

Computational design of quinone electrolytes for redox flow batteries using machine learning and high-throughput theoretical calculations

Fei Wang,¹ Jipeng Li,² Zheng Liu,¹ Tong Qiu¹, Jianzhong Wu^{,3}, Diannan Lu^{*,†}*

¹ Department of Chemical Engineering, Tsinghua University, Beijing, 100084

² School of Materials Science and Engineering, Hainan University, Haikou 570228

³ Department of Chemical and Environmental Engineering and Department of Mathematics, University of California, Riverside, California 92521, United States

E-mail: ludiannan@tsinghua.edu.cn; jwu@engr.ucr.edu

Abstract

Molecular design of redox-active materials with higher solubility and greater redox potential windows is instrumental in enhancing the performance of redox flow batteries (RFBs). Here we propose a sweeping computational procedure for a systematic evaluation of organic redox-active species by combining machine learning, quantum-mechanical and classical-density-function-theory calculations. 1517 small quinone molecules were generated from the building blocks of benzoquinone, naphthoquinone, and anthraquinone with different substituent groups. The physics-based methods were used to predict HOMO-LUMO gaps and solvation free energies that account for the redox potential differences and aqueous solubility, respectively. The high-throughput calculations were augmented with the quantitative structure-property relationship

(QSPR) analyses and machine learning/graph network modeling to evaluate the materials overall behavior. The computational procedure was able to reproduce high-performance cathode electrolyte materials consistent with experimental observations and identify new electrolytes for RFBs by screening 100,000 di-substituted quinone molecules, the largest library of redox-active molecules ever investigated. The efficient computational platform may facilitate better understanding of the structure-function relationship of quinone molecules and advance the design and application of all-organic active materials for RFBs.

Keywords: quinones, redox flow battery, machine learning

Introduction

Large-scale, stationary energy-storage techniques are imperative for the widespread applicability of green energy such as wind and solar power¹. A redox flow battery (RFB) is an electrochemical energy storage device², in which catholytes and anolytes are stored in separate external tanks and transported to the battery for energy conversions. The RFB power is determined by the capacity of electrodes while its energy density depends on the volume, the composition, and the concentration of the redox-active electrolytes³⁻⁵. The complete decoupling of the power and energy density makes RFBs ideal for high-capacity energy storage with operational flexibility.

Several types of RFBs have been proposed, such as all-vanadium flow batteries, Fe/Cr flow batteries, and zinc bromide flow batteries⁶. All-vanadium RFBs utilize the same electrolytes for

both half-cells and are advantageous in preventing contamination and thereby improving the battery life cycle^{4,7,8}. The cost and energy efficiency of all-vanadium RFBs, however, need to be improved for industrial applications^{9,10}. Lithium-ion flow batteries¹¹ have the advantage of high current density, capacity and flexibility, but their performance is often compromised by the availability of the lithium conducting and storage materials. Aqueous organic redox flow batteries (AORFBs) may overcome some of the major hurdles of the metal ion-based RFBs by using Earth-abundant elements as C, H, O, N, S.^{12–17} The tunability of redox-active molecules offers new avenues for engineering design and optimization by selecting materials with wider voltage windows, higher aqueous solubility, increased electrochemical stability and faster electrode reaction kinetics. For example, quinones have a backbone structure that allows for a two-electron redox reaction in an aqueous solution. Their electrochemical properties and solution behavior can be tuned by substitutions at the benzene rings. An open circuit voltage of 1.2V was achieved by utilizing 2,6-DHAQ¹⁴ as the electrolyte under alkaline conditions, but the replacement of 2,6-DHAQ with PEGAQs¹⁵ reduces the voltage to 1.0V. Given the inexhaustible types of quinone organics that can be obtained by varying the substitution positions and substituent groups, a high-throughput molecular design method is needed for the development of all-quinone RFBs.

One key objective of molecular quinone design is to identify substitutions leading to a higher redox potential difference and higher solubility such that the RFBs can achieve a higher energy density. Whereas physics-based models are available to predict such properties, the direct

application is often limited by their demanding computational cost in evaluating a large library of chemical species under diverse thermodynamic conditions. Exploring the chemical space requires new computational strategies to improve the design efficiency. The machine-learning (ML) methods provide complementary alternatives with particular strength in high-throughput screening and molecular design. To prepare the “big data” required for training ML models, we combine quantum-mechanical (QM) and classical density function theory (cDFT) calculations with quantitative structure-property relationship (QSPR) analyses such that the physicochemical properties of quinones can be systematically evaluated.

A number of previous studies have been reported on the molecular design of redox-active materials for RFBs. For example, Allam et al.¹⁶ applied artificial neural networks(ANN), gradient-boosting regression(GBR), and kernel ridge regression(KRR) to predict the redox potentials of quinones based on the electron affinity and the number of bound lithium atoms. Li et al.¹⁷ applied machine-learning methods to predict the cost and efficiency of all-vanadium flow batteries. Lin et al.¹⁸ used an extreme learning machine(ELM) model to predict the properties of a novel redox flow battery—the single flow Zinc-Nickle battery. However, to our knowledge, there is no proposal for seeking high-performance quinone molecules used in RFBs. In this work, we firstly constructed 1517 small quinone molecules using benzoquinone, naphthoquinone, and anthraquinone as the building blocks. QM and cDFT calculations were carried out to evaluate their HOMO-LUMO gaps and solvation free energies. The molecular features of promising electrolytes were extracted from

the perspectives of molecular descriptors and graph networks for machine-learning models. We demonstrated the effectiveness of the proposed method with experimental results for the existing quinone electrolytes and its potential use for the computational design of redox-active materials by screening 100,000 di-substituted quinone molecules.

Models and Methods

The computational scheme for quinone design

Figure 1 presents the overall computational scheme for the molecular design of quinones that may be used as catholytes in RFBs. Following previous work^{19–22}, we chose benzoquinone, naphthoquinone, and anthraquinone as the building blocks to construct a library of 1517 mono- or fully-substituted quinone molecules by the addition of different substituent groups at different substitution positions. For most of the virtual quinone molecules, the experimental data for the solvation free energies and the HOMO-LUMO gaps have not been determined. Thus, the accuracy of our theoretical methods was calibrated with polycyclic aromatic hydrocarbons that are similar quinones. The calculated results agree well with the experimental data²⁰. Specifically, we used a quantum-chemistry package (ORCA) and classic density functional theory (cDFT) to calculate the HOMO-LUMO gaps and solvation free energies of these substances. The molecular properties provide a quantitative measure of the redox potential difference and solubility of the electrolyte in AORFBs. and also determine the capacity and performance of the flow battery. The theoretical results are then analyzed with the QSPR method and machine learning/graph network models.

