)$$

where PZ symbolizes the parent phenazine, XPZ represents the substituted phenazine molecules, $E_{1(\text{ref})}^0$ is the reported redox potential of parent PZ,²⁷ $\Delta G_{(\text{rxn},\text{sol})}$ corresponds to the free energy change of the reaction, F is the Faraday constant, n is number of electrons involved in the reduction, and $G_{(\text{sol})}^0$ represents the final composite free energy of individual species, which was calculated by adding the free energy contribution computed at the B3LYP level of theory, $G_{(\text{therm},\text{gas})}^{(\text{B3LYP})}$, to the single-point energies calculated at the M06-2X level of theory: $E_{(\text{sol})}^{\text{M06-2X}}$.

2.2. Data Generation. **2.2.1. Training-Set and Internal Test-Test.** These data sets were obtained from work reported by Mavrandonakis and co-workers.²⁷ In their report, the redox potentials of 189 phenazine derivatives were computed using DFT in the DME solvent. These DFT redox potentials were used as a target property in this work during training and testing. 20 unique electron-withdrawing and electron-donating functional groups were present in the data set [–N(CH₃)₂, –NH₂, –OH, –OCH₃, –P(CH₃)₂, –SCH₃, –SH, –CH₃, –C₆H₅, –CH=CH₂, –F, –Cl, –CHO, –COCH₃, –CONH₂,

Table 3. Representative Structures from Multiple Functional Group Test-Sets^a

		
Mol ID: 19	Mol ID: 9	Mol ID: 7
		
Mol ID: 12	Mol ID: 14	Mol ID: 5

^aMol IDs were assigned to identify derivatives from the corresponding data set.

–COOCH₃, –COOH, –CF₃, –CN, and –NO₂]. It should be noted that phenazine derivatives in this data set contain only one type of functional group per molecule. The optimized 3D structures of derivatives in neutral and in anionic states were also provided. However, only neutral structures were used in this study. Unfortunately, not all compounds were supplied with their neutral structure, those compounds were modeled, and their optimized structures were added to the data set. Next, 208 different types of features were generated using the RDKit python library.⁵³ The list of all features is given in Table S1 of Supporting Information. The features were scaled using the “StandardScaler” class of the scikit-learn library,⁶² removing the mean and scaling each feature to unit variance. Finally, the whole data set was shuffled and split randomly into a training-set and test-set in an 8:2 ratio (151 samples in the training-set and 38 samples in the test-set). A few phenazine derivatives from the training-set/internal test-set are shown in Table 1.

2.2.2. External Test-Set. This data set was compiled from different reports studying various properties of phenazine derivatives.^{63–67} Their redox potentials were computed using DFT and used as a target property during testing. We gathered a total of 30 phenazine derivatives. Derivatives containing five or more substituted rings were removed. Also, derivatives having drastically different neutral and anion structures were removed. In the end, 22 diverse phenazine derivatives with multiple types of functional groups remained in the external test-set. Table 2 shows some of the structures from this data set. It can be seen that this data set contains unique and different structures compared to the training-set.

2.2.3. Multiple Functional Group Test-Sets. This data set contains two test-sets: (i) two functional group test-set and (ii) three functional group test-set. These test-sets were generated by randomly choosing the position and the type of the functional group from this list [–N(CH₃)₂, –NH₂, –OH, –OCH₃, –P(CH₃)₂, –SCH₃, –SH, –CH₃, –C₆H₅, –CH=CH₂, –F, –Cl, –CHO, –COCH₃, –CONH₂, –COOCH₃, –COOH, –CF₃, –CN, and –NO₂]. 20 derivatives having two different types of functional groups per molecule were generated for two functional group test-set. Similarly, 20 derivatives having three different types of functional groups per molecule were generated for three functional group test-set. Their redox potentials were computed using DFT and used as a target property during testing. Five derivatives from two and three functional group test-sets were removed to form an external validation set. Thus, the final size of two and three

functional group test-sets was reduced from 20 to 15. In this report, the term “multiple” refers to the derivatives containing different types and more than one functional group. Similarly, the terms “two functional groups” and “three functional groups” refer to the derivatives containing two different types of functional groups and three different types of functional groups per molecule, respectively. A few representative structures from these test-sets are shown in Table 3.

2.2.4. External Validation Set. An external validation set of 10 phenazine derivatives was compiled from two and three functional group test-sets. Five derivatives from two functional group test-set and five derivatives from three functional group test-set were selected. Their redox potentials were computed using DFT and used as a target property. This validation set does not come from the training-set. Therefore, it is termed as an external validation set. It was used for feature selection and hyperparameter optimization. External validation set improves generalization by transferring knowledge from the test-set to models through hyperparameters.

2.3. Hyperparameter Optimization. Hyperparameters of the models were optimized using the “GridSearchCV” class of the scikit-learn library.⁶² During hyperparameter optimization, models were trained on the training-set and evaluated on the external validation set. Mean squared error (MSE) was used as an evaluation metric for hyperparameter optimization. The grid of hyperparameters for each model is given in Table S2 of Supporting Information. The parameter grid was adjusted manually.

2.4. ML Models. Following four ML models were investigated in this study. These models were chosen due to their ability to generalize from small data sets. Models were implemented with the scikit-learn python library.⁶² First, models were trained on the training-set containing all 208 features, followed by hyperparameter optimization. Then, the models were re-trained on different subsets of features to identify the set of features having the highest average performance on the external validation set. Once the optimum features were identified, hyperparameter optimization was performed with the selected features to improve the model performance further.

2.4.1. Automatic Relevance Determination Regression (ARDR). This is a probabilistic model related to the sparse Bayesian learning (SBL) framework. It assumes axis-parallel, elliptical Gaussian distribution for each coefficient. The precision of each Gaussian distribution is drawn from the

prior distribution (gamma distribution); therefore, it can lead to sparser coefficients. Thus, it is an effective tool to remove irrelevant features.^{68,69}

2.4.2. Gaussian Process Regression (GP). It is a non-parametric Bayesian model. The nonparametric Bayesian model provides the probability distribution of parameters over all possible functions that fit the data. The prior in a Gaussian process is specified on the function space. Gaussian process prior is a multivariate normal distribution whose mean is obtained from the data, and covariance is specified using the kernel function. The hyperparameters of the kernel are optimized during the training.^{70,71} We used a combination of *WhiteKernel* and *RBF* kernel. *WhiteKernel* is used for specifying the noise level, and *RBF* kernel is a very popular kernel used in many algorithms.

2.4.3. Kernel Ridge Regression (KRR). It is the extension of ridge regression with kernel trick. In ridge regression, a linear model is leaned with the l2-norm regularization. Using the kernel trick, KRR learns a linear function in the high dimensional non-linear space without actually transforming the data.⁷²

2.4.4. Support Vector Regression (SVR). This model is the regression form of the support vector machine (SVM), a popular algorithm for classification tasks. Analogous to the SVM, SVR depends on the subset of training data and ignores the points whose prediction is close to their true value. SVM also utilizes the kernel trick and learns a hyperplane in the high dimensional space.⁷³

2.5. Evaluation Metrics. The following metrics were used for evaluating the model performance. In the formulas below, N denotes the number of data points, \hat{y}_i denotes the predicted value of i th sample, and the y_i denotes the corresponding true value.

- Coefficient of determination (R^2):

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}$$

where

$$\bar{y} = \frac{\sum_{i=1}^N y_i}{N}$$

- Mean Squared Error (MSE):

$$MSE = \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}$$

- Mean Absolute Error (MAE):

$$MAE = \frac{\sum_{i=1}^N |y_i - \hat{y}_i|}{N}$$

The use of terms “Accuracy” and “Performance” in this report is contextual and refers to one or more metrics defined above.

