Journal of Materials Chemistry A

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Published on 11 April 2024. Downloaded by Changchun Institute of Applied Chemistry, CAS on 5/26/2024 9:15:38 AM.

Check for updates

Cite this: J. Mater. Chem. A, 2024, 12, 12034

Data-driven discovery of carbonyl organic electrode molecules: machine learning and experiment†

Jiayi Du,‡[ab] Jun Guo,‡[ac] Qiqi Sun,[ad] Wei Liu,[a] Tong Liu,*[ab] Gang Huang[ab] and Xinbo Zhang *[ab]

Received 7th January 2024
Accepted 8th April 2024

DOI: 10.1039/d4ta00136b

rsc.li/materials-a

Carbonyl organic electrode molecules have broad prospects for application in lithium-ion batteries due to their environmental friendliness and cost-effective merit. To overcome the drawbacks associated with traditional time-consuming and costly trial and error experiments, herein, high-throughput calculations and machine learning methods have been employed to accelerate the development of high-performance carbonyl organic electrode molecules by evaluating one million molecules. Hierarchical clustering has been introduced into the selection process to find those target molecules and help us eliminate non-ring molecules. As the reduction potential is a crucial factor in evaluating the performance of electrode materials, based on the created dataset of organic electrode molecules by high-throughput calculations, we have built a machine learning model whose coefficient of determination can reach 0.88 for predicting the reduction potential. With the above efforts, naphthalene-1,4,5,8-tetraone with high reduction potential and energy density has been screened out and indeed exhibits a long cycle life of 2500 cycles at 1 A g^{-1} and a high discharge voltage of 2.5 V. The approach developed in this work offers new insight to filter advanced organic electrode molecules accurately and rapidly for Li-ion batteries.

1 Introduction

With the ever-increasing demand for electric vehicles, electronic devices and grid-scale energy storage, the development of lithium-ion batteries (LIBs) reaches an unprecedented level. However, Fe, Co and Ni-based cathode materials of traditional LIBs are confronted with high cost and shortage of resources. By virtue of inexpensive and readily accessible sources as well as environmental friendliness, organic electrode molecules (OEMs) have emerged as a focal point for next-generation electrode materials.[1–3] Generally, OEMs are composed of non-metallic elements (*e.g.*, H, C, N, O).[4] The soft molecular skeleton of OEMs is suitable for flexible electronic devices in comparison with traditional inorganic electrode materials.[5–7] The theoretical capacity (TC, mA h g^{-1}) of OEMs is designable by adjusting the number of redox blocks. For example, cyclohexanehexone (C_6O_6) consisting of six redox blocks (carbonyls) without other supporting blocks served as a cathode material for LIBs with the highest capacity of 902 mA h g^{-1} among carbonyl-based compounds.[8] Besides, the reduction potential (RP) of OEMs can be modified by introducing electron-withdrawing or electron-donating functional groups near the redox blocks, leading to an increase or decrease in the RP.[9] Zheng *et al.* introduced two Br atoms into tetraoxapentacene,[10] yielding discharge plateaus that were notably elevated, at approximately 4.4 V (*vs.* Li^+/Li), and surpassing those of most OEMs. In addition, incorporation of diverse electron-withdrawing heteroaromatic building blocks, particularly pyridazine, into phenanthrenequinone has been observed to enhance the discharge potential through density functional theory (DFT) calculations.[11] However, traditional strategies such as experiments and DFT calculations to tailor OEMs are not only time-consuming but also costly. This inefficiency significantly hampers research progress.

In recent years, machine learning (ML), an emerging new interdiscipline, has significantly contributed to the fields of chemistry and materials science,[12–14] offering practical solutions to various challenges in these domains. Furthermore, ML can be employed to generate and explore enormous chemical space among which optimal candidates can be swiftly identified without manual labor.[15,16] For example, Shree Sowndarya S. V. and colleagues have developed a multi-objective reinforcement

[a]State Key Laboratory of Rare Earth Resource Utilization, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun 130022, PR China. E-mail: tongliu@ciac.ac.cn; xbzhang@ciac.ac.cn

[b]School of Applied Chemistry and Engineering, University of Science and Technology of China, Hefei 230026, PR China

[c]School of Materials Science and Engineering, Changchun University of Science and Technology, Changchun 130022, PR China

[d]National & Local United Engineering Laboratory for Power Battery, Department of Chemistry, Northeast Normal University, Changchun, 130024, PR China

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d4ta00136b

‡ These authors contributed equally to this work.

This journal is © The Royal Society of Chemistry 2024

Published on 11 April 2024. Downloaded by Changchun Institute of Applied Chemistry, CAS on 5/26/2024 9:15:38 AM.

learning framework trained with approximately 100 000 quantum chemistry simulations. This framework can simultaneously predict the stability, RP and synthesizability of organic radicals.[17] Wang *et al.* have synthesized three photosensitizers successfully by active learning among 7 million molecules.[18] Therefore, owing to various building blocks, more advanced OEMs can be quickly explored *via* ML instead of chemists' intuition or experience. However, there are few reports on the utilization of ML for developing OEMs.[19–21] On one hand, the electrode preparation process and the operating environment of OEMs differ significantly, making it challenging to establish a unified experiment database. On the other hand, the quantum chemical database of OEMs can be readily implemented according to the uniform standards. Therefore, it is worthwhile to develop a quantum chemical database of OEMs and integrate it using ML to dig out more potential OEMs.

Herein, we have illustrated a large-scale strategy of identifying advanced OEMs for LIBs (Fig. 1). Initially, one million organic molecules have been gathered from PubChem. Then, based on the characteristics of the reported OEMs, 1524 molecules have been selected as the candidates of OEMs with the consideration of the type of atom, number of active sites, and hierarchical clustering algorithm. Subsequently, the RP of 1200 molecules randomly selected from the 1524 molecules has been calculated by high-throughput calculations (HTCs), which serves as the training set for a support vector regression (SVR) ML model. Furthermore, the structure–property relationship mapping RP has further been excavated from the ML data, incorporating descriptors such as the number of conjugated double bonds, to define ideal OEMs. From the predictions of the remaining molecules, naphthalene-1,4,5,8-tetraone (NT) with high RP and TC has been singled out for application in LIBs, exhibiting high energy density and long cycle life.