Next, surrogate models were constructed to reproduce the HOMO-LUMO gaps and solvation free energies of quinone molecules and analyze the dependence of such properties on substituents at different backbone positions. Compared with mono- and multi-substituted quinones, di-substituted quinones hold advantages in terms of a balance of stability, diversity and synthesizability. Thus, our molecular design is focused on the properties of all possible di-substituents with the same backbone. Quinone molecules promising for AORFBs have been identified from over 100,000 di-substituted quinones.

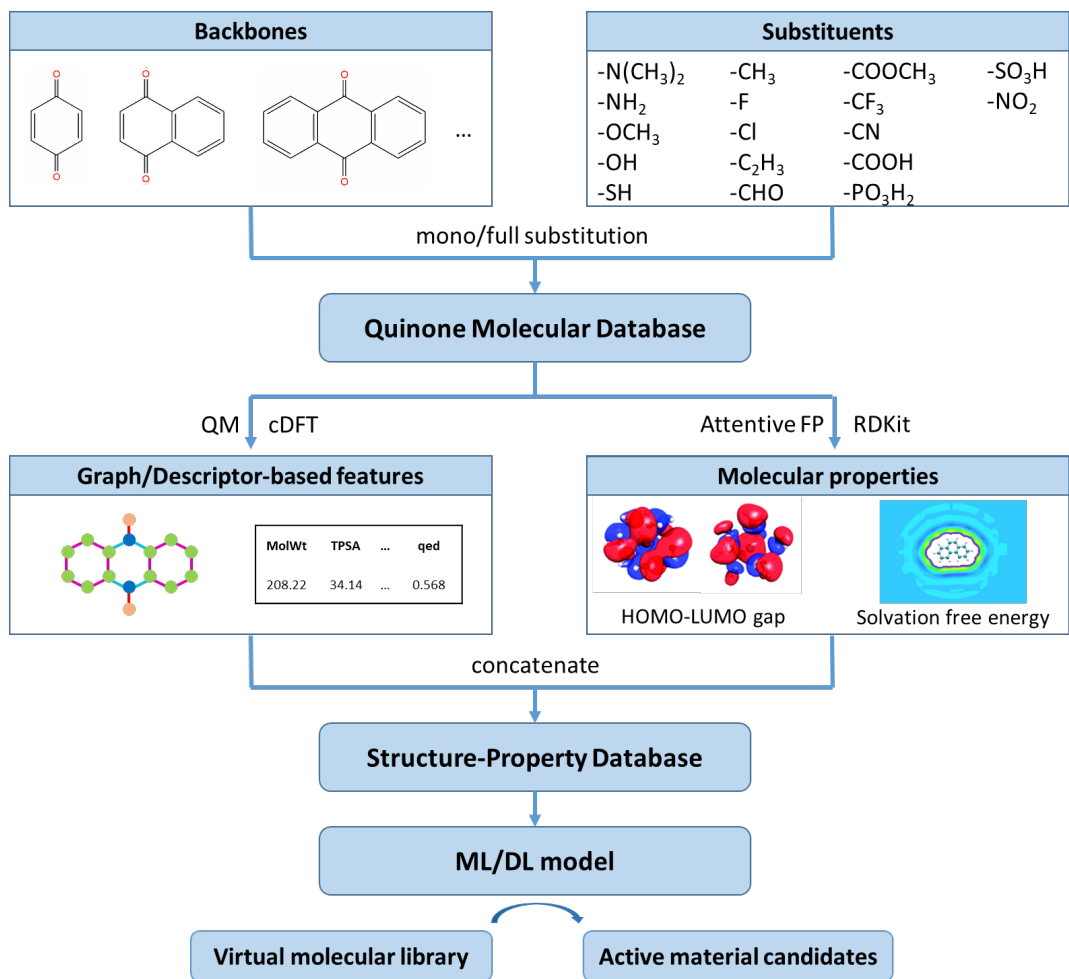


Figure 1. The computational scheme for molecular quinone design

Results and discussion

The choice of ML models

We firstly constructed 1516 quinone molecules with 17 different substituents using benzoquinone, naphthoquinone, and anthraquinone as building blocks (Supplementary Figure S1). The theoretical predictions for the solvation free energies and HOMO-LUMO gaps are used as a database for training two machine-learning (ML) models, eXtreme Gradient Boosting (Xgboost) and Attentive Fingerprints (FP). The molecular properties, along with the predicted values from machine learning, are listed in Supplementary Table S2. In testing the ML models, the data are divided into a training set, a validation set, and a test set with a ratio of 8:1:1.