2.6. Feature Selection. As the number of features obtained from the RDKit library was more than the size of training-set, it was necessary to implement a feature selection strategy. It has been observed that the training-set containing more features than data points leads to overfitting.³⁰ Feature

selection was implemented using the “*SelectKBest*” class of the scikit-learn library.⁷⁴ The parameter “ k ” of the “*SelectKBest*” class was obtained by evaluating the average performance of models on the external validation set at different values of “ k .” First, models were trained on the training-set containing all features, followed by hyperparameter optimization. Then, the models were re-trained on the subsets of features selected using “*SelectKBest*” class at different values of “ k .” These values for “ k ” were tested: 50, 75, 100, 125, 150, and 208. The average model performance at different values of “ k ” on the external validation set is shown in Table 4. It can be seen that

Table 4. Average Model Performance on External Validation Set at Different Values of “ k ”

performance metric	values of “ k ”					
	50	75	100	125	150	208
R^2	0.45	0.42	0.57	0.55	0.54	0.54
MSE	0.02	0.02	0.02	0.02	0.02	0.02
MAE	0.12	0.12	0.10	0.10	0.10	0.10

the models trained on 100 selected features show the highest average performance in terms of R^2 . Therefore, these 100 features were selected for the subsequent analysis. The models trained on 100 selected features were further improved through hyperparameter optimization.

2.7. Feature Importance Analysis. Feature importance analysis was performed using the technique known as permutation importance. In this technique, values of the feature to be assessed are randomly shuffled (permuted). Then, prediction accuracy is computed on the shuffled data set. Shuffling feature values is equivalent to replacing the feature with noise, thereby removing its information from the data set. Therefore, the model is expected to perform poorly on the shuffled data set if the feature is important. The degree of importance depends on the amount of variation in the accuracy. This technique does not re-train the model; therefore, a trained model is required. The permutation importance was computed using “*permutation_importance*” class of the scikit-learn library and the training-set.⁷⁵ This procedure was repeated 100 times to obtain reliable estimates. The feature importance scores were rescaled between 0 to 1. The mean and standard deviation of the feature scores were reported. The mean feature score was used for the ranking of individual features. The terms “Feature” and “Descriptor” are used interchangeably in this report.

3. RESULTS AND DISCUSSION

3.1. Test-Set Performance. We assessed the generalizability of the trained models (i.e., performance on the unseen data) using internal and external test-sets. Please refer to Section 2 for the preparation of internal and external test-sets. As the internal test-set comes from the same source, it is very similar to the training-set and contains derivatives with only one type of functional group per molecule. However, the external test-set is compiled from multiple sources, therefore, it has very diverse phenazine derivatives with different types of functional groups. It also contains functional groups and structures not present in the training-set (e.g., $-NHP$, $-Br$, and extended conjugation). Figure 2 shows the performance on the internal test-set, and Figure 3 shows performance on the external test-set. It can be seen that all models have excellent accuracy on the internal test-set ($R^2 > 0.98$) and high accuracy

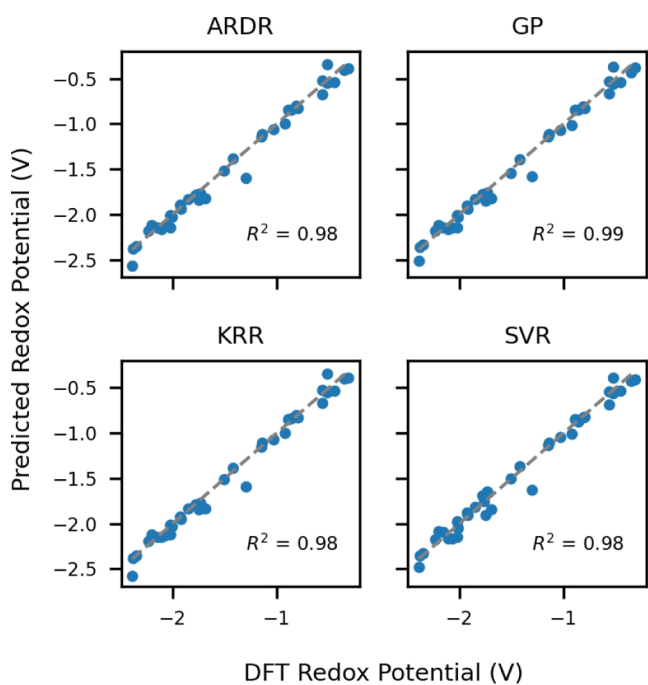


Figure 2. Plots showing ML predictions on internal test-set (*y*-axis) vs DFT redox potentials (*x*-axis). Gray dashed line corresponds to the perfect predictions.

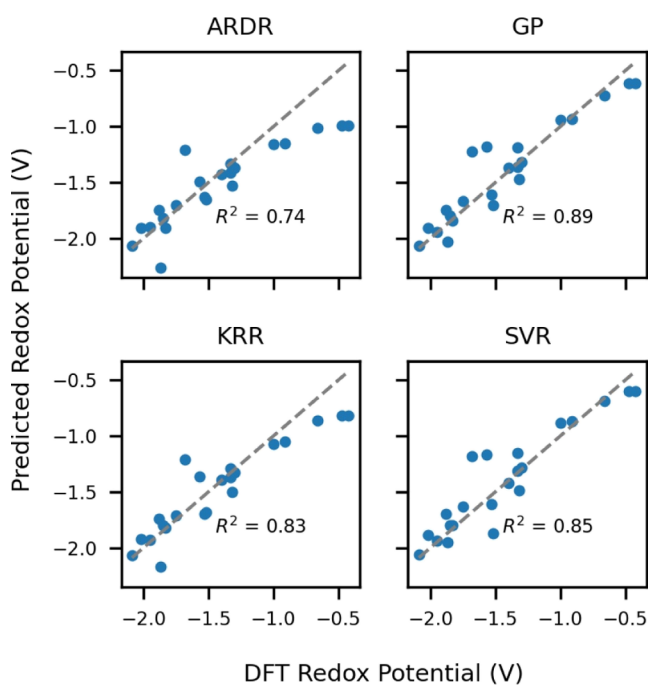


Figure 3. Plots showing ML predictions on external test-set (*y*-axis) vs DFT redox potentials (*x*-axis). Gray dashed line corresponds to the perfect predictions.

on the external test-set set ($R^2 > 0.74$). The GP model achieved the highest R^2 of 0.89 on the external test-set. After deep analysis in Section 3.3, it was revealed that GP is not a stable model, whereas relatively low-performing models KRR ($R^2 = 0.83$) and SVR ($R^2 = 0.85$) are more stable. Therefore, one should be careful while using the high-performing model, and the stability of the model should also be considered. The values of performance metrics on internal and external tests are

shown in Table 5. Such a performance on the external test-set is surprising as models were trained on the phenazine

Table 5. Values of Performance Metrics on Internal and External Test-Sets^a

Model name	Internal test-set			External test-set		
	R^2	MSE	MAE	R^2	MSE	MAE
ARDR	0.98	0.01	0.06	0.74	0.06	0.18
GP	0.99	0.01	0.05	0.89	0.03	0.11
KRR	0.98	0.01	0.05	0.83	0.04	0.14
SVR	0.98	0.01	0.07	0.85	0.03	0.13

^aNumbers were rounded up to two decimals.

derivatives having only one type of functional group. These results show that ML models are capable of generalizing from a very small and simple data set.

3.2. Prediction on Multiple Functional Group Test-Sets. Next, we assessed the model performance on the phenazine derivatives substituted with different types of functional groups per molecule. These test-sets were generated randomly; please refer to Section 2 for the generation of this data set. Figures 4 and 5 show the performance on the

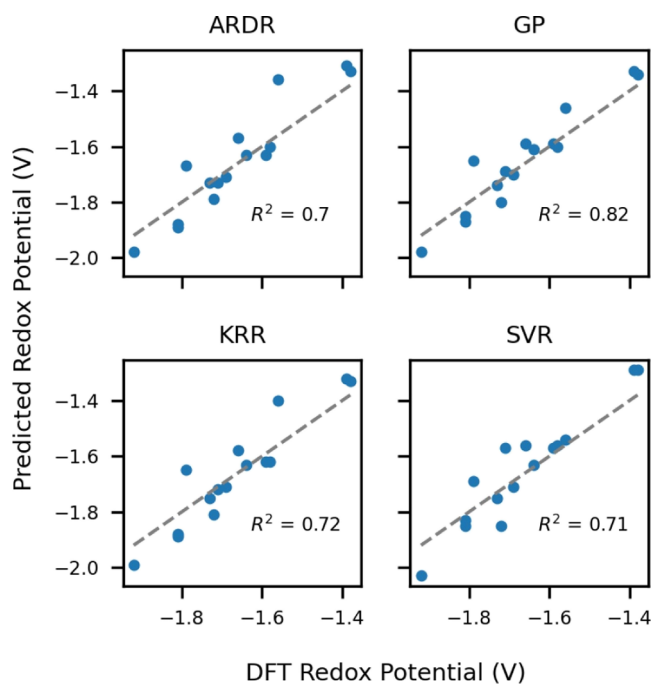


Figure 4. Plots showing ML predictions on two functional group test-set (*y*-axis) vs DFT redox potentials (*x*-axis). Gray dashed line corresponds to the perfect predictions.