2 Results and discussion

2.1 Collecting data

In the past decades, considerable papers have reported on organic electrode materials. Hence, we gathered 137 OEMs from hundreds of high-quality papers with comparatively complete experiment data on the Web of Science, using the keyword 'organic electrode'. Most of these materials indicated a capacity of less than 400 mA h g^{-1} and a voltage of less than 3 V (Fig. 2a). Further analysis revealed that 119 of these are organic conjugated carbonyl molecules, forming the carbonyl molecule dataset (CMDS), which constitutes 86.86% of the 137 OEMs (Fig. 2b). It is obvious that carbonyl-based OEMs are star materials and possessed of the merits of high capacity and fast kinetics. Therefore, in this work, our focus is also directed towards carbonyl-based OEMs. As the OEMs in the CMDS contain non-metal heavy atoms such as C, N, and O, along with a few S and Cl atoms (Fig. 2c), most molecules in CMDS exhibit a relative molecular mass of less than 500 (Fig. 2d), and the ratio of carbonyl bonds is less than 20% (Fig. 2e). These molecules also feature a high ratio of aromatic bonds and heteroatoms (Fig. S1†).

2.2 Screening database

Besides collecting labeled OEMs from the *Web of Science*, we have gathered one million organic molecules from PubChem[22]

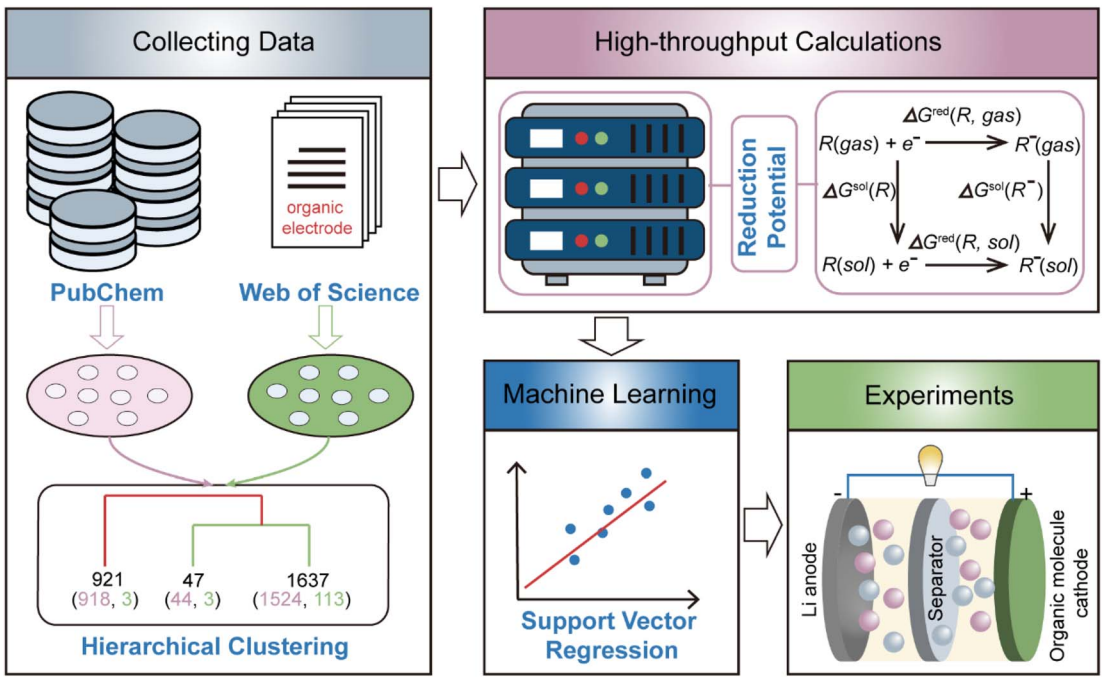

Fig. 1 Overview of the high energy density OEMs found by HTCs and ML from PubChem. First, the OEMs collected from PubChem are screened. Then, the RP of 1524 OEMs is calculated. After this, the machine learning model is built utilizing the results of high-throughput calculation. Finally, an organic molecule is selected to serve as the cathode for LIBs.

Published on 11 April 2024. Downloaded by Changchun Institute of Applied Chemistry, CAS on 5/26/2024 9:15:38 AM.

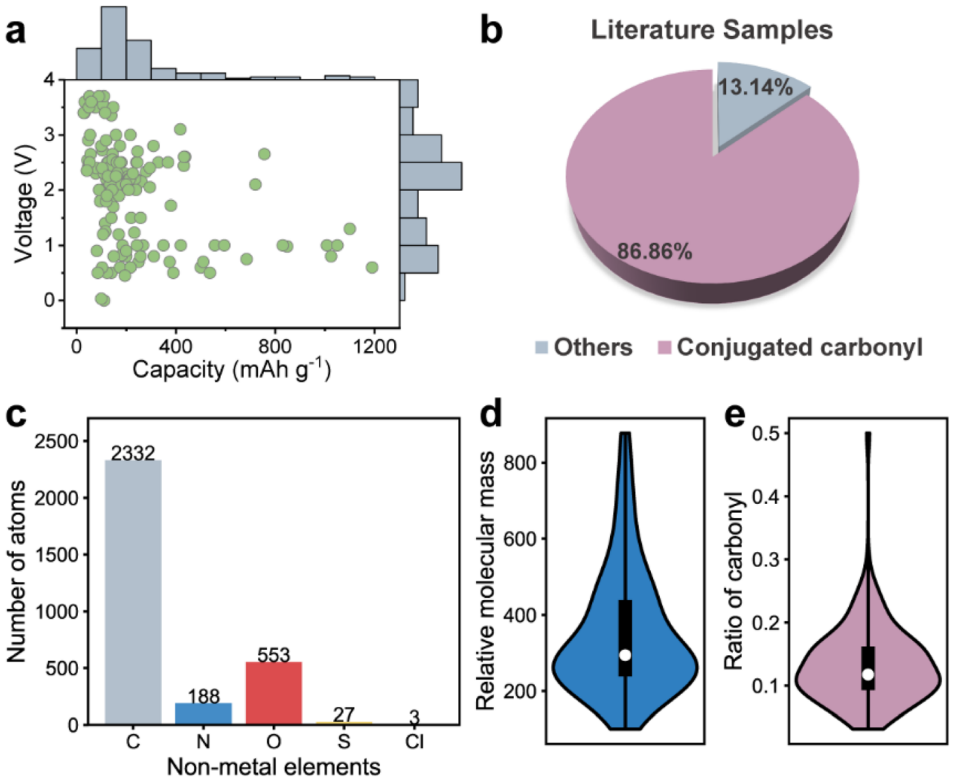

Fig. 2 Analysis of the collected OEM dataset from the reported papers. (a) Voltage and capacity of OEMs. (b) The proportion of the molecules containing conjugated carbonyl groups in the 137 OEMs. (c) The element distribution of 119 organic conjugated carbonyl molecules. The distribution of relative molecular mass (d) and the carbonyl bond ratio (e) in CMDS.

to explore unexploited OEMs (more details are shown in ESI Note 1†).