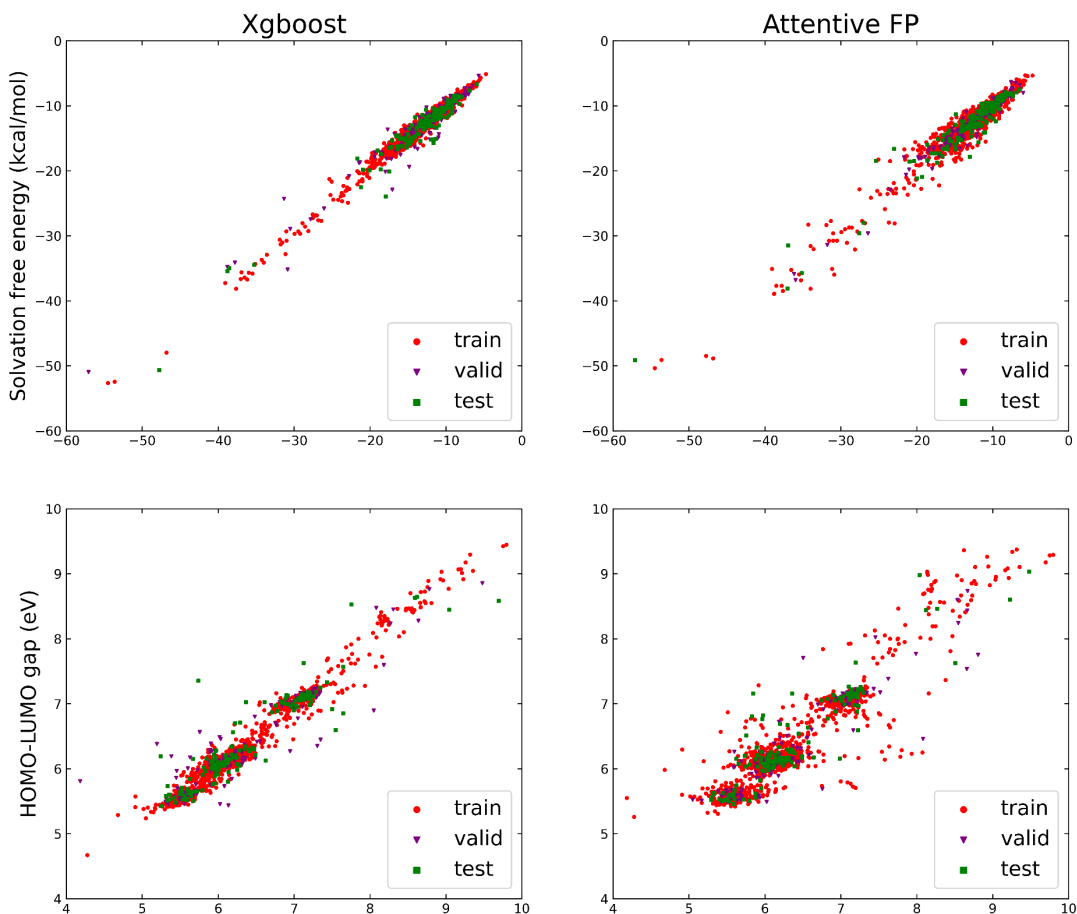


Figure 2. A comparison of Xgboost and Attentive FP methods to predict the solvation-free energies and HOMO-LUMO gaps of 1516 quinone molecules. (a) Xgboost for predicting the solvation free energies; (b) Xgboost for predicting the HOMO-LUMO gaps; (c) Attentive FP for predicting the solvation free energies; and (d) Attentive FP for predicting the HOMO-LUMO gaps.

Table 1. Various indicators of data fitting with Xgboost and Attentive FP methods.

Property	Model	R ²	MAE	RMSE
----------	-------	----------------	-----	------

Solvation free energy	Xgboost	0.9777	0.4506	0.2030
	Attentive FP	0.9408	0.8425	0.7098
HOMO-LUMO gap	Xgboost	0.9632	0.1034	0.0107
	Attentive FP	0.8596	0.1800	0.0324

As shown in Figure 2, Xgboost is more accurate than Attentive FP for fitting both the solvation free energies and the HOMO-LUMO gaps. Table 1 lists various indicators of the data correlation. In addition, the two methods have made higher-precision predictions of the solvation free energy, and the predicted value has a higher linear correlation with the experimental value calculated by the theory. Attentive FP's model gives a poor prediction of the HOMO-LUMO gap, in which most of the points formed by the experimental value and the predicted value deviated from the diagonal line. It is anticipated that the better prediction by the Xgboost model may attribute to feature selection.

In fitting the solvation free energies and the HOMO-LUMO gaps of quinone molecules with the Xgboost model, we have used 200 molecular descriptors These descriptors show good correlation with both the solubility and orbital properties of the quinone molecule. By contrast, the Attentive FP model is based on the molecular characteristics constructed on the basis of the graph network. Such characteristics are more derived from the surrounding environment of each atom rather than the overall features of the molecule, which explains its relatively poor predictive ability in comparison with Xgboost. In general, the performance of the machine learning models depends on the choice of hyperparameters (viz.,...), the quality of the data, and the selection

of features. As we can see in Figure 2, the selected features better reflect the solvation free energies of quinone molecules than their HOMO-LUMO gaps. It appears that the latter requires a larger data set or deeper feature engineering to have better correlations. Moreover, it can be seen from Table 1 that the prediction errors of the Xgboost model are acceptable for high-throughput screening. In the following, our analyses of quinone molecules for RFBs are all based on the Xgboost model.

The effects of substituent position and groups

We have constructed 17 molecules using benzoquinone, naphthoquinone, and anthraquinone as the building blocks. To study the influence of substituent position and substituent type on its molecular characteristics, we used the 1,10-anthraquinone (AQ) as a representative backbone and compared the difference between theoretical results and ML correlation data according to the substituent position and substituent type. Figure 3 shows the results, and Table 2 presents the corresponding substituent positions.