derivatives containing two and three different functional groups, respectively. It can be seen that the models performed reasonably well ($R^2 > 0.7$) even though molecules used for the training had only one type of functional group per molecule. In particular, GP model achieved the highest performance of $R^2 = 0.82$ on two functional groups test-set. However, automatic relevance determination regression (ARDR) achieved the highest performance of $R^2 = 0.82$ on three functional groups test-set. A deeper analysis of GP and ARDR in Section 3.3 suggests that GP and ARDR are not very reliable models. Although KRR and SVR have relatively low performance, they

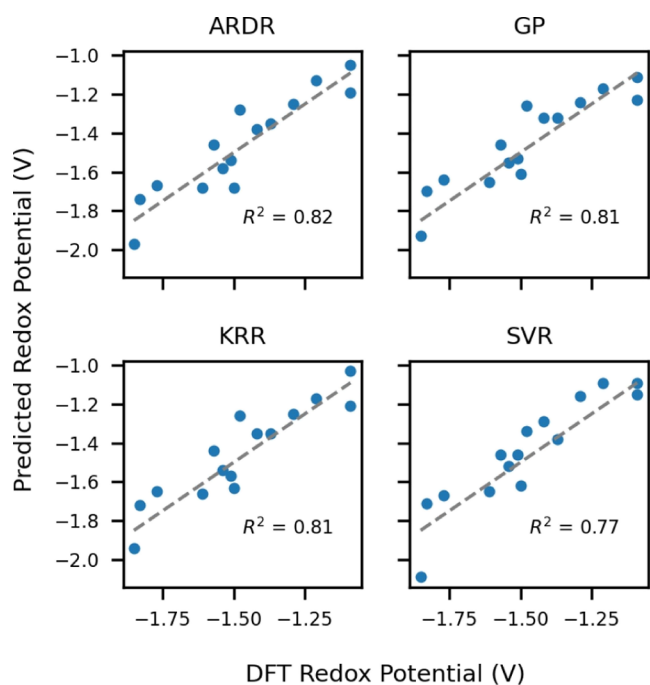


Figure 5. Plots showing ML predictions on three functional group test-set (*y*-axis) vs DFT redox potentials (*x*-axis). Gray dashed line corresponds to the perfect predictions.

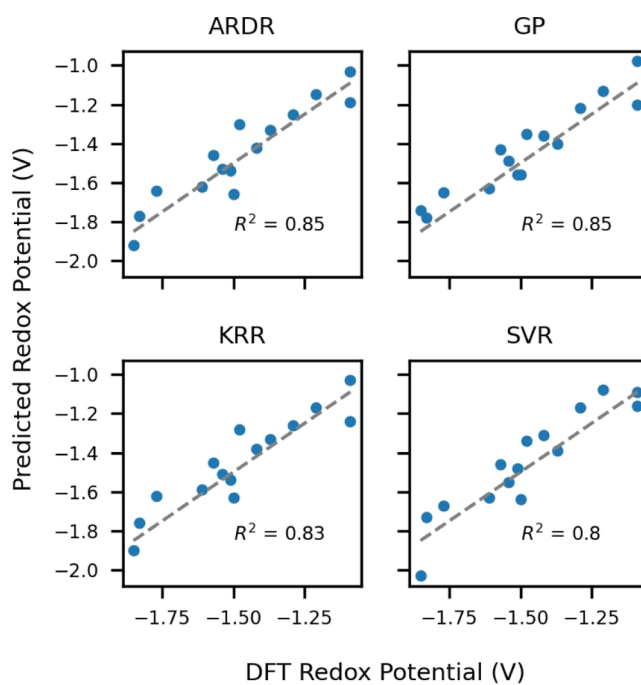


Figure 6. Plots showing ML predictions on three functional group test-set (*y*-axis) vs DFT redox potentials (*x*-axis). The combined data set (training-set + two functional group test-set) was used for the training. Gray dashed line corresponds to the perfect predictions.

are more reliable. Therefore, one should be careful while using a high-performing model, and the model's reliability and stability should also be considered. Nevertheless, these results again show the surprising generalization power of ML models.

Furthermore, we added all 15 derivatives from two functional group test-set to the training-set and re-trained the models on this new data set of 166 derivatives. The predictive performance of this combined data set was assessed on the same data set of 15 derivatives containing three different types of functional groups. The results of this analysis are shown in Figure 6. It can be seen that the model performance has improved with the addition of more data in the training-set.

3.3. Feature Importance Analysis. We carried out feature importance analysis using permutation importance. Please refer to Section 2 for the details on the technique. In order to understand how model performance changes with the number of descriptors, we re-trained the models on the subset of features and assessed their performance on the internal test-set. Top 50 features based on their permutation importance score were used. R^2 was used as a performance metric. The result of this analysis is shown in Figure 7. It can be seen that most of the models show a jump in the R^2 and have $R^2 > 0.9$ around the top 10 features. The unusual behavior of the GP model is attributed to the instability of the model for a small number of features. The plots in Figure 8 show the histograms of the top 10 important features from each model. Although models show variation in feature importance, they all agree in terms of the most important feature that is, "PEOE_VSA1." Interestingly, most of the features in ARDR have small weights as ARDR tries to prune the large number of irrelevant features, leading to a sparse model.^{69,76} Five out of 10 features—"MaxAbsPartialCharge," "PEOE_VSA1," "fr_ArN," "fr_NH0," and "fr_NH2" are common to all models. Variations in the feature importance scores could be attributed to the difference

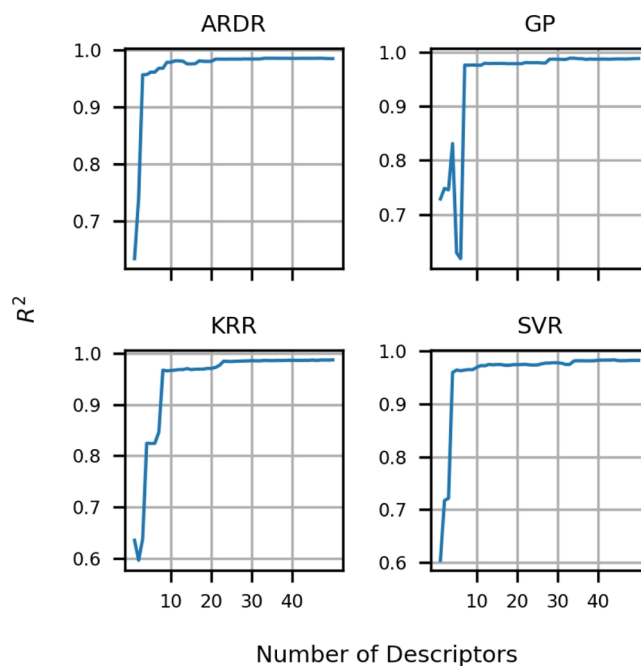


Figure 7. R^2 vs number of descriptors. R^2 was computed using the internal test-set. In this study, we identified a few issues with ARDR and GP. Despite high predictive performance, ARDR is not a reliable model as it places very high weight on one feature (i.e., "PEOE_VSA1"). Similarly, GP is not a reliable model as it becomes unstable when the small number of features are used. We encountered divided by zero errors in the kernel function during the analysis with the GP model.

in the internal structures of the models. Here, we discuss some of the common features from Figure 8.

