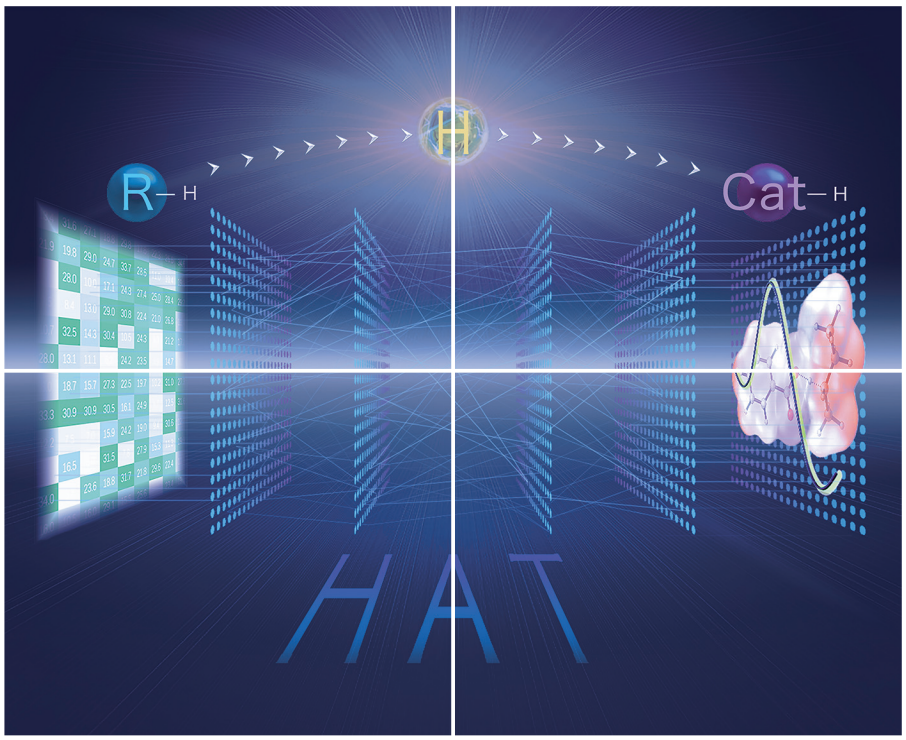
Volume 8 | Number 22 | 21 November 2021



ORGANIC CHEMISTRY

FRONTIERS

CCS 中国化学会 CHINESE CHEMICAL SOCIETY
SIOC 1950
ROYAL SOCIETY OF CHEMISTRY
rsc.li/frontiers-organic

ORGANIC CHEMISTRY FRONTIERS

RESEARCH ARTICLE

View Article Online
View Journal | View Issue

Cite this: *Org. Chem. Front.*, 2021, **8**, 6187

Machine learning prediction of hydrogen atom transfer reactivity in photoredox-mediated C–H functionalization†

Li-Cheng Yang, Xin Li, Shuo-Qing Zhang and Xin Hong *

Photoredox-mediated hydrogen atom transfer (HAT) catalysis has reshaped the synthetic strategy of C–H bond functionalization. The rationalization and prediction of HAT reactivity are crucial for the reaction design of photoredox-mediated C–H functionalization. In this work, we report the development of a machine learning model that can predict the HAT barrier of photoredox-mediated HAT catalysis using the physical organic descriptors of the ground state substrate and radical. Based on 2926 DFT-computed HAT barriers of the designed chemical space, the trained AdaBoost model is able to predict the HAT barrier with a mean absolute error of 0.60 kcal mol^{-1} in the out-of-sample test set. The applicability of the machine learning model is further validated by comparing the prediction against the DFT-computed reactivities on scaffolds and substituents that are not present in the designed chemical space, as well as experimental kinetics data of HAT reaction with the cumyloxyl radical. This work provides a machine learning approach for reactivity prediction from physical organic descriptors and DFT-computed statistics, offering a useful technique that can be directly applied in the experimental designs of photoredox-mediated HAT catalysis.