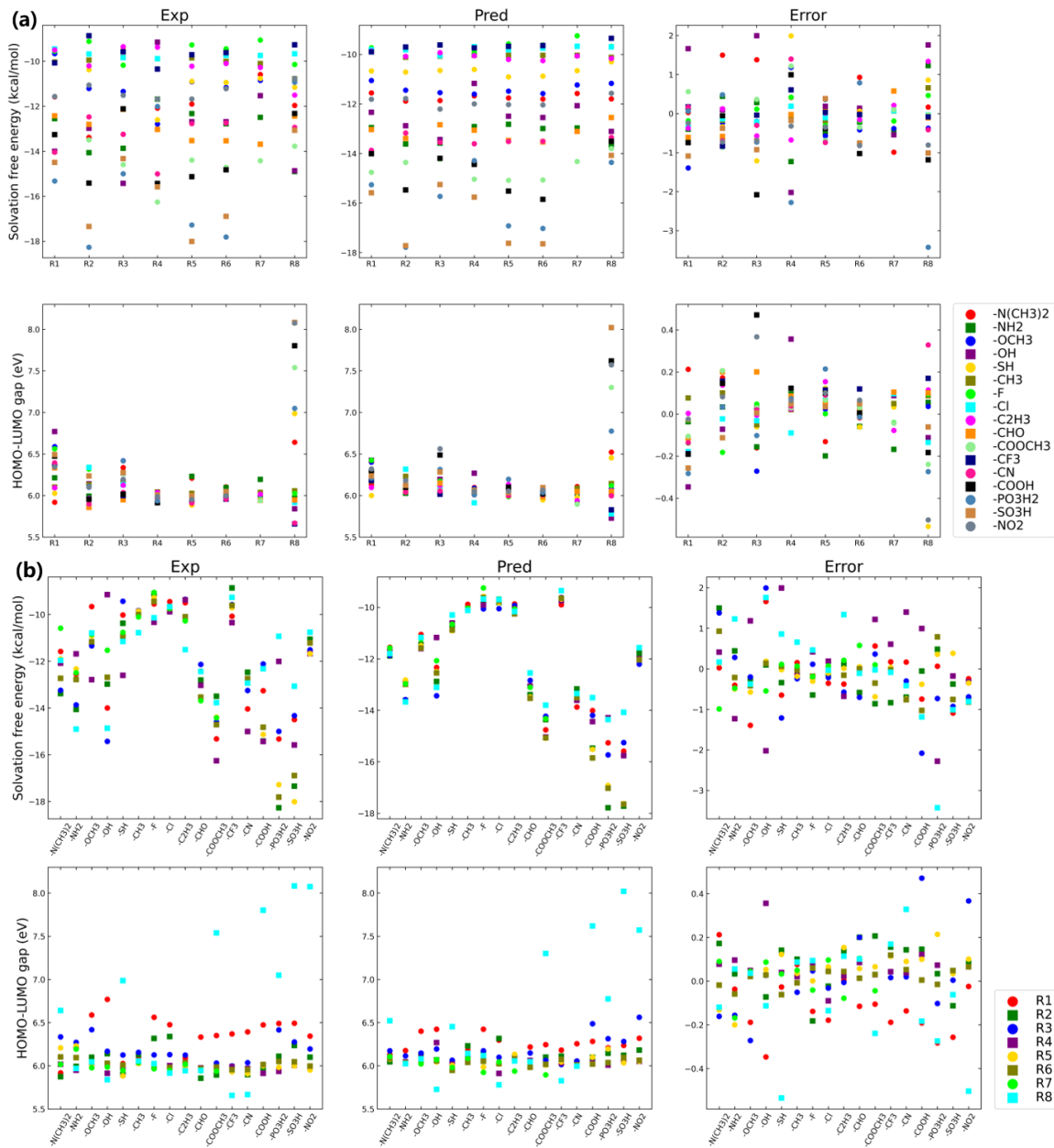


Figure 3. Solvation-free energies and HOMO-LUMO gaps of different quinones. (a) Effect of different substituent types on target properties. (b) Effect of different substituent positions on target properties.

Table 2. The positions of substituent for a few representative quinone molecules.

R1	R2	R3	R4	R5	R6	R7	R8
----	----	----	----	----	----	----	----

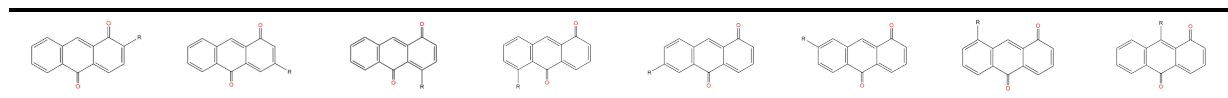


Figure 3a shows the effects of different substituent groups at the same substitution position. It can be found that substitution with $-\text{PO}_3\text{H}_2$, $-\text{SO}_3\text{H}$, $-\text{COOCH}_3$, $-\text{COOH}$ groups reduces the solvation free energy regardless of the substituent positions. The ranking of different substituents reproduces that by quantum-chemistry calculations. For the HOMO-LUMO gaps, different substituent groups have less influence on the target properties at the substitution positions of R1-R7, while at the R8 substitution position, the HOMO-LUMO gaps are sensitive to different substituent groups. Specifically, all substitutions at the R1 position result in a higher HOMO-LUMO gap compared to substitution at other positions, while substitution with groups such as $-\text{COOCH}_3$, $-\text{COOH}$, $-\text{SO}_3\text{H}$, $-\text{NO}_2$ at the R8 position can greatly improve the HOMO-LUMO gap.

Figure 3b shows the molecular properties of quinones with the same substituent group but at different substitution positions. The results predicted by the ML model are similar to those from the physics-based calculations. Regardless of the substitution positions, substitution with $-\text{F}$, $-\text{Cl}$, and $-\text{CH}_3$ groups has little impact on the solvation free energies of the quinone molecules. On the other hand, substitution with $-\text{PO}_3\text{H}_2$ and $-\text{SO}_3\text{H}$ groups leads to a significant reduction of the solvation free energy. In this case, the result is highly dependent on the position of the substituent. The lowest solvation free energy can be obtained when the substitution takes place on positions R2, R5, and R6. In terms of the HOMO-LUMO gaps, most of the substituents have better performance in the substitution positions R1 and R8. For some poorly performing substituent

groups, satisfactory target properties cannot be obtained even if they are in the R8 substitution position. The above results indicate the substituents at the R1 position can improve the HOMO-LUMO gap, and the more refined selection of the substituents at the R8 position may achieve more extraordinary effects.

The error graphs in Figure 3 suggest that the machine-learning (ML) model has a fairly high level of fidelity in reproducing the theoretical results. In the mono-substituted quinones with 1,10-AQ as the backbone, the prediction error of the ML model for the solvation free energy is between -2 and 3 kcal/mol, and the prediction error of the HOMO-LUMO gap is mostly between -0.2 and 0.2 eV. However, the prediction error of the model for the R8 substitute HOMO-LUMO gap is more significant, about $??$. For a high-throughput screening tool, this relative error is still within our acceptable range.

Screening from over 100,000 candidates

After establishing the computational method, please describe how it can be used as a design tool to identify new quinone molecules. Specify the targets and explain connection between the molecular properties (i.e., HOMO-LUMO gaps and solvation free energies) and the performance of RFBs.

To demonstrate the high-through capabilities of our computational framework, we used the Xgboost model to predict the solvation free energies and the HOMO-LUMO gaps for over 100,000 di-substitutions derived from the 17 quinone-based backbones. It took about 10 seconds to

complete the predictions and rank all molecules according to In contrast, it took over 30 minutes to complete the calculation of the solvation free energy and the HOMO-LUMO gap of a single molecule. We see an increase in calculation speed by 7 orders of magnitude!