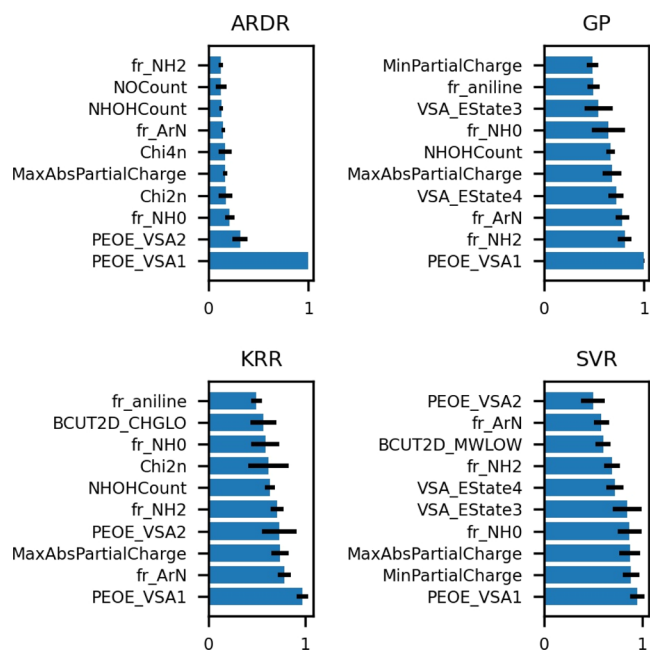


Figure 8. Top 10 features (*y*-axis) vs mean feature importance score (*x*-axis). Feature importance scores were rescaled between 0 to 1. Error bars represent standard deviation from 100 repetitions.

3.3.1. PEOE_VSA1. This is the sum of the approximate accessible van der Waals surface area (i.e., VSA in \AA^2) of the atoms having partial charge less than -0.30 .^{77–79} The partial charges are computed using the partial equalization of orbital electronegativities (PEOE) method developed by Gasteiger and Marsili in 1980. Please refer to the discussion of *MaxAbsPartialCharge* descriptor for the PEOE method. Thus, this descriptor captures the information related to molecular size and the number of electron-donating functional groups.

3.3.2. MaxAbsPartialCharge. This is the maximum value of the absolute Gasteiger partial charges present in the molecule. In 1980, Gasteiger and Marsili gave the procedure to calculate the partial charges in a molecule. That procedure is known as PEOE. In this method, the charge is transferred between

bonded atoms until equilibrium. Gasteiger partial charges depend on the connectivity and orbital electronegativity, thus capturing the electron-donating and electron-withdrawing power of the atoms.⁸⁰ Electronegativity is essential information as electron-donating groups decrease the redox potential, and electron-withdrawing groups increase the redox potential.²⁷

3.3.3. MinPartialCharge. This is the minimum value of the Gasteiger partial charges present in the molecule. Please refer to the discussion of *MaxAbsPartialCharge* descriptor for the properties of Gasteiger partial charges.

3.3.4. fr_NH0. It is the number of tertiary amines present in the molecule.

3.3.5. fr_ArN. It is the number of *N* functional groups attached to aromatic rings.

3.3.6. fr_NH2. It is the number of primary amines.

3.3.7. NHOHCount. It is the number of N–H and O–H bonds present in the molecule.

From the analysis in this section, we realized that there are some issues with the ARDR and GP which are outlined below. One should be very careful while using ARDR and GP models.

3.3.8. Issues with the ARDR Model. As ARDR is related to the SBL framework, it reduces the number of irrelevant features. Unfortunately, in this case, ARDR has put a lot of weight on only one feature, that is, “PEOE_VSA1” (Figure 8). Surprisingly, ARDR also archives an accuracy of more than $0.95R^2$ only with the two features (Figure 7). Although it has shown good performance on the data set used in this work, it may not work for the broad chemical space. This type of behavior reduces the reliability of the model.

3.3.9. Issues with the GP Model. From Figure 7, it can be seen that the model’s accuracy decreases with more features, and at around 10 features, there is a significant drop in the performance. We also encountered divided by zero errors in the kernel function during this analysis. This shows that GP may not be a very stable model in this case.

3.4. Structure–Functional Relationship. “PEOE_VSA1” is the most important descriptor common to all models. It is computed by summing over the approximate accessible VSA (i.e., in \AA^2) of the atoms having partial charge less than -0.30 .^{77–79} Thus, the “PEOE_VSA1” descriptor captures the information related to molecular size and the number of

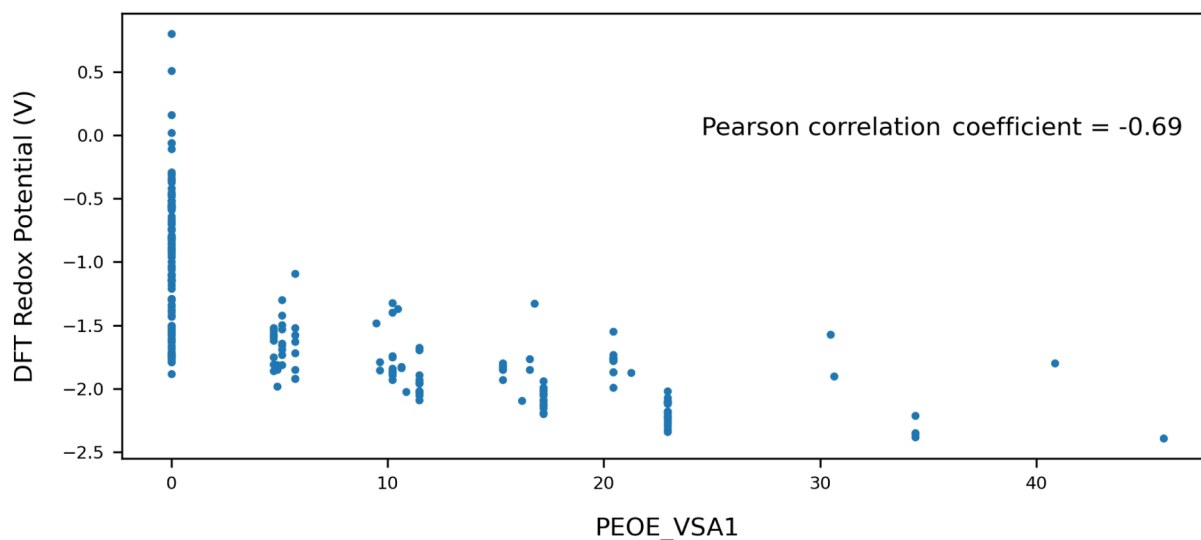


Figure 9. Redox potential vs “PEOE_VSA1.”

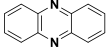
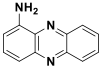
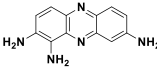
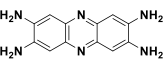
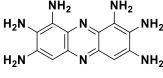
				
Mol ID: 0	Mol ID: 3	Mol ID: 105	Mol ID: 134	Mol ID: 136
PEOE_VSA1: 0	PEOE_VSA1: 5.73	PEOE_VSA1: 17.20	PEOE_VSA1: 22.93	PEOE_VSA1: 34.40
Potential: -1.74	Potential: -1.85	Potential: -2.09	Potential: -2.32	Potential: -2.36
PEOE_VSA1 increases		→		
Delocalization increases		→		
Redox Potential decreases		→		

Figure 10. Examples from the training-set showing the effect of charge delocalization on “PEOE_VSA1.” Values of “PEOE_VSA1” and DFT redox potentials in volts are also shown. Mol IDs were assigned to identify derivatives from the corresponding data set.

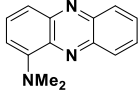
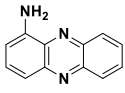
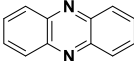
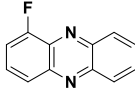
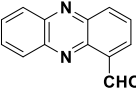
				
Mol ID: 1	Mol ID: 3	Mol ID: 0	Mol ID: 21	Mol ID: 26
Potential: -1.85	Potential: -1.85	Potential: -1.74	Potential: -1.63	Potential: -1.50
Shift: -0.11	Shift: -0.11	Shift: 0	Shift: 0.11	Shift: 0.24
Electron-donating groups			Electron-withdrawing groups	

Figure 11. Examples showing positive and negative shifts with respect to parent phenazine. DFT redox potentials and shifts in volts are also shown. Mol IDs were assigned to identify derivatives from the corresponding data set.