Received 6th September 2021,
Accepted 18th September 2021

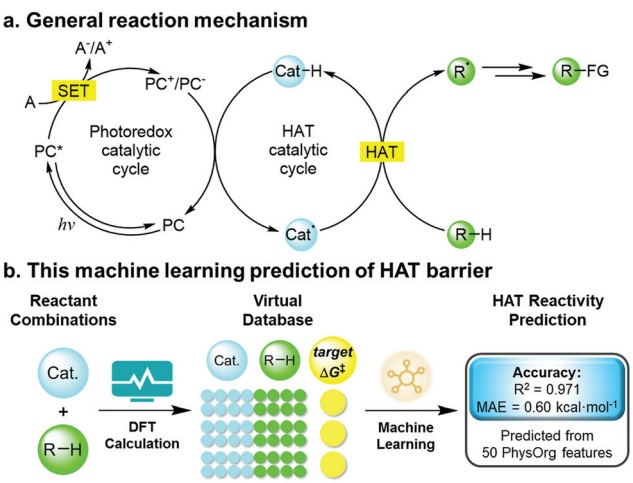
DOI: 10.1039/d1qo01325d

rsc.li/frontiers-organic

Introduction

Photoredox-mediated hydrogen atom transfer (HAT) catalysis has emerged as a powerful strategy to functionalize native C–H bonds. This reaction has revolutionized the synthetic potential of the ubiquitous C–H bonds and breathed new life into the classic HAT transformation.[1–4] Through the fruitful studies, this transformation has reached a remarkable substrate scope, and is able to efficiently functionalize benzylic,[5,6] allylic,[7] aldehyde[8–10] and alkyl[11–14] C–H bonds. The operative mechanism generally involves cooperative photoredox and HAT catalytic cycles,[1–3] which lead to the generation of the key catalytic radical component that cleaves the inert C–H bond *via* a HAT step (Scheme 1a). Therefore, the understanding of the HAT kinetics is essential for the rational design of photoredox-mediated HAT catalysis.

Towards the understanding and prediction of HAT reactivity, fruitful mechanistic studies have been performed over the last few decades.[15–17] The representative strategies include the valence bond model approach developed by Shaik,[18] which can predict the HAT kinetics with bond energy-related parameters, and the Marcus cross relation approach developed by Mayer,[19] which relates the HAT barrier to the kinetics of self-exchange rate constant and thermodynamic equilibrium. In addition to these theoretical frameworks, "polarity match" is a widely applied empirical concept in the HAT design of photoredox-mediated HAT catalysis.[20] Despite these inspiring



Scheme 1 Photoredox-mediated HAT catalysis.

Center of Chemistry for Frontier Technologies, Department of Chemistry, State Key Laboratory of Clean Energy Utilization, Zhejiang University, Hangzhou 310027, China. E-mail: hxchem@zju.edu.cn

†Electronic supplementary information (ESI) available: Detailed computational methods, descriptors, machine learning models and model performance on the test set. See DOI: 10.1039/d1qo01325d

studies that laid the mechanistic basis of HAT reactivity prediction, an efficient and quantitative approach to predict the HAT reactivity in photoredox-mediated HAT catalysis is yet to be achieved.

In light of the stimulating development of machine learning application in organic synthesis, we envisioned that this strategy would serve as a powerful method to connect the readily accessible properties of reactants and the desired information on reaction kinetics. Independent studies by Doyle,[21,22] Denmark,[23] Grzybowski,[24,25] Jensen,[26,27] Glorius,[28] Sigman,[29] Sunoj,[30,31] Norrby,[32] Buttar,[33] Ess[34,35] and others[36,37] revealed the exciting potential of machine learning in the reactivity and selectivity prediction of synthetic transformations. We recently demonstrated that mechanism-based computational statistics can serve as an alternative data source for machine learning of organic synthesis, which led to a regioselectivity prediction model of the radical C–H functionalization of heteroarenes.[38] Based on the computational statistics, herein we report the development of a machine learning model that can predict the HAT barrier from the physical organic features of the substrate and radical (Scheme 1b). The trained AdaBoost model can predict the HAT barrier with a mean absolute error (MAE) of 0.60 kcal mol^{-1} in the out-of-sample test set. The model performance was further validated by the satisfactory comparison with the experimental kinetics data. This work provides a straightforward approach to evaluate the design of potential HAT transformations in the photoredox-mediated C–H functionalization.