The predicted results are shown in Figure 4. Here the predicted solvation free energies for all di-substituted quinone molecules is generally within -25 and -10 kcal/mol, and the predicted value of the HOMO-LUMO gap is between 5 and 9 eV. The potential of the molecule as a battery active material is scored by $0.5 \times \text{Solvation Free Energy} / -30 + 0.5 \times \text{HOMO-LUMO gap} / 10$. The detailed calculation results are available in Table S3.

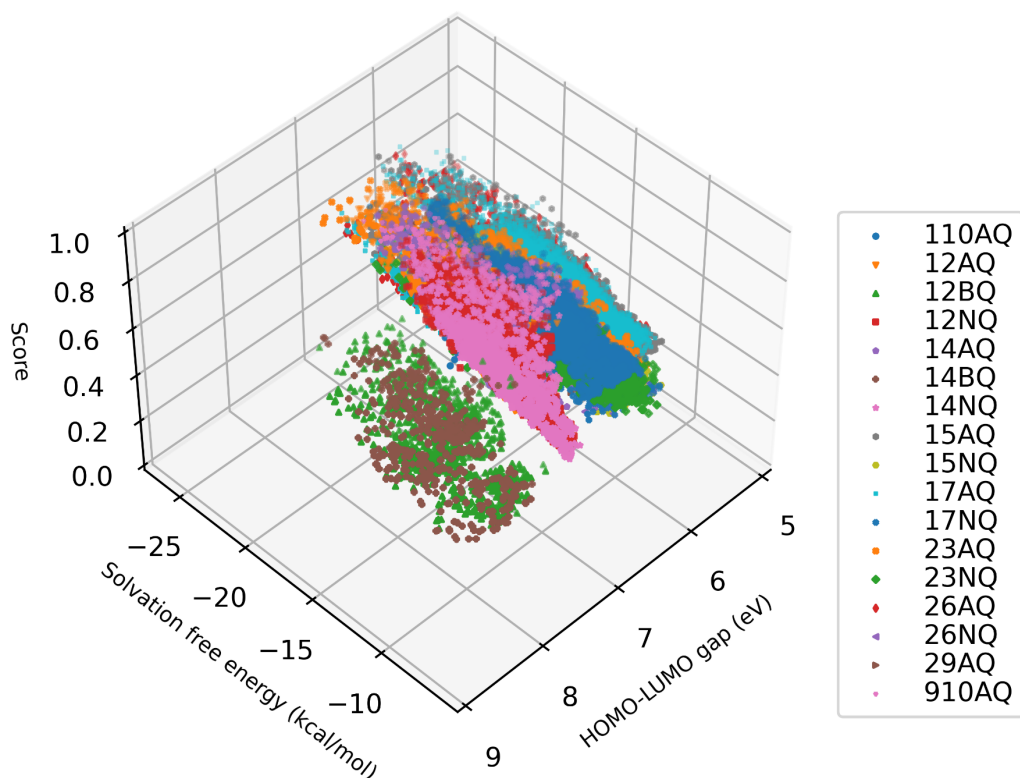


Figure 4. A score of over 100,000 virtual quinone molecules.

As shown in Figure 4, different quinone-based backbones have a great impact on molecular properties. The di-substituents of the 1,2-benzoquinone(BQ) and 1,4-BQ backbones generally yield high HOMO-LUMO gaps, but their solvation-free energies are not as low as other backbones. The backbones with better overall performance include 1,2-AQ, 1,4-AQ, 1,5-AQ. These quinone molecules can attain low solvation-free energy and high HOMO-LUMO gap simultaneously, and thus can be used as candidate materials for quinone-based flow batteries for experimental verification.

To further explore the common characteristics of quinone molecules with different scores, we selected 10,000 molecules with the highest, lowest, and medium scores from 100,000+ molecules, and analyzed the effects of backbones and substituents, as shown in Figure 5. We did not perform an in-depth analysis of the relationship between the score and the position of the substituents due to the difficulty in unifying the positions of the substituents on different backbones.

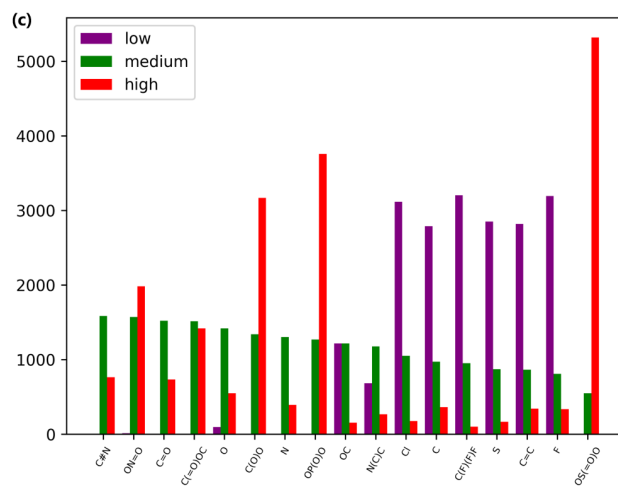
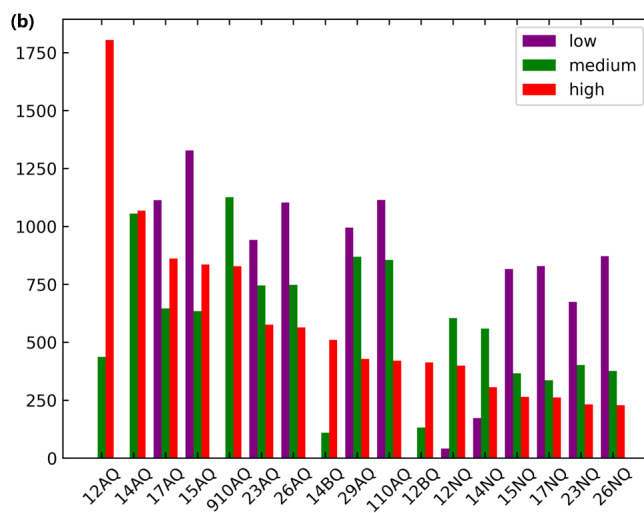
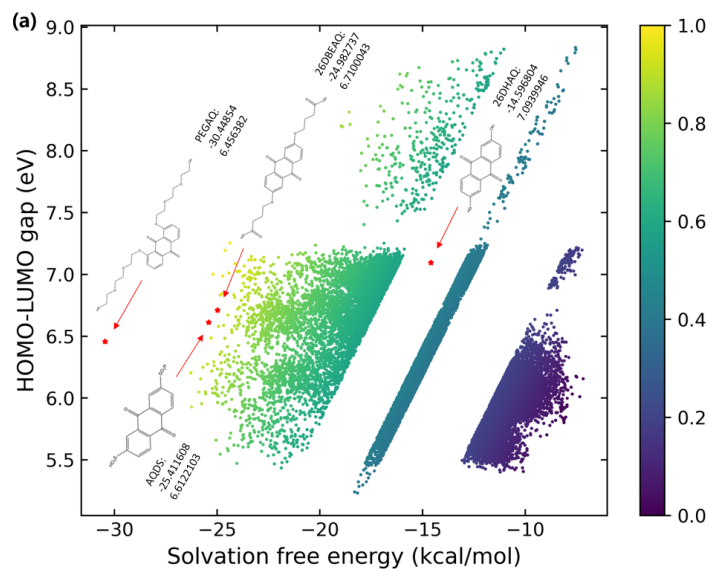