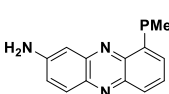
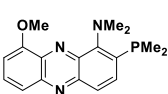
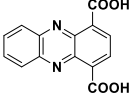
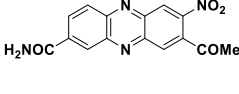
			
Mol ID: 9	Mol ID: 8	Mol ID: 7	Mol ID: 15
Potential: -1.92	Potential: -1.85	Potential: -1.40	Potential: -1.09
Shift: -0.18	Shift: -0.11	Shift: 0.34	Shift: 0.65
Electron-donating groups		Electron-withdrawing groups	

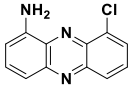
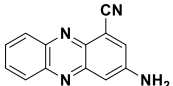
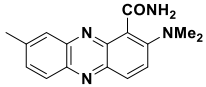
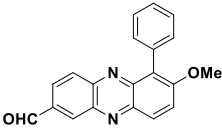
Figure 12. Examples showing the effect of similar type of functional groups on the redox potential. DFT redox potentials and shifts in volts are also shown. Mol IDs were assigned to identify derivatives from the corresponding data set.

electron-donating functional groups present in the molecule. From Figure 9, we can see that the redox potential of phenazine derivatives decreases with the increasing value of “PEOE_VSA1.” The Pearson correlation coefficient between “PEOE_VSA1” and redox potential is -0.69 , supporting the observation. We observed that the value of “PEOE_VSA1” is higher for the systems having delocalization of negative partial charge. The delocalized system contains more atoms with the negative partial charge than the corresponding localized system. Thus, the number of atoms contributing to “PEOE_VSA1” in delocalized systems is higher than localized ones. The effect of delocalization of partial charge on “PEOE_VSA1” is shown in Figure 10 with a few examples from the training-set. Thus, for designing better anolytes, it is suggested to increase the delocalization of negative partial charge in the phenazine derivatives.

The redox potential of phenazine derivative depends on the type of functional group, the position of attachment, and the number of functional groups. Two types of functional groups have been investigated in this study: (i) electron-donating, and

(ii) electron-withdrawing. The redox potential of parent phenazine without any functional group is -1.74 V. When the redox potential of the derivative decreases (i.e., less than -1.74 V) after the attachment of functional groups, then it is called a negative shift. Similarly, if it increases, it is called a positive shift. The shift is quantified as the difference between the redox potential of phenazine derivative and parent phenazine. After sorting phenazine derivatives based on the redox potential, it was observed that electron-donating groups show a negative shift, whereas electron-withdrawing groups show a positive shift. Thus, the shift corresponding to electron-donating groups is negative and electron-withdrawing groups is positive. The redox potentials of phenazine derivatives were computed using the approach discussed in Section 2.1. Equation 2 shows that functional groups stabilizing the anionic form of phenazine derivatives have high redox potential. In contrast, those that destabilize the anionic form have low redox potential. Therefore, electron-withdrawing groups show a positive shift as they stabilize the anionic form, and electron-donating groups show a negative shift as they destabilize the

Table 6. Examples Showing the Effect of Absolute Values of Single Functional Group Shift on the Redox Potential of Derivatives Containing Different Types of Functional Groups^a

	Phenazine Derivative	Details of the phenazine derivative	Redox potential of the corresponding single functional group derivative	Shift of the corresponding single functional group derivative
A.		Mol ID: 1 Potential: -1.71 Shift: 0.03	-NH ₂ : -1.85 -Cl: -1.61	-NH ₂ : -0.11 -Cl: 0.13
B.		Mol ID: 7 Potential: -1.58 Shift: 0.16	-NH ₂ : -1.92 -CN: -1.42	-NH ₂ : -0.18 -CN: 0.32
C.		Mol ID: 12 Potential: -1.83 Shift: -0.09	-N(CH ₃) ₂ : -1.98 -CH ₃ : -1.79 -CONH ₂ : -1.52	-N(CH ₃) ₂ : -0.24 -CH ₃ : -0.05 -CONH ₂ : 0.22
D.		Mol ID: 4 Potential: -1.54 Shift: 0.20	-OCH ₃ : -1.86 -C ₆ H ₅ : -1.76 -CHO: -1.51	-OCH ₃ : -0.12 -C ₆ H ₅ : -0.02 -CHO: 0.23

^aDFT redox potentials and shifts in volts are also shown. Mol IDs were assigned to identify derivatives from the corresponding data set.

anionic form. A few examples showing positive and negative shifts with respect to parent phenazine are shown in Figure 11.

In the case of derivatives with multiple functional groups, if all groups are similar, then shift also corresponds to their type. For example, when the derivative contains all electron-donating groups, it shows a negative shift. Similarly, the shift is positive when the derivative contains all electron-withdrawing groups. A few examples having similar types of functional groups are shown in Figure 12.

When derivatives contain more than one functional group that differ in their type, the shift is determined by the group showing the highest absolute shift in the corresponding single functional group derivative. For example, derivative A in Table 6 contains -NH₂, an electron-donating group which has a shift of -0.11 V and -Cl, an electron-withdrawing group which has a shift of 0.13 V. The absolute of the shift for -Cl is more than -NH₂; therefore, derivative A shows a positive shift of 0.03 V, supporting our claim. A similar analysis is applicable to derivative B, which also shows a positive shift. Derivative C contains -N(CH₃)₂ and -CH₃, two electron-donating groups, and -CO(NH₂), an electron-withdrawing group. An absolute shift of -N(CH₃)₂ is -0.24 V, which is the highest among all three groups. Therefore, derivative C shows a negative shift of -0.09 V. Derivative D contains -OCH₃ and -C₆H₅, two electron-donating groups, and -CHO, one electron-withdrawing group. However, derivative D shows a positive shift as the absolute shift of -CHO is more than both electron-donating groups. Thus, the redox potential of phenazine derivatives containing multiple functional groups is determined by the relative strength of electron-donating or electron-withdrawing power of the functional groups.

The effect of position on the redox potential of single functional group derivatives has been studied by Mavrandonakis and co-workers.²⁷ They showed that the position does not have a significant effect for electron-withdrawing groups. However, electron-donating groups which are capable of intramolecular hydrogen bonding show more negative shift when

attached at position 2 compared to position 1. The position numbers in phenazine derivatives are shown in Figure 13. They

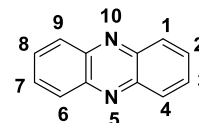
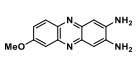
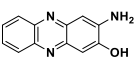
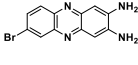
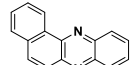
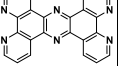
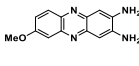
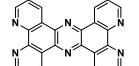
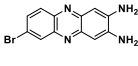
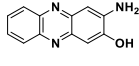
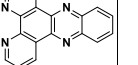
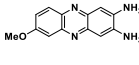
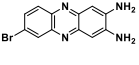
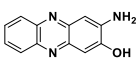
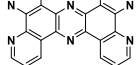
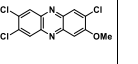


Figure 13. Numbering of the positions in phenazine derivatives.

also investigated the effect of the number of functional groups on redox potential. It was shown that the addition of more electron-withdrawing groups shifts the redox potential continuously toward positive values. However, this effect is less significant for electron-donating groups. The difference between the phenazine derivative with four amino groups and eight amino groups is very small (~0.05 V). The difference between phenazine derivative with four cyano groups and eight cyano groups is ~1.23 V.

3.5. Identification of Promising Phenazine Derivatives for the Anolyte. In this section, we identify the top five promising candidates for the anolyte using the trained ML models. Models developed in this study are based on features that do not require electronic structure calculations. Therefore, these models could screen millions of molecules in a significantly small amount of time. Then, experimentation or DFT calculations could be performed on the reduced number of molecules to identify the best redox-active molecules, saving computational and experimental costs. Using this hybrid DFT-ML approach, we have identified promising phenazine derivatives for the anolyte in RFBs. These promising candidates would provide a good starting point for the experimentalists. Electron-donating molecules with negative redox potential are preferred candidates for the anolyte. As KRR and SVR are stable models, the predictions here are based on them. The values of redox potentials are averaged over 100 independent iterations of data splitting and model training. Table 7 lists the top five phenazine derivatives from

Table 7. Top Five Anolyte Candidates Predicted Using DFT, KRR, and SVR from the External Test-Set^a

DFT	 Mol ID: 13 DFT: -2.09	 Mol ID: 29 DFT: -2.02	 Mol ID: 12 DFT: -1.95	 Mol ID: 1 DFT: -1.88	 Mol ID: 5 DFT: -1.87
KRR	 Mol ID: 13 ML: -2.09 DFT: -2.09	 Mol ID: 5 ML: -2.09 DFT: -1.87	 Mol ID: 12 ML: -1.98 DFT: -1.95	 Mol ID: 29 ML: -1.95 DFT: -2.02	 Mol ID: 4 ML: -1.78 DFT: -1.83
SVR	 Mol ID: 13 ML: -2.06 DFT: -2.09	 Mol ID: 12 ML: -1.96 DFT: -1.95	 Mol ID: 29 ML: -1.91 DFT: -2.02	 Mol ID: 5 ML: -1.89 DFT: -1.87	 Mol ID: 3 ML: -1.81 DFT: -1.52

^aSVR and KRR were trained on the phenazine derivatives containing single type of functional group per derivative. Mol IDs and redox potentials predicted from DFT and ML models are shown below the respective candidates. Mol IDs were assigned to identify derivatives from the corresponding test-set. Derivatives are arranged in increasing order of redox potential. Redox potentials are given in the unit of volts.

the external test-set with the most negative redox potentials obtained from DFT and two ML models. Four out of five predictions from KRR and SVR match with DFT predictions. The top five promising candidates from multiple functional groups test-sets are shown in Tables S3–S5 [Supporting Information](#).