However, the sheer volume of one million organic molecules made it impractical for manual recognition of those with the potential to become OEMs. Therefore, we set up screening conditions to increase the proportion of effective OEMs. As shown in Fig. 3a, we initially selected molecules containing only H, C, N, O, S, and Cl atoms, and these molecules were required to possess conjugated carbonyl. Then, referring to Fig. 2a, we identified that only a few molecules displayed a capacity of more than 400 mA h g^{-1}. Consequently, our focus shifted to molecules with high TC exceeding 400 mA h g^{-1} (more details about the calculating method of TC are shown in ESI Note 2†). Sequentially, after (i), (ii) and (iii) stages of selection by RDKit,[23] used for identifying elements and functional groups and counting the conjugated carbonyl bonds, a total of 2483 molecules, designed as PubChem Dataset (PubDS), met our previous conditions. However, it is conceivable that not all 2483 molecules can serve as OEMs. Therefore, additional measures should be taken to narrow the range of potential OEMs to reduce time and cost. At this point, we anticipate finding internal connections among OEMs in CMDS so that molecules in PubDS similar to OEMs can be screened out.

Clustering methods are usually used to find the similarity of unlabelled data and group them accordingly. Zhang *et al.* utilized modified X-ray diffraction on crystals to categorize Li-containing compounds into 7 clusters, identifying lithium ion conductors in the IV and V clusters with higher conductivities.[24] In addition, clustering methods are also found to have extensive use in the realm of organic molecules. Based on Simplified Molecular Input Line Entry System (SMILES) sequences,[25] the clustering method can provide effective discrimination for both hydrophobic and hydrophilic monomers.[26] In this work, hierarchical clustering is leveraged to further screen out those molecules with more promise to become OEMs.

Herein, PubDS and CMDS were combined to create a new carbonyl molecule database, referred to as CMDB. Initially, the molecules in CMDB, stored in SMILES format, were converted into another computer-readable MOL format. Subsequently, Morgan Fingerprints (MF),[27] which contained molecular structure information of the local environment, were extracted from these molecules to serve as input for hierarchical clustering. The radius and n_{bits} were set to 2 and 2048, respectively, by RDKit. The Euclidean distance was used to calculate MF during hierarchical clustering, and the complete dendrogram is shown in Fig. S2.† Furthermore, the optimal number of clusters, as determined by both the elbow method and Calinski Harabasz method, was found to be three (Fig. 3b). The three clusters (cluster I, II, and III) contained 921, 44, and 1637 molecules, respectively (Fig. 3c). The distribution of CMDS and PubDS within these three clusters is displayed in Fig. 3d and e. Notably, almost all CMDS molecules are concentrated in cluster III, and 1524 molecules from PubDS are also located in cluster III. To some extent, this result indeed demonstrates the good internal

 This journal is © The Royal Society of Chemistry 2024

Published on 11 April 2024. Downloaded by Changchun Institute of Applied Chemistry, CAS on 5/26/2024 9:15:38 AM.