Figure 5. Virtual molecular library scoring and analysis of results. (a) Scoring chart for di-substituted quinones and reported quinone-based electrode materials; (b) Backbone distribution in three types of scoring molecules; (c) Distribution of substituent types in three types of scoring molecules.

Figure 5 shows the statistics of the backbones and substituent types according to three different scores of di-substituted molecule populations. As discussed above, most high-scoring molecules consist of the di-substitutions of 1,2-AQ and 1,4-AQ backbones. While all 17 backbones can get medium and high scores of quinone molecules, no di-substitutions constructed by the five backbones of 1,2-AQ, 1,4-AQ, 9,10-AQ, 1,4-BQ and 1,2-BQ appear in the low-scoring group. In terms of the choice of substituent types, the large-scale tests can better reflect the influence of different substituents on molecular properties. In the high-scoring group, substituents with $-\text{SO}_3\text{H}$, $-\text{PO}_3\text{H}$ and $-\text{COOH}$ groups account for nearly half of the molecules, and the molecules obtained by their substitution will not be given low scores. The effect of $-\text{Cl}$, $-\text{CH}_3$, $-\text{CF}_3$, $-\text{SH}$, $-\text{F}$, $-\text{C}_2\text{H}_5$ is obviously not suitable substituents for quinone-based electrolytes, and the scores of molecules generated by their substitution are generally at a low level.

Combining the above two points, adjusting the $-\text{SO}_3\text{H}$, $-\text{PO}_3\text{H}$, and $-\text{COOH}$ groups on the basis of the 1,2-AQ or 1,4-AQ backbone is expected to obtain battery materials with the best properties. From Figure 6a, we can also see that among all the above 100,000 quinone molecules, almost none have achieved good results in terms of both solvation-free energy and the HOMO-

LUMO gap at the same time. In terms of our scoring system for the entire disubstituted quinones, for a good battery material alternative molecule, it should have at least a HOMO-LUMO gap of more than 7 eV and a solvation free energy of less than -20 kcal/mol, which means it must be in the blank region of the current material in the upper left corner of the scoring chart. Unfortunately, by now there is no such quinone to achieve these criteria. In addition, we also used the same model to predict the corresponding properties of quinone-based battery materials that have been reported in the literature^{14,15,35,36}. The full cell performance data consisting of the above material-based composition is compiled in Table S4. As shown in Figure 5a, all the materials that have been verified by experiments, but not in our data set samples, get high scores under the evaluation of this model, which means that our machine learning model has certain accuracy for the screening ability of battery materials. For PGEAQ, the modification of the PEG group greatly reduces the solvation free energy of the anthraquinone (AQ) molecules, which makes the molecule obtain a high evaluation score, but its HOMO-LUMO gap is actually in the lower-middle position in the whole evaluation system. For 26DHAQ, although it does not appear to be in a dominant position, it actually has the highest predicted voltage, except that the hydroxyl modification cannot bring a great improvement in solubility for the backbone molecule. In fact, it is reasonable to consider whether the modification of the molecule in the upper right corner of the scoring diagram with PEG could result in a new molecule that is good in both target properties. In addition, there are a number of molecules in the sample set of di-substituted substances that we have constructed to

obtain scores comparable to them, showing that there is still a lot of room to find in the screening of battery materials, not to mention that the data set we constructed is only within the limited backbone and limited substituent types.

We also calculated the corresponding properties for the above four molecules reported in the literature using quantitative calculations and the results are presented in Table 3. Although the property prediction model is generated based on the property data of single substituents, resulting in a reasonable error in predicting the properties of these di-substitutes, its prediction is still informative in terms of ranking ability compared to quantitative calculations.

Table 3. Comparison of quantitative calculation results with ML predicted results.

SPECIES	QC		ML	
	SFE	HOMO-LUMO gap	SFE	HOMO-LUMO gap
26DHAQ	-14.2380	8.127	-14.5968	7.094
PEGAQ	-24.8238	7.525	-30.4485	6.456
26DBEAQ	-25.0677	7.463	-24.9827	6.710
AQDS	-34.7494	7.487	-25.4116	6.612

Conclusion

In this work, we constructed a virtual database of quinone small molecules with quinone molecular backbones and predefined substituents, and further constructed a database of their structural properties by combining theoretical calculations. Then we combined machine learning tools to build an artificial intelligence model for fast prediction of target properties from molecular structure and verified the accuracy of the machine learning model for predicting the solvation free energy and HOMO-LUMO gap of quinone molecules, as well as its ability to reduce quantification

and density functional theory calculation results. Through the analysis of the calculation results, we discussed the influence of the quinone backbone, the position of the substituent, and the type of the substituent on the characteristics of the quinone molecule, which proved that the machine learning method can be used as a high-throughput screening method for flow battery active materials with much higher efficiency than the theoretical calculation. Based on a model for rapid property prediction, we constructed over 100,000 virtual quinone molecules, evaluated their performance as active substances in redox flow batteries, and compared them with some molecules reported in the literature. Actually, all molecules, both virtual and real, have not been able to obtain sufficient advantages in solubility and open-circuit voltage at the same time, which also indicates that there is still much room for research on the development of active materials for redox flow batteries. The predictive model of molecular target properties and the effect of substituents and substitution positions on the properties of the backbone molecule can quickly help us to find the direction of molecular property improvement and assist in the design and development of new molecules with better performance.