4. CONCLUSIONS

In this study, four ML models were employed to predict the redox potentials of phenazine derivatives in DME using DFT. Models were trained on a small data set of 151 phenazine derivatives having only one type of functional group per molecule (20 unique functional groups). The trained models achieved high accuracies ($R^2 > 0.74$) on internal and external test-sets containing diverse phenazine derivatives. We also showed that despite being trained on derivatives with a single type of functional groups, models were able to predict the redox potentials of the derivatives containing multiple and different types of functional groups with good accuracies ($R^2 > 0.7$). Feature selection and hyperparameter optimization using the validation set were critical strategies for performance improvement. Feature selection removed the unnecessary and noisy features. Hyperparameter optimization using an external validation set helped improve the generalizability of the models. The addition of 15 derivatives from two functional group test-sets in the training-set improved the accuracy on three functional group test-sets. It was observed that the “PEOE_VSA1” descriptor was the most important molecular feature as it contains information related to molecular size and the partial charges. Deeper analysis showed that one should not rely only on the model performance but also investigate the stability and reliability of the models. Through the structure–functional relationship, we observed that the redox potential of derivatives containing multiple functional groups is

influenced by the functional group having either strong electron-donating or strong electron-withdrawing power. Models developed in this study are based on features that do not require electronic structure calculations. Therefore, these models could screen millions of molecules in a significantly small amount of time. Then, experimentation or DFT calculations could be performed on the screened candidates to identify the best molecules, saving computational and experimental costs. Using this hybrid DFT-ML approach, we have identified promising phenazine derivatives for the anolyte in RFBs. These promising candidates would provide a good starting point for the experimentalists. This study shows that it is possible to develop reasonably accurate ML models for complex quantities such as redox potential using small and simple data sets.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsomega.1c06856>.

Additional Computation details; list of all features; and parameter grids used during hyperparameter optimization; top five candidates for the anolyte from multiple functional group test-sets (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Siddharth Ghule – *Physical and Materials Chemistry Division, CSIR-National Chemical Laboratory (CSIR-NCL), Pune 411008, India; Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, India; orcid.org/0000-0003-0864-0777; Phone: +91-20-25903095; Email: ss.ghule@ncl.res.in*

Kavita Joshi – Physical and Materials Chemistry Division, CSIR-National Chemical Laboratory (CSIR-NCL), Pune 411008, India; Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, India; Phone: +91-20-25902476; Email: kjoshi@ncl.res.in

Kumar Vanka – Physical and Materials Chemistry Division, CSIR-National Chemical Laboratory (CSIR-NCL), Pune 411008, India; Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, India; orcid.org/0000-0001-7301-7573; Phone: +91-20-25903095; Email: k.vanka@ncl.res.in

Authors

Soumya Ranjan Dash – Physical and Materials Chemistry Division, CSIR-National Chemical Laboratory (CSIR-NCL), Pune 411008, India; Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, India; orcid.org/0000-0001-7267-6104

Sayan Bagchi – Physical and Materials Chemistry Division, CSIR-National Chemical Laboratory (CSIR-NCL), Pune 411008, India; Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, India; orcid.org/0000-0001-6932-3113

Complete contact information is available at: <https://pubs.acs.org/10.1021/acsomega.1c06856>

Author Contributions

S.G. and S.D. carried out the research work and written the manuscript with the guidance and supervision of K.V., K.J., and S.B.

Notes

The authors declare no competing financial interest. The MOL files of all phenazine derivatives and compiled DFT data are available at https://github.com/siddharth-ncl-work/ml_redox_potential-DATA.git.

ACKNOWLEDGMENTS

K.V. is grateful to the Department of Science and Technology (DST) (EMR/2014/000013) for providing financial assistance. S.B. thanks CSIR-NCL and SERB, India (SR/S2/RJN-142/2012 and EMR/2016/000576), for financial support. S.G. acknowledges Council of Scientific and Industrial Research (CSIR) for providing Research Fellowship. S.D. acknowledges CSIR-NCL (MLP101026) for providing a fellowship. The support and the resources provided by the “PARAM Brahma Facility” under the National Supercomputing Mission, Government of India at the Indian Institute of Science Education and Research (IISER) Pune are gratefully acknowledged.

REFERENCES

- (1) Shafiee, S.; Topal, E. When Will Fossil Fuel Reserves Be Diminished? *Energy Policy* **2009**, *37*, 181–189.
- (2) Dehghani-Sanij, A. R.; Tharumalingam, E.; Dusseault, M. B.; Fraser, R. Study of Energy Storage Systems and Environmental Challenges of Batteries. *Renewable Sustainable Energy Rev.* **2019**, *104*, 192–208.
- (3) Höök, M.; Tang, X. Depletion of Fossil Fuels and Anthropogenic Climate Change—A Review. *Energy Policy* **2013**, *52*, 797–809.
- (4) Gür, T. M. Review of Electrical Energy Storage Technologies, Materials and Systems: Challenges and Prospects for Large-Scale Grid Storage. *Energy Environ. Sci.* **2018**, *11*, 2696–2767.
- (5) Chu, W.-S.; Chun, D.-M.; Ahn, S.-H. Research Advancement of Green Technologies. *Int. J. Precis. Eng. Manuf.* **2014**, *15*, 973–977.