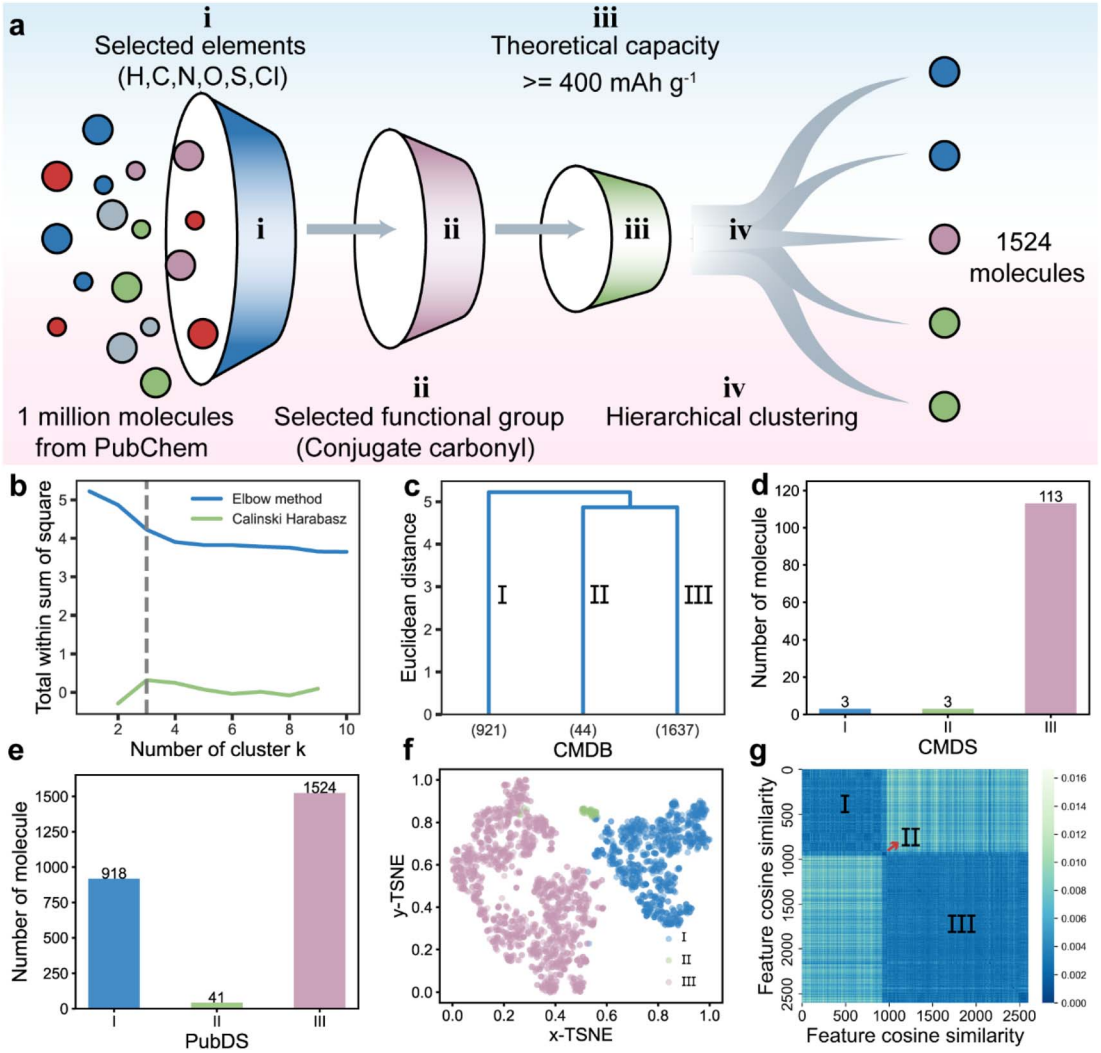

Fig. 3 The screening process and the analysis of hierarchical clustering. (a) The scheme of the screening process with four stages. (b) The evaluation result of the optimal numbers of clustering. (c) The dendrogram of hierarchical clustering at the optimal clustering number 3. The molecules of CMDS (d) and PubDS (e) distributed at three clusters, I, II, and III. (f) Visualization of CMDB molecular space at the optimal clustering number 3 by t-SNE. (g) Cosine similarity matrix between molecules in CMDB.

similarity of CMDS *via* hierarchical clustering. Therefore, the 1524 molecules in cluster III exhibit significant potential to serve as OEMs. Simultaneously, we have endeavoured to interpret the reasons behind the classification. The t-Distributed Stochastic Neighbour Embedding (t-SNE) was implemented to visualize the two-dimensional molecular space of CMDB, and three separated clusters confirmed the rationality of hierarchical clustering (Fig. 3f). Besides, in Fig. 3g, three clusters were located in three deeply blue square areas, indicating that molecules in each cluster were closer by cosine matrix and had more similar MFs. Therefore, both methods showed that the 1524 molecules had more similar features with OEMs in CMDS. Last but not least, we also display some molecules from each cluster in Fig. S3.† It is apparent that clusters I, II, and III primarily consist of branch chain molecules, straight chain molecules, and ring molecules, respectively, corresponding to Fig. 3f and g. In brief, hierarchical clustering successfully divides PubDS into three parts, among which cluster III is more similar to OEMs in CMDS, aiding in narrowing down the potential OEMs and identifying the 1524 molecules.

2.3 High-throughput calculations

To pursue the OEMs with high theoretical energy density, 1200 molecules were randomly selected from 1524 molecules and high-throughput calculations (HTCs) were carried out to determine their RP (ESI Note 3†). Every molecule underwent calculations in four states, as illustrated in Fig. S4,† resulting in a total of 4800 tasks. Finally, we obtained the RP values for 1003 molecules, while 197 molecules could not be calculated due to optimization failures. The results of RP from HTCs are shown in Fig. 4a. Examining the distribution of relative molecular mass (Fig. 4b), these selected molecules, on the whole, are characterized by their small size. The TC of most molecules ranges between 400 and 600 mA h g^{-1} (Fig. 4c). From the RP perspective,

Published on 11 April 2024. Downloaded by Changchun Institute of Applied Chemistry, CAS on 5/26/2024 9:15:38 AM.

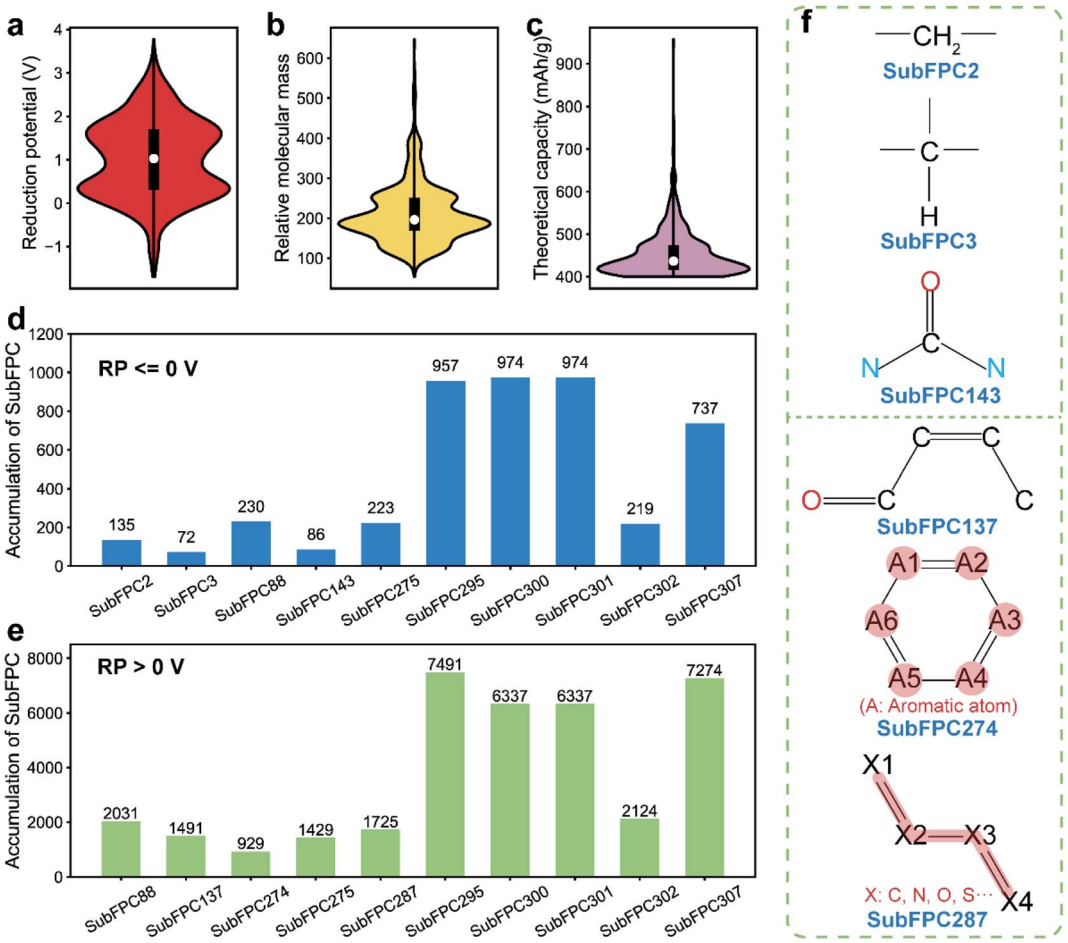

Fig. 4 The results of high-throughput calculations. The violin plot of RP (a), relative molecular mass (b), and TC (c). The top 10 functional groups according to the accumulative frequency of 307 functional groups from SubFPC for molecules with RP ≤ 0 V (d) and molecules with RP > 0 V (e). (f) The schematics of SubFPC2, SubFPC3, SubFPC143, SubFPC137, SubFPC274, and SubFPC287.