Supporting Information

It provides computational details of the machine learning model building process, including the construction of a virtual quinone-like molecular library, quantitative computational results of molecular properties, hyperparameter selection for the Xgboost model, and model prediction results of molecular properties.

Acknowledgment

This work is supported and sponsored by the National Natural Science Foundation of China (No. U1862204). The numerical calculations were performed at the “Tianhe-2” platform supported by Guangzhou Supercomputer Centre.

Reference

- (1) Hasewend, B.; Jokic, T. How the European Green Deal Promotes Sustainable Energy Research and Innovation. In *Solar Energy Conversion in Communities*; Visa, I., Duta, A., Eds.; Springer Proceedings in Energy; Springer International Publishing: Cham, 2020; pp 455–456. https://doi.org/10.1007/978-3-030-55757-7_32.
- (2) Eckroad, S.; Gyuk, I. EPRI-DOE Handbook of Energy Storage for Transmission & Distribution Applications. *Electric Power Research Institute, Inc* 2003, 3–35.
- (3) Zhou, H.; Zhang, H.; Zhao, P.; Yi, B. A Comparative Study of Carbon Felt and Activated Carbon Based Electrodes for Sodium Polysulfide/Bromine Redox Flow Battery. *Electrochimica Acta* 2006, 51 (28), 6304–6312. <https://doi.org/10.1016/j.electacta.2006.03.106>.
- (4) Skyllas-Kazacos, M.; Grossmith, F. Efficient Vanadium Redox Flow Cell. *J. Electrochem. Soc.* 1987, 134 (12), 2950. <https://doi.org/10.1149/1.2100321>.
- (5) Remick, R. J.; Ang, P. G. P. Electrically Rechargeable Anionically Active Reduction-Oxidation Electrical Storage-Supply System. US4485154A, November 27, 1984.

- (6) Ye, J.; Xia, L.; Wu, C.; Ding, M.; Jia, C.; Wang, Q. Redox Targeting-Based Flow Batteries. *J. Phys. D: Appl. Phys.* 2019, 52 (44), 443001. <https://doi.org/10.1088/1361-6463/ab3251>.
- (7) Tsekouras, G.; Anastasopoulos, C.; Kanellos, F.; Kontargyri, V.; Karanasiou, I.; Salis, A.; Mastorakis, N. A Demand Side Management Program of Vanadium Redox Energy Storage System for an Interconnected Power System; 2008; pp 26–28.
- (8) Hawkins, J. M.; Robbins, T. A Field Trial of a Vanadium Energy Storage System; IET, 2001; pp 652–656.
- (9) Ma, X.; Zhang, H.; Sun, C.; Zou, Y.; Zhang, T. An Optimal Strategy of Electrolyte Flow Rate for Vanadium Redox Flow Battery. *Journal of power sources* 2012, 203, 153–158.
- (10) Li, S.; Huang, K.; Liu, S.; Fang, D.; Wu, X.; Lu, D.; Wu, T. Effect of Organic Additives on Positive Electrolyte for Vanadium Redox Battery. *Electrochimica Acta* 2011, 56 (16), 5483–5487.
- (11) Pham-Truong, T.; Wang, Q.; Ghilane, J.; Randriamahazaka, H. Recent Advances in the Development of Organic and Organometallic Redox Shuttles for Lithium-ion Redox Flow Batteries. *ChemSusChem* 2020, 13 (9), 2142–2159.
- (12) Huskinson, B.; Marshak, M. P.; Suh, C.; Er, S.; Gerhardt, M. R.; Galvin, C. J.; Chen, X.; Aspuru-Guzik, A.; Gordon, R. G.; Aziz, M. J. A Metal-Free Organic–Inorganic Aqueous Flow Battery. *Nature* 2014, 505 (7482), 195–198.

- (13) Darling, R. M.; Gallagher, K. G.; Kowalski, J. A.; Ha, S.; Brushett, F. R. Pathways to Low-Cost Electrochemical Energy Storage: A Comparison of Aqueous and Nonaqueous Flow Batteries. *Energy & Environmental Science* 2014, 7 (11), 3459–3477.
- (14) Kwabi, D. G.; Lin, K.; Ji, Y.; Kerr, E. F.; Goulet, M.-A.; De Porcellinis, D.; Tabor, D. P.; Pollack, D. A.; Aspuru-Guzik, A.; Gordon, R. G. Alkaline Quinone Flow Battery with Long Lifetime at PH 12. *Joule* 2018, 2 (9), 1894–1906.
- (15) Jin, S.; Jing, Y.; Kwabi, D. G.; Ji, Y.; Tong, L.; De Porcellinis, D.; Goulet, M.-A.; Pollack, D. A.; Gordon, R. G.; Aziz, M. J. A Water-Miscible Quinone Flow Battery with High Volumetric Capacity and Energy Density. *ACS Energy Letters* 2019, 4 (6), 1342–1348.
- (16) Allam, O.; Kuramshin, R.; Stoichev, Z.; Cho, B.; Lee, S.; Jang, S. Molecular Structure–Redox Potential Relationship for Organic Electrode Materials: Density Functional Theory–Machine Learning Approach. *Materials Today Energy* 2020, 17, 100482.
- (17) Li, T.; Xing, F.; Liu, T.; Sun, J.; Shi, D.; Zhang, H.; Li, X. Cost, Performance Prediction and Optimization of a Vanadium Flow Battery by Machine-Learning. *Energy & Environmental Science* 2020, 13 (11), 4353–4361.
- (18) Lin, X.; Guo, Y.; Cheng, J.; Guo, Z.; Yan, X. An Improved Extreme Learning Machine Model and State-of-Charge Estimation of Single Flow Zinc-Nickle Battery; Springer, 2017; pp 613–622.