- (6) Balat, H. Green Power for a Sustainable Future. *Energy Explor. Exploit.* **2007**, *25*, 1–25.
- (7) Demirbas, A. Electrical Power Production Facilities from Green Energy Sources. *Energy Sources, Part B* **2006**, *1*, 291–301.
- (8) Dunn, B.; Kamath, H.; Tarascon, J.-M. Electrical Energy Storage for the Grid: A Battery of Choices. *Science* **2011**, *334*, 928–935.
- (9) Chung, E. What Caused the Deadly Power Outages in Texas and How Canada’s Grid Compares. CBC News. 2021. <https://www.cbc.ca/news/technology/power-outages-texas-canada-1.5920833> (accessed March 30, 2021).
- (10) Larcher, D.; Tarascon, J.-M. Towards Greener and More Sustainable Batteries for Electrical Energy Storage. *Nat. Chem.* **2015**, *7*, 19–29.
- (11) Koochi-Fayegh, S.; Rosen, M. A. A Review of Energy Storage Types, Applications and Recent Developments. *J. Energy Storage* **2020**, *27*, 101047.
- (12) Deng, D. Li-ion Batteries: Basics, Progress, and Challenges. *Energy Sci. Eng.* **2015**, *3*, 385–418.
- (13) Skyllas-Kazacos, M.; Chakrabarti, M. H.; Hajimolana, S. A.; Mjalli, F. S.; Saleem, M. Progress in Flow Battery Research and Development. *J. Electrochem. Soc.* **2011**, *158*, R55.
- (14) Leung, P.; Li, X.; Ponce de León, C.; Berlouis, L.; Low, C. T. J.; Walsh, F. C. Progress in Redox Flow Batteries, Remaining Challenges and Their Applications in Energy Storage. *RSC Adv.* **2012**, *2*, 10125–10156.
- (15) Sánchez-Díez, E.; Ventosa, E.; Guarnieri, M.; Trovò, A.; Flox, C.; Marcilla, R.; Soavi, F.; Mazur, P.; Aranzabe, E.; Ferret, R. Redox Flow Batteries: Status and Perspective towards Sustainable Stationary Energy Storage. *J. Power Sources* **2021**, *481*, 228804.
- (16) Ha, S.; Gallagher, K. G. Estimating the System Price of Redox Flow Batteries for Grid Storage. *J. Power Sources* **2015**, *296*, 122–132.
- (17) Whitehead, A. H.; Rabbow, T. J.; Trampert, M.; Pokorny, P. Critical Safety Features of the Vanadium Redox Flow Battery. *J. Power Sources* **2017**, *351*, 1–7.
- (18) Chen, Y.; Kang, Y.; Zhao, Y.; Wang, L.; Liu, J.; Li, Y.; Liang, Z.; He, X.; Li, X.; Tavajohi, N.; Li, B. A Review of Lithium-Ion Battery Safety Concerns: The Issues, Strategies, and Testing Standards. *J. Energy Chem.* **2021**, *59*, 83–99.
- (19) Díaz-Ramírez, M. C.; Ferreira, V. J.; García-Armingol, T.; López-Sabirón, A. M.; Ferreira, G. Environmental Assessment of Electrochemical Energy Storage Device Manufacturing to Identify Drivers for Attaining Goals of Sustainable Materials 4.0. *Sustain* **2020**, *12*, 342.
- (20) da Silva Lima, L.; Quartier, M.; Buchmayr, A.; Sanjuan-Delmás, D.; Laget, H.; Corbisier, D.; Mertens, J.; Dewulf, J. Life Cycle Assessment of Lithium-Ion Batteries and Vanadium Redox Flow Batteries-Based Renewable Energy Storage Systems. *Sustain. Energy Technol. Assess.* **2021**, *46*, 101286.
- (21) Kear, G.; Shah, A. A.; Walsh, F. C. Development of the All-Vanadium Redox Flow Battery for Energy Storage: A Review of Technological, Financial and Policy Aspects. *Int. J. Energy Res.* **2012**, *36*, 1105–1120.
- (22) Alotto, P.; Guarnieri, M.; Moro, F. Redox Flow Batteries for the Storage of Renewable Energy: A Review. *Renewable Sustainable Energy Rev.* **2014**, *29*, 325–335.
- (23) Weber, A. Z.; Mench, M. M.; Meyers, J. P.; Ross, P. N.; Gostick, J. T.; Liu, Q. Redox Flow Batteries: A Review. *J. Appl. Electrochem.* **2011**, *41*, 1137–1164.
- (24) Qi, Z.; Koenig, G. M. Review Article: Flow Battery Systems with Solid Electroactive Materials. *J. Vac. Sci. Technol., B: Nanotechnol. Microelectron.: Mater., Process., Phenom.* **2017**, *35*, 040801.
- (25) Sánchez-Díez, E.; Ventosa, E.; Guarnieri, M.; Trovò, A.; Flox, C.; Marcilla, R.; Soavi, F.; Mazur, P.; Aranzabe, E.; Ferret, R. Redox Flow Batteries: Status and Perspective towards Sustainable Stationary Energy Storage. *J. Power Sources* **2021**, *481*, 228804.
- (26) Xi, J.; Xiao, S.; Yu, L.; Wu, L.; Liu, L.; Qiu, X. Broad Temperature Adaptability of Vanadium Redox Flow Battery—Part 2: Cell Research. *Electrochim. Acta* **2016**, *191*, 695–704.

- (27) De La Cruz, C.; Molina, A.; Patil, N.; Ventosa, E.; Marcilla, R.; Mavrandonakis, A. New Insights into Phenazine-Based Organic Redox Flow Batteries by Using High-Throughput DFT Modelling. *Sustainable Energy Fuels* **2020**, *4*, 5513–5521.
- (28) Gentil, S.; Reynard, D.; Girault, H. H. Aqueous Organic and Redox-Mediated Redox Flow Batteries: A Review. *Curr. Opin. Electrochem.* **2020**, *21*, 7–13.
- (29) Leung, P.; Shah, A. A.; Sanz, L.; Flox, C.; Morante, J. R.; Xu, Q.; Mohamed, M. R.; Ponce de León, C.; Walsh, F. C. Recent Developments in Organic Redox Flow Batteries: A Critical Review. *J. Power Sources* **2017**, *360*, 243–283.
- (30) Cao, J.; Tian, J.; Xu, J.; Wang, Y. Organic Flow Batteries: Recent Progress and Perspectives. *Energy Fuels* **2020**, *34*, 13384–13411.
- (31) Li, M.; Rhodes, Z.; Cabrera-Pardo, J. R.; Minter, S. D. Recent Advancements in Rational Design of Non-Aqueous Organic Redox Flow Batteries. *Sustainable Energy Fuels* **2020**, *4*, 4370–4389.
- (32) Romadina, E. I.; Komarov, D. S.; Stevenson, K. J.; Troshin, P. A.; Romadina, E. I.; Komarov, D. S.; Stevenson, K. J.; Troshin, P. A. New Phenazine Based Anolyte Material for High Voltage Organic Redox Flow Batteries. *Chem. Commun.* **2021**, *57*, 2986–2989.
- (33) 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.
- (34) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine Learning for Molecular and Materials Science. *Nature* **2018**, *559*, 547–555.
- (35) Schmidt, J.; Marques, M. R. G.; Botti, S.; Marques, M. A. L. Recent Advances and Applications of Machine Learning in Solid-State Materials Science. *npj Comput. Mater.* **2019**, *5*, 83.
- (36) Wei, J.; Chu, X.; Sun, X. Y.; Xu, K.; Deng, H. X.; Chen, J.; Wei, Z.; Lei, M. Machine Learning in Materials Science. *InfoMat* **2019**, *1*, 338–358.
- (37) Pilia, G.; Wang, C.; Jiang, X.; Rajasekaran, S.; Ramprasad, R. Accelerating Materials Property Predictions Using Machine Learning. *Sci. Rep.* **2013**, *3*, 2810.
- (38) Batra, R. Accurate Machine Learning in Materials Science Facilitated by Using Diverse Data Sources. *Nature* **2021**, *589*, 524–525.
- (39) 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.; Markopoulos, G.; Jeon, S.; Kang, H.; Miyazaki, H.; Numata, M.; Kim, S.; Huang, W.; Hong, S. I.; Baldo, M.; Adams, R. P.; Aspuru-Guzik, A. Design of Efficient Molecular Organic Light-Emitting Diodes by a High-Throughput Virtual Screening and Experimental Approach. *Nat. Mater.* **2016**, *15*, 1120–1127.
- (40) 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*, 3762–3767.
- (41) Faber, F. A.; Lindmaa, A.; Von Lilienfeld, O. A.; Armiento, R. Machine Learning Energies of 2 Million Elpasolite (ABC₂D₆) Crystals. *Phys. Rev. Lett.* **2016**, *117*, 135502.
- (42) Carrasquilla, J.; Melko, R. G. Machine Learning Phases of Matter. *Nat. Phys.* **2017**, *13*, 431–434.
- (43) Cavasotto, C. N.; Di Filippo, J. I. Artificial Intelligence in the Early Stages of Drug Discovery. *Arch. Biochem. Biophys.* **2021**, *698*, 108730.
- (44) Peyton, B. G.; Briggs, C.; D’Cunha, R.; Margraf, J. T.; Crawford, T. D. Machine-Learning Coupled Cluster Properties through a Density Tensor Representation. *J. Phys. Chem. A* **2020**, *124*, 4861–4871.
- (45) Seko, A.; Hayashi, H.; Nakayama, K.; Takahashi, A.; Tanaka, I. Representation of Compounds for Machine-Learning Prediction of Physical Properties. *Phys. Rev. B* **2017**, *95*, 144110.
- (46) Sahoo, S.; Adhikari, C.; Kuanar, M.; Mishra, B. A Short Review of the Generation of Molecular Descriptors and Their Applications in Quantitative Structure Property/Activity Relationships. *Curr. Comput.-Aided Drug Des.* **2016**, *12*, 181–205.
- (47) Fisher, D.; Lukow, S. R.; Berezutskiy, G.; Gil, I.; Levy, T.; Zeiri, Y. Machine Learning Improves Trace Explosive Selectivity: Application to Nitrate-Based Explosives. *J. Phys. Chem. A* **2020**, *124*, 9656–9664.
- (48) Nayak, S.; Bhattacharjee, S.; Choi, J.-H.; Lee, S. C. Machine Learning and Scaling Laws for Prediction of Accurate Adsorption Energy. *J. Phys. Chem. A* **2019**, *124*, 247–254.
- (49) Wei, Y.; Chin, K.; Barge, L. M.; Hermis, N.; Wei, T.; Wei, T. Machine Learning Analysis of the Thermodynamic Responses of In Situ Dielectric Spectroscopy Data in Amino Acids and Inorganic Electrolytes. *J. Phys. Chem. B* **2020**, *124*, 11491–11500.
- (50) Nisbet, M. L.; Nolis, G. M.; Schrier, J.; Norquist, A. J.; Poeppelmeier, K. R.; Cabana, J. Machine-Learning-Assisted Synthesis of Polar Racemates. *J. Am. Chem. Soc.* **2020**, *142*, 7555–7566.
- (51) Wexler, R. B.; Martirez, J. M. P.; Rappe, A. M. Chemical Pressure-Driven Enhancement of the Hydrogen Evolving Activity of Ni₂P from Nonmetal Surface Doping Interpreted via Machine Learning. *J. Am. Chem. Soc.* **2018**, *140*, 4678–4683.
- (52) Lee, M.-H. Identification of Host-Guest Systems in Green TADF-Based OLEDs with Energy Level Matching Based on a Machine-Learning Study. *Phys. Chem. Chem. Phys.* **2020**, *22*, 16378–16386.
- (53) Landrum, G. RDKit: Open-source cheminformatics <https://www.rdkit.org/> (accessed Oct 23, 2021).
- (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.; Li, X.; Caricato, M.; Marenich, A.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ort, J. V.; Fox, D. J. *Gaussian 09*; Gaussian, Inc.: Wallingford CT 2016.
- (55) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785.
- (56) Becke, A. D. Density-functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648.
- (57) Tirado-Rives, J.; Jorgensen, W. L. Performance of B3LYP Density Functional Methods for a Large Set of Organic Molecules. *J. Chem. Theory Comput.* **2008**, *4*, 297–306.
- (58) Schuchardt, K. L.; Didier, B. T.; Elsethagen, T.; Sun, L.; Gurumoorthi, V.; Chase, J.; Li, J.; Windus, T. L. Basis Set Exchange: A Community Database for Computational Sciences. *J. Chem. Inf. Model.* **2007**, *47*, 1045–1052.
- (59) Zhao, Y.; Truhlar, D. G. Density Functionals with Broad Applicability in Chemistry. *Acc. Chem. Res.* **2008**, *41*, 157–167.
- (60) Papajak, E.; Truhlar, D. G. Efficient Diffuse Basis Sets for Density Functional Theory. *J. Chem. Theory Comput.* **2010**, *6*, 597–601.
- (61) Treitel, N.; Shenhar, R.; Aprahamian, I.; Sheradsky, T.; Rabinovitz, M. Calculations of PAH Anions: When Are Diffuse Functions Necessary? *Phys. Chem. Chem. Phys.* **2004**, *6*, 1113–1121.
- (62) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
- (63) Nakagawa, R.; Nishina, Y. Simulating the Redox Potentials of Unexplored Phenazine Derivatives as Electron Mediators for Biofuel Cells. *J. Phys. Energy* **2021**, *3*, 034008.
- (64) Miao, L.; Liu, L.; Zhang, K.; Chen, J. Molecular Design Strategy for High-Redox-Potential and Poorly Soluble n-Type Phenazine Derivatives as Cathode Materials for Lithium Batteries. *ChemSusChem* **2020**, *13*, 2337–2344.
- (65) Sousa, A. C.; Martins, L. O.; Robalo, M. P. Laccases: Versatile Biocatalysts for the Synthesis of Heterocyclic Cores. *Mol.* **2021**, *26*, 3719.
- (66) Castro, K. P.; Clikeman, T. T.; DeWeerd, N. J.; Bukovsky, E. V.; Rippey, K. C.; Kuvychko, I. V.; Hou, G. L.; Chen, Y. S.; Wang, X.