some molecules have extremely low RP, even less than 0 V, making them unsuitable candidates for OEMs. Hence, the Substructure Fingerprint Count (SubFPC)[28] descriptor, composed of 307 sub-structures, is used to investigate the difference between molecules with RP > 0 V and RP ≤ 0 V. The cumulative frequency of 307 sub-structures is calculated, and the top 10 sub-structures are exhibited in Fig. 4d and e. Among them, SubFPC88, SubFPC275, SubFPC295, SubFPC300, SubFPC301, SubFPC302, and SubFPC307 account for a high proportion in both types of molecules. However, SubFPC2, SubFPC3, and SubFPC143 have a higher proportion in the molecules with RP ≤ 0 V, while SubFPC137, SubFPC274, and SubFPC287 possess a higher proportion in the molecules with RP > 0 V. From Fig. 4f, it can be observed that the SubFPC2 and SubFPC3 are alkyl chains, and SubFPC143 represents the carbonic acid derivatives. In contrast, the SubFPC137 stands for a conjugated double bond with the carbonyl group, SubFPC274 is an aromatic atom, and SubFPC287 denotes a conjugated double bond. The more accurate definition of these six functional groups is listed in Table S1.† By comparison, it is found that the presence of the conjugated double bond indeed plays an important role in RP and endows OEMs with stronger reducibility.

2.4 Machine learning

Based on HTCs, we attempted to build a machine learning model for more efficiently predicting the RP of OEMs without DFT calculations. First of all, appropriate descriptors should be considered to express molecules' inherent features. SubFPC and Substructure Fingerprint (SubFP) were used to count the sub-structures. As outlined in Section 2.3, most functional groups among 307 sub-structures were uncommon. Therefore, we removed those sub-structures that were not present in 90% of the molecules, retaining 22 sub-structures. In addition, the global descriptors such as Extended Topochemical Atom (ETA, 43 sub-descriptors), Topological Charge (TopoC, 21 sub-descriptors), and BCUT (6 sub-descriptors) generated by Padel,[28] along with 7 single descriptors generated by RDKit (Table S2†) were also included. In a word, a total of 121 descriptors were constructed. Specifically, the ETA is related to the molecular size, electronegativity of atoms, and electronic contribution. TopoC is the charge distribution on the topological molecular surface, and BCUT is the bond-type combination of adjacent and nonadjacent atoms.

This journal is © The Royal Society of Chemistry 2024

Published on 11 April 2024. Downloaded by Changchun Institute of Applied Chemistry, CAS on 5/26/2024 9:15:38 AM.

To understand the relationship between the 121 descriptors and RP, we quantified their dependency based on the mutual information algorithm.[29] A higher Mutual Information Value (MIV) indicated a stronger dependency. Therefore, according to MIV of the 121 descriptors, useless descriptors could be removed to reduce the dimensionality of feature and enhance the performance of ML. Furthermore, the SVR served as the initial training model to reduce the number of descriptors according to the ranking of MIV for the 121 descriptors. The performance of the SVR model was evaluated using the 5-fold cross-validation (Fig. S5†). During the ML training process, the values of descriptors were normalized. When the first 48 descriptors were selected, the R^2 reached the highest score of 0.812 ± 0.027. Finally, these 48 descriptors remained (Table S3†). Moreover, Pearson correlation coefficient matrix (PCCM) between the selected 121 descriptors clearly shows the positive and negative correlations (Fig. S6†). After the mutual information method of reducing dimensionality, based on the PCCM of finally selected 48 descriptors (Fig. S7†), it is obvious that the successful removal of redundant and irrelevant descriptors with low correlation contributes to an improved score of SVR.

We then studied three ML models, SVR, Multi-Layer Perceptron (MLP), and XGBoost (XGB) to predict the RP. The optimal parameters of these three models were determined through the grid-search method and 5-fold cross-validation. The learning curves of these three models are shown in Fig. S8.† When the training set and test set were split into 8 : 2, the R^2 scores for the test set were 0.882, 0.843, and 0.819 for SVR, MLP, and XGB, respectively (Fig. 5a–c). Also, it was clear that the curves of training and test of SVR were higher when the training size was beyond 0.6, illustrating the better performance of SVR. In addition, the test set of SVR had a smaller mean absolute error (MAE) of 0.243 and root mean square error (RMSE) of 0.334 (Table S4†). Therefore, the SVR model was more appropriate to predict the RP in our work.

Among these 48 descriptors, ETA accounts for 45.8% of them, with half of its sub-descriptors retained following the mutual information selection process (Fig. 5d). ETA comprises recording of various expressions like related to size or bulk, electronegativity, and the electronic contribution of atoms,[30] making it possible to accurately describe the electron structure and atomic composition of OEMs. The first MIV and second MIV are SubFPC287 and ETA_dBetaP, respectively. Fig. 5e illustrates that more conjugated double bonds enhance the RP, which is consistent with the finding in Section 2.3. As for ETA_dBetaP (Fig. 5f), it is determined by the following equation:

$$\mathrm{ETA}_{\mathrm{dBetaP}} = \frac{\sum \beta_{\mathrm{ns}} - \sum \beta_{\mathrm{s}}}{N_{\mathrm{V}}} \quad (1)$$

where β_{ns} represents the non-sigma electron, β_{s} stands for the sigma electron, and N_{V} denotes the molecular volume. The sigma electron originates from the σ-bond, so ETA_dBetaP also demonstrates that an increase in single bonds corresponds to a decrease in RP. Other partial descriptors are shown in Fig. S9.† Furthermore, individual condition expectation (ICE) is introduced to show how different samples are affected by those descriptors (Fig. S10†). Taking SubFPC287 as an example, all samples show an upward trend with increasing SubFPC287, indicating that the conjugated double bond consistently contributes positively to RP. However, high RP molecules exhibit minimal change with ETA_dBetaP, which means that increasing ETA_dBetaP is not an effective way to improve the RP for high RP molecules. Similar patterns could also be observed for other descriptors such as ETA_BetaP_ns ($\sum\beta_{\mathrm{ns}}/N_{\mathrm{V}}$),