- (19) Er, S.; Suh, C.; Marshak, M. P.; Aspuru-Guzik, A. Computational Design of Molecules for an All-Quinone Redox Flow Battery. *Chemical science* 2015, 6 (2), 885–893.
- (20) Li, J.; Fu, J.; Huang, X.; Lu, D.; Wu, J. Predicting Hydration Free Energies of Amphetamine-Type Stimulants with a Customized Molecular Model. *Journal of Physics: Condensed Matter* 2016, 28 (34), 344001.
- (21) Li, J.; Xu, H.; Wang, J.; Wang, Y.; Lu, D.; Liu, J.; Wu, J. Theoretical Insights on the Hydration of Quinones as Catholytes in Aqueous Redox Flow Batteries. *Chinese Journal of Chemical Engineering* 2021, 37, 72–78.
- (22) Li, J.; Wang, J.; Wang, Y.; Lu, D.; Wu, J. A Multiscale Procedure for Predicting the Hydration Free Energies of Polycyclic Aromatic Hydrocarbons. *Journal of Chemical & Engineering Data* 2020, 65 (4), 2206–2211.
- (23) Liu, Y.; Zhao, S.; Wu, J. A Site Density Functional Theory for Water: Application to Solvation of Amino Acid Side Chains. *Journal of chemical theory and computation* 2013, 9 (4), 1896–1908.
- (24) Liu, Y.; Fu, J.; Wu, J. High-Throughput Prediction of the Hydration Free Energies of Small Molecules from a Classical Density Functional Theory. *The Journal of Physical Chemistry Letters* 2013, 4 (21), 3687–3691.
- (25) Berendsen, H.; Grigera, J.; Straatsma, T. The Missing Term in Effective Pair Potentials. *Journal of Physical Chemistry* 1987, 91 (24), 6269–6271.

- (26) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *Journal of computational chemistry* 2002, 23 (16), 1623–1641.
- (27) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *Journal of computational chemistry* 2004, 25 (9), 1157–1174.
- (28) Neese, F. The ORCA Program System. Wiley Interdisciplinary Reviews: Computational Molecular Science 2012, 2 (1), 73–78.
- (29) Chen, T.; He, T.; Benesty, M.; Khotilovich, V.; Tang, Y.; Cho, H.; Chen, K. Xgboost: Extreme Gradient Boosting. *R package version 0.4-2* 2015, 1 (4), 1–4.
- (30) Pesantez-Narvaez, J.; Guillen, M.; Alcañiz, M. Predicting Motor Insurance Claims Using Telematics Data—XGBoost versus Logistic Regression. *Risks* 2019, 7 (2), 70.
- (31) Pan, B. Application of XGBoost Algorithm in Hourly PM_{2.5} Concentration Prediction; IOP publishing, 2018; Vol. 113, p 012127.
- (32) Zhang, X.; Yan, C.; Gao, C.; Malin, B. A.; Chen, Y. Predicting Missing Values in Medical Data via XGBoost Regression. *Journal of healthcare informatics research* 2020, 4 (4), 383–394.
- (33) Xiong, Z.; Wang, D.; Liu, X.; Zhong, F.; Wan, X.; Li, X.; Li, Z.; Luo, X.; Chen, K.; Jiang, H. Pushing the Boundaries of Molecular Representation for Drug Discovery with the Graph Attention Mechanism. *Journal of medicinal chemistry* 2019, 63 (16), 8749–8760.

- (34) Bergstra, J.; Yamins, D.; Cox, D. Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures; PMLR, 2013; pp 115–123.
- (35) Chen, Q.; Gerhardt, M. R.; Hartle, L.; Aziz, M. J. A Quinone-Bromide Flow Battery with 1 W/Cm² Power Density. *Journal of the Electrochemical Society* 2015, *163* (1), A5010.
- (36) Lin, K.; Chen, Q.; Gerhardt, M. R.; Tong, L.; Kim, S. B.; Eisenach, L.; Valle, A. W.; Hardee, D.; Gordon, R. G.; Aziz, M. J. Alkaline Quinone Flow Battery. *Science* 2015, *349* (6255), 1529–1532.

Supporting Information

Calculation of solvation free energy

Solvation-free energy is the most fundamental property of molecules which can represent the actual solubility of the molecule to some extent. It refers to the energy change of the entire system during the process of dissolving solute molecules from vacuum to solvent at a fixed temperature and pressure. The method for calculating the solvation free energy by a combination of quantum mechanical (QM) calculations and classical density functional theory (cDFT) has been established in our previous work^{22–24}. Specifically, the SPC/E²⁵ water model is used for calculation, and the solute structure is determined using the Hartree–Fock (HF) method by energy minimization in a vacuum. The AM1-BCC²⁶ atomic charges are assigned to calculate electrostatic interactions and the GAFF²⁷ force-field parameters are chosen to describe the van der Waals (VDW) interactions.

Calculation of HOMO-LUMO gap

HOMO-LUMO gap refers to the energy difference between HOMO and LUMO in molecular orbital theory, which can be used to measure whether a molecule is easily excited. Thus HOMO-LUMO gap is used to characterize the open-circuit voltage when the substance molecule is used as the active material of the flow battery. All QM calculations are based on the ORCA ab initio quantum chemistry package²⁸.

Machine learning based on Xgboost

To express the molecular structure as digital parameters that can be used by the machine learning model and unify the input features, we used RDKit (version 2017.09.1) python package to calculate 200 two-dimensional molecular descriptors for each molecule, which were used as input features of machine learning model. Xgboost²⁹ (version 1.4.1) is a machine learning algorithm that has been optimized in algorithm and engineering based on the Gradient Boosting Decision Tree (GBDT). By continuously generating new decision trees to fit the residuals of the last prediction, Xgboost's practicality has been verified in many tasks³⁰⁻³².

Machine learning based on Attentive FP

Attentive Fingerprints (FP)³³ is a representation method of a molecular graph network based on an attention mechanism. Unlike machine learning methods such as Xgboost, it represents the entire molecule as a graph. The atoms in the molecule are the nodes of the graph, and the bonds between the atoms are the undirected edges between the nodes. Through the calculation of the surrounding environment of each node atom in the molecular graph (the number and type of neighboring atoms, etc.), the characteristics of each node are obtained, and finally, the characteristics of a molecule as a whole are integrated, which is used to train the deep learning model based on the PyTorch to predict the molecule's various properties. Both the calculation of molecular descriptors and molecular graph networks are carried out based on the SMILES expression of molecules.

Hyperparameter optimization

In this work, we use the Tree Parzen Estimator algorithm in Hyperopt³⁴ to optimize the learning rate, dropout rate, the number of trees, and other parameters in the model. The detailed hyperparameters for optimization are listed in Supplementary Table S1.