B.; Strauss, S. H.; Boltalina, O. V. Incremental Tuning Up of Fluorous Phenazine Acceptors. *Chem.—Eur. J.* **2016**, *22*, 3930–3936.

(67) Wang, C.; Li, X.; Yu, B.; Wang, Y.; Yang, Z.; Wang, H.; Lin, H.; Ma, J.; Li, G.; Jin, Z. Molecular Design of Fused-Ring Phenazine Derivatives for Long-Cycling Alkaline Redox Flow Batteries. *ACS Energy Lett.* **2020**, *5*, 411–417.

(68) 1.1. Linear Models—scikit-learn 1.0 documentation https://scikit-learn.org/stable/modules/linear_model.html#bayesian-regression (accessed Oct 23, 2021).

(69) Wipf, D.; Nagarajan, S. A New View of Automatic Relevance Determination. *Advances in Neural Information Processing Systems 20, Proceedings of the Twenty-First Annual Conference on Neural Information Processing Systems*. 2007.

(70) 1.7. Gaussian Processes—scikit-learn 1.0 documentation https://scikit-learn.org/stable/modules/gaussian_process.html (accessed Oct 23, 2021).

(71) Sit, H.. Quick Start to Gaussian Process Regression <https://towardsdatascience.com/quick-start-to-gaussian-process-regression-36d838810319> (accessed Oct 23, 2021).

(72) 1.3. Kernel ridge regression—scikit-learn 1.0 documentation https://scikit-learn.org/stable/modules/kernel_ridge.html (accessed Oct 23, 2021).

(73) 1.4. Support Vector Machines—scikit-learn 1.0 documentation <https://scikit-learn.org/stable/modules/svm.html#svm-regression> (accessed Oct 23, 2021).

(74) sklearn.feature_selection.SelectKBest—scikit-learn 1.0.2 documentation https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.SelectKBest.html (accessed Jan 7, 2022).

(75) 4.2. Permutation feature importance—scikit-learn 1.0 documentation https://scikit-learn.org/stable/modules/permutation_importance.html#permutation-importance (accessed Oct 23, 2021).

(76) 1.1. Linear Models—scikit-learn 1.0 documentation https://scikit-learn.org/stable/modules/linear_model.html (accessed Oct 21, 2021).

(77) Landrum, G. Getting Started with the RDKit in Python—The RDKit 2020.03.1 documentation <https://www.rdkit.org/docs/GettingStartedInPython.html#list-of-available-descriptors> (accessed March 31, 2021).

(78) QuaSAR-Descriptor <http://www.cadaster.eu/sites/cadaster.eu/files/challenge/descr.htm> (accessed Oct 22, 2021).

(79) rdkit.Chem.MolSurf module—The RDKit 2021.09.1 documentation <https://www.rdkit.org/docs/source/rdkit.Chem.MolSurf.html> (accessed Dec 29, 2021).

(80) Gasteiger, J.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativity—a Rapid Access to Atomic Charges. *Tetrahedron* **1980**, *36*, 3219–3228.

Recommended by ACS

Machine Learning Determination of New Hammett's Constants for *meta*- and *para*-Substituted Benzoic Acid Derivatives Employing Quantum Chemical Atomic Char...

Gabriel Monteiro-de-Castro, Itamar Borges Jr., *et al.*

JULY 11, 2023

THE JOURNAL OF ORGANIC CHEMISTRY

READ 

Toward Universal Substituent Constants: Relating QAIM Functional Group Descriptors to Substituent Effect Proxies

Kevin M. Lefrancois-Gagnon and Robert C. Mawhinney

SEPTEMBER 20, 2023

JOURNAL OF CHEMICAL INFORMATION AND MODELING

READ 

PubChemQC B3LYP/6-31G**/PM6 Data Set: The Electronic Structures of 86 Million Molecules Using B3LYP/6-31G* Calculations

Maho Nakata and Toshiyuki Maeda

SEPTEMBER 07, 2023

JOURNAL OF CHEMICAL INFORMATION AND MODELING

READ 

High-Throughput Screening of Promising Redox-Active Molecules with MolGAT

Mesfin Diro Chaka, Yedilfana Setarge Mekonnen, *et al.*

JUNE 30, 2023

ACS OMEGA

READ 

Get More Suggestions >