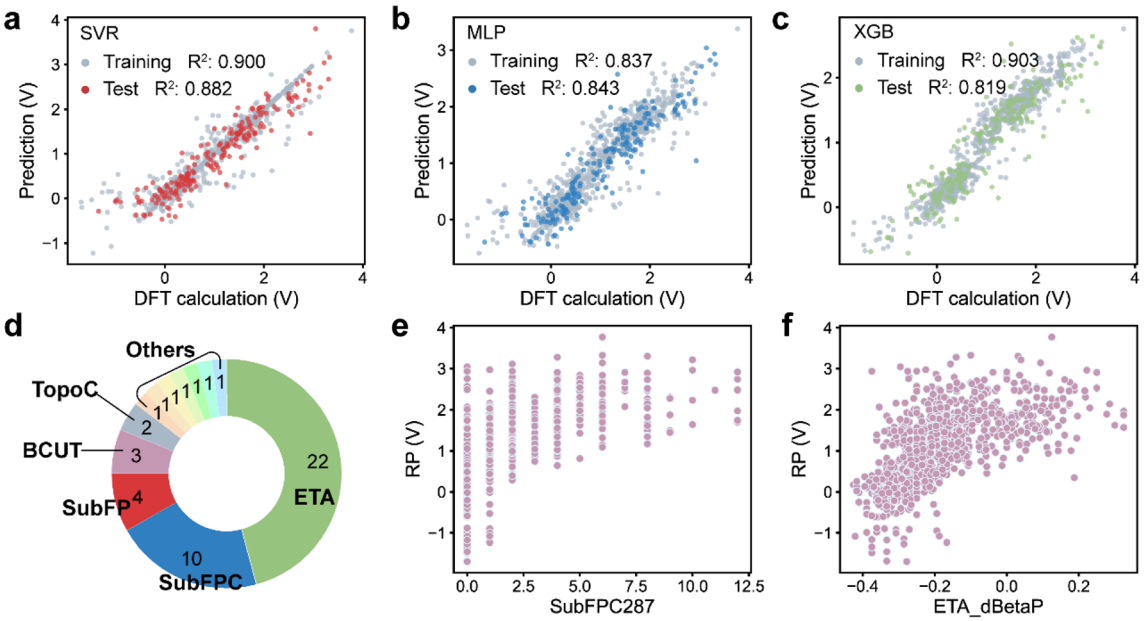

Fig. 5 The descriptors and machine learning models. The prediction results of SVR (a), MLP (b), and XGB (c) models. (d) The distribution of 48 descriptors after the selection of mutual information. (e) The relationship between the SubFPC287 and RP. (f) The relationship between the ETA_dBetaP and RP.

Published on 11 April 2024. Downloaded by Changchun Institute of Applied Chemistry, CAS on 5/26/2024 9:15:38 AM.

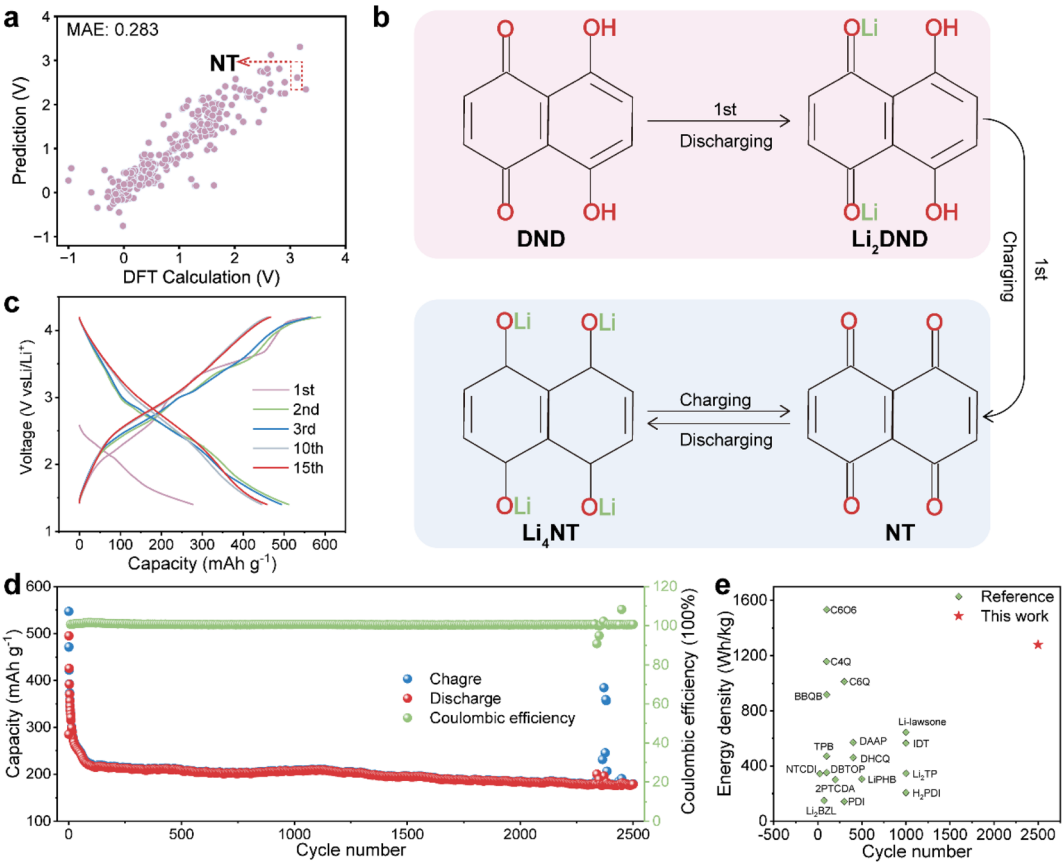

Fig. 6 Electrochemical performance of NT. (a) The validation of the remaining 279 molecules using the SVR model and HTCs. (b) Schematic diagram of the charging and discharging process of NT. (c) The full discharge and charge curves of the 1st, 2nd, 3rd, 10th, and 15th cycles at 0.1 A g^{-1}. (d) The long-term cycling performance of NT at 1 A g^{-1}. (e) Performance comparison of NT with the reported organic carbonyl cathode materials.

max partial charge (Gasteiger charges), and ETA_dBeta ($\sum\beta_{ns} - \sum\beta_s$). Specifically, SubFPC137 has a negative effect on high RP molecules. Therefore, by analyzing the ICE of descriptors, it is clear that adjusting these descriptors can guide the design of OEMs with high RP.

2.5 Experimental

To further verify the accuracy of the SVR model, it was employed to predict the remaining 324 molecules from the above 1524 molecules (Section 2.3), and the RP of these molecules was calculated by HTCs. The obtained RP of 279 molecules and the prediction results with MAE of 0.283 are shown in Fig. 6a. Notably, the molecule with the highest theoretical energy density in Fig. 6a, highlighted in the red box, is NT. Following this, NT was selected out for successful application in a LIB cathode. First, 5,8-dihydroxynaphthalene-1,4-dione (DND) as a precursor of NT was synthesized (Fig. S11 and S12†). Subsequently DND was made for the cathode material (ESI Note 5†). After the first cycle discharge and charge in LIBs, NT was formed (Fig. 6b). During the first cycle of discharge, the discharge capacity was only 277 mA h g^{-1} at 0.1 A g^{-1} (Fig. 6c), which was close to the theoretical capacity of 282 mA h g^{-1} of DND. Then, in the second discharge cycle, the capacity increased to 511 mA h g^{-1} approaching the theoretical capacity of 570 mA h g^{-1}. This illustrates the successful conversion of the DND electrode into NT after the first charge process. X-ray photoelectron spectroscopy (XPS) spectra further confirms this conversion (Fig. S13†). In addition, NT exhibits excellent cycling performance of 2500 cycles at 1 A g^{-1}, reaching a discharge voltage of up to 2.5 V. Moreover, compared with other organic carbonyl cathode materials, NT exhibits a pretty high energy density and cycle life[8,10,31–45] (Fig. 6e, Table S5†). The successful application of NT in a LIB cathode suggests the viability of our method in exploring new OEMs.

3 Conclusion

In summary, we have collected one million molecules from PubChem and after four rounds of filtration we identified 1524 molecules as potential OEMs. Throughout the screening process, hierarchical clustering has been introduced for the first time, and the reliability of classification has been confirmed by the t-SNE and cosine matrix analyses. This innovative approach enables the rapid and large-scale screening of potential OEMs. Moreover, employing HTCs, the SVR model with high accuracy in predicting RP has been built. Upon analysing the SVR

 This journal is © The Royal Society of Chemistry 2024

Published on 11 April 2024. Downloaded by Changchun Institute of Applied Chemistry, CAS on 5/26/2024 9:15:38 AM.

descriptors, double bonds, especially the conjugated double bonds, significantly contribute to the RP of OEMs. After the above steps, we have selected the promising molecule NT as the cathode material of LIBs, achieving an impressive energy density of ~1200 W h kg^{-1} and verifying the good prediction performance of the SVR model. The approach reported in this work could efficiently screen and design an extensive array of organic cathode materials, showcasing the potential of data-driven method for accelerating the development of advanced electrode materials.

Data availability

The datasets generated during this study and ML codes are available from **https://zenodo.org/records/10828373**.

Author contributions

J. Du: conceptualization, investigation, formal analysis, and writing – original draft. J. Guo: resources, investigation, formal analysis. W. L.: methodology, writing – review & editing. T. Liu: writing – review & editing. G. Huang: conceptualization, methodology. X. Zhang: conceptualization, supervision.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was financially supported by the National Key Research and Development Program of China (2022YFB2402200), National Natural Science Foundation of China (52171194, 52071311), Jilin Province Science and Technology Development Plan Funding Project (Grant YDZJ202301-ZYTS545), Youth Innovation Promotion Association CAS (Grant 2020230), and National Natural Science Foundation of China Outstanding Youth Science Foundation of China (Overseas). The Supercomputing USTC is acknowledged for computational support.

References

1 Y. Lu, Q. Zhang, L. Li, Z. Niu and J. Chen, *Chem*, 2018, **4**, 2786–2813.
2 L. Liu, L. Miao, L. Li, F. Li, Y. Lu, Z. Shang and J. Chen, *J. Phys. Chem. Lett.*, 2018, **9**, 3573–3579.
3 Z. Lin, H. Y. Shi, L. Lin, X. Yang, W. Wu and X. Sun, *Nat. Commun.*, 2021, **12**, 4424.
4 X. Chen, X. Yin, J. Aslam, W. Sun and Y. Wang, *Electrochem. Energy Rev.*, 2022, **5**, 12.
5 Z. Wu, Q. Liu, P. Yang, H. Chen, Q. Zhang, S. Li, Y. Tang and S. Zhang, *Electrochem. Energy Rev.*, 2022, **5**, 26.
6 J. Peng, D. Wu, H. Li, L. Chen and F. Wu, *Battery Energy*, 2023, **2**, 20220059.
7 T. Yang, Y. Niu, Q. Liu and M. Xu, *Nano Mater. Sci.*, 2023, **5**, 119–140.
8 Y. Lu, X. Hou, L. Miao, L. Li, R. Shi, L. Liu and J. Chen, *Angew. Chem., Int. Ed.*, 2019, **58**, 7020–7024.
9 G. Kwon, Y. Ko, Y. Kim, K. Kim and K. Kang, *Acc. Chem. Res.*, 2021, **54**, 4423–4433.
10 Y. Zheng, H. Ji, J. Liu, Z. Wang, J. Zhou, T. Qian and C. Yan, *Nano Lett.*, 2022, **22**, 3473–3479.
11 L. B. Zhao, S. T. Gao, R. He, W. Shen and M. Li, *ChemSusChem*, 2018, **11**, 1215–1222.
12 P. Raccuglia, K. C. Elbert, P. D. Adler, C. Falk, M. B. Wenny, A. Mollo, M. Zeller, S. A. Friedler, J. Schrier and A. J. Norquist, *Nature*, 2016, **533**, 73–76.
13 M. Zhong, K. Tran, Y. Min, C. Wang, Z. Wang, C. T. Dinh, P. De Luna, Z. Yu, A. S. Rasouli, P. Brodersen, S. Sun, O. Voznyy, C. S. Tan, M. Askerka, F. Che, M. Liu, A. Seifitokaldani, Y. Pang, S. C. Lo, A. Ip, Z. Ulissi and E. H. Sargent, *Nature*, 2020, **581**, 178–183.
14 T. Xie and J. C. Grossman, *Phys. Rev. Lett.*, 2018, **120**, 145301.
15 J. Peng, D. Schwalbe Koda, K. Akkiraju, T. Xie, L. Giordano, Y. Yu, C. J. Eom, J. R. Lunger, D. J. Zheng, R. R. Rao, S. Muy, J. C. Grossman, K. Reuter, G. B. Rafael and S. H. Yang, *Nat. Rev. Mater.*, 2022, **7**, 991–1009.
16 Z. Rao, P. Tung, R. Xie, Y. Wei, H. Zhang, A. Ferrari, T. P. C. Klaver, F. Körmann, P. T. Sukumar, A. K. d. Silva, Y. Chen, Z. Li, D. Ponge, J. Neugebauer, O. Gutfleisch, S. Bauer and D. Raabe, *Science*, 2022, **378**, 78–85.
17 S. V. Shree Sowndarya, J. N. Law, C. E. Tripp, D. Duplyakin, E. Skordilis, D. Biagioni, R. S. Paton and P. C. St John, *Nat. Mach. Intell.*, 2022, **4**, 720–730.
18 S. Xu, J. Li, P. Cai, X. Liu, B. Liu and X. Wang, *J. Am. Chem. Soc.*, 2021, **143**, 19769–19777.
19 K. Sakano, Y. Igarashi, H. Imai, S. Miyakawa, T. Saito, Y. Takayanagi, K. Nishiyama and Y. Oaki, *ACS Appl. Energy Mater.*, 2022, **5**, 2074–2082.
20 R. P. Carvalho, C. F. N. Marchiori, D. Brandell and C. M. Araujo, *Energy Storage Mater.*, 2022, **44**, 313–325.
21 O. Allam, R. Kuramshin, Z. Stoichev, B. W. Cho, S. W. Lee and S. S. Jang, *Mater. Today Energy*, 2020, **17**, 100482.
22 S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B. A. Shoemaker, P. A. Thiessen, B. Yu, L. Zaslavsky, J. Zhang and E. E. Bolton, *Nucleic Acids Res.*, 2021, **49**, D1388–D1395.
23 G. A. Landrum, *RDKit: Open-Source Cheminformatics*, 2014, **https://www.rdkit.org**.
24 Y. Zhang, X. He, Z. Chen, Q. Bai, A. M. Nolan, C. A. Roberts, D. Banerjee, T. Matsunaga, Y. Mo and C. Ling, *Nat. Commun.*, 2019, **10**, 5260.
25 D. Weininger, *J. Chem. Inf. Comput. Sci.*, 1988, **28**, 31–36.
26 L. Liu, W. Chen, T. Liu, X. Kong, J. Zheng and Y. Li, *J. Mater. Chem. A*, 2019, **7**, 11847–11857.
27 D. Rogers and M. Hahn, *J. Chem. Inf. Model.*, 2010, **50**, 742–754.
28 C. W. Yap, *J. Comput. Chem.*, 2011, **32**, 1466–1474.
29 A. Kraskov, H. Stögbauer and P. Grassberger, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2004, **69**, 066138.
30 K. Roy and R. N. Das, *SAR QSAR Environ. Res.*, 2011, **22**, 451–472.

Published on 11 April 2024. Downloaded by Changchun Institute of Applied Chemistry, CAS on 5/26/2024 9:15:38 AM.

31 W. Huang, X. Zhang, S. Zheng, W. Zhou, J. Xie, Z. Yang and Q. Zhang, *Sci. China Mater.*, 2019, **63**, 339–346.
32 L. Chen, S. Liu, L. Zhao and Y. Zhao, *Electrochim. Acta*, 2017, **258**, 677–683.
33 S. Zhang, S. Ren, D. Han, M. Xiao, S. Wang and Y. Meng, *J. Power Sources*, 2019, **438**, 227007.
34 W. Huang, M. Zhang, H. Cui, B. Yan, Y. Liu and Q. Zhang, *Chem.–Asian J.*, 2019, **14**, 4164–4168.
35 J. Yang, P. Xiong, Y. Shi, P. Sun, Z. Wang, Z. Chen and Y. Xu, *Adv. Funct. Mater.*, 2020, **30**, 109597.
36 J. Lee and M. J. Park, *Adv. Energy Mater.*, 2017, **7**, 1602279.
37 Z. Luo, L. Liu, Q. Zhao, F. Li and J. Chen, *Angew. Chem., Int. Ed.*, 2017, **56**, 12561–12565.
38 V. Medabalmi and K. Ramanujam, *J. Electrochem. Soc.*, 2017, **164**, A1720–A1725.
39 H. Zhang, R. Zhang, F. Ding, C. Shi and N. Zhao, *Energy Storage Mater.*, 2022, **51**, 172–180.
40 S. More, N. Khupse, M. Bhosale, J. Ambekar, M. Kulkarni and B. Kale, *ChemistrySelect*, 2020, **5**, 2157–2163.
41 H. Peng, P. Chen, X. Yang, Z. Xue, S. Wang, J. Na, J. Yu and Y. Yamauchi, *J. Mater. Chem. A*, 2020, **8**, 11521–11528.
42 Y. Hong, J. Hu, W. Tang, B. Wei, M. Guo, S. Jia and C. Fan, *Energy Storage Mater.*, 2022, **52**, 61–68.
43 M. Rajesh, F. Dolhem, C. Davoisne and M. Becuwe, *ChemSusChem*, 2020, **13**, 2364–2370.
44 V. Medabalmi and K. Ramanujam, *ChemistrySelect*, 2018, **3**, 10657–10662.
45 J. Aher, A. Graefenstein, G. Deshmukh, K. Subramani, B. Krueger, M. Haensch, J. Schwenzel, K. Krishnamoorthy and G. Wittstock, *ChemElectroChem*, 2020, **7**, 1160–1165.

This journal is © The Royal Society of Chemistry 2024