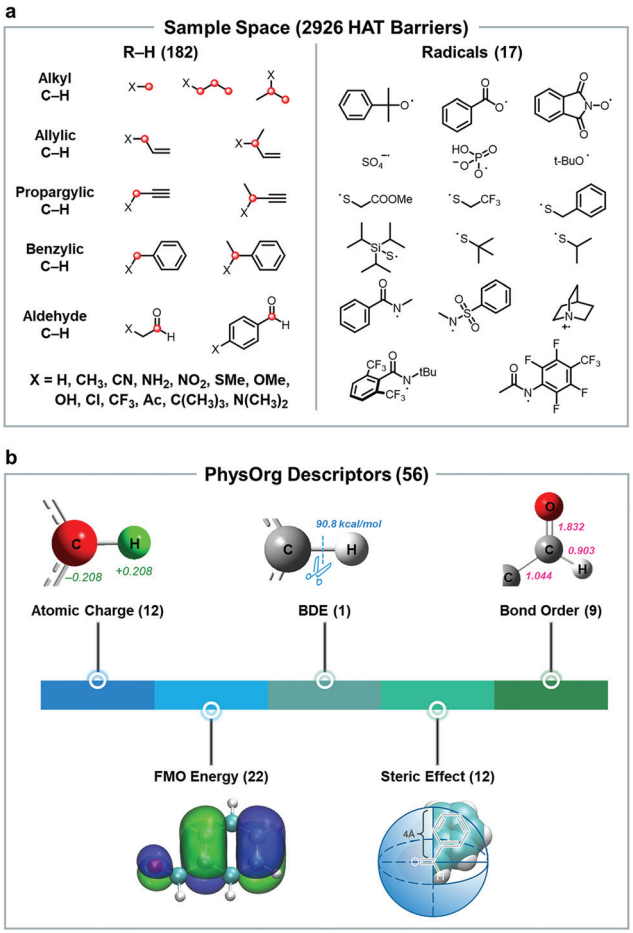


Fig. 1 Designed chemical space and physical organic descriptors for model training of target HAT reaction. (a) Designed chemical space of HAT reaction involves 182 distinctive C–H positions, which arise from 5 types of C–H bonds and 13 kinds of substituents. Each studied C–H position is labeled in red. (b) 56 physical organic descriptors were used to encode the electronic and steric effects of the ground state substrate and radical. These descriptors include five categories, and the number of descriptors in each category is shown in parentheses.

Results and discussion

Building the machine learning model

To describe the structure–activity relationship of the HAT reaction in photoredox-mediated HAT catalysis, we first designed a sample space based on the experimental advances of related transformations.[1–3,16] The selected substrates for C–H bond cleavage cover the general native C–H bonds that are of potential interest for reaction development, which includes the allylic, propargylic, benzylic, aldehyde and the less reactive alkyl C–H bonds (Fig. 1a). The allylic, benzylic, aldehyde and alkyl C–H functionalization has been reported in related experimental studies.[1–3,16] The propargylic C–H scaffold was included considering the chemical similarity. Based on these scaffolds, twelve representative substituents that cover both the electronic and steric effects were considered to support the generalization ability of the machine learning model, which led to a collection of 182 distinctive C–H bond positions for the HAT reaction. For the radical component, 17 radicals were selected including the widely used radical catalysts that are applied in the photoredox-mediated HAT catalysis and several additional radicals that were used in C–H functionalization.[1–3,16] These radical catalysts include the quinuclidine radical cation,[13,39] benzoyloxy radical,[11] and the sulphate radical anion,[40] which were designed by experimentalists to achieve the desired HAT ability and redox potential for the cooperative catalysis. The combinations of the selected C–H substrates and radicals created a collection of 3094 HAT reactions, which would allow the machine learning model to map the variation of the target structure–activity relationship.

We next performed density functional theory (DFT) calculations on the designed sample space to provide the required statistics for model training. The experimental measurement of the kinetic rate of the single HAT step in the photoredox-mediated HAT catalysis is challenging. Thus, there is a lack of related experimental data for the desired chemical space, and we applied the computational approach to obtain the target structure–activity relationship data. The computations were performed at the B3LYP[41,42] level of theory with the 6-31+G(d, p) basis set (computational details are included in the ESI†). The acetonitrile solvent was chosen for the calculation of solvation energy due to its wide applications in the photoredox-mediated HAT catalysis.[1–3] The computed barriers in aceto-

nitrile were also compared with those in acetone and DMSO for a representative subset, and satisfactory linear correlations were identified (Fig. S1†). Therefore, we believe that the computed barriers in acetonitrile and the derived machine learning model are applicable for the prediction of relative HAT reactivities in additional polar solvents. The C–H substrates and radicals were optimized individually, and the subsequent transition state calculations were performed in a semiautomatic fashion with the aid of customized scripts. The detailed flowchart of the computational data generation is provided in the ESI (Fig. S2†). Conformations of the reactants and the HAT transition states were carefully considered to ensure the reliability of the computed kinetics data. 2926 transition states of the 3094 designed HAT reactions were located, which created the dataset for model training.

Physical organic (PhysOrg) descriptors were utilized to encode the ground state reactants. The 56 selected PhysOrg descriptors capture both the local and global properties of the C–H substrate and radical (Fig. 1b). Four categories of descriptors describe the local properties of the reacting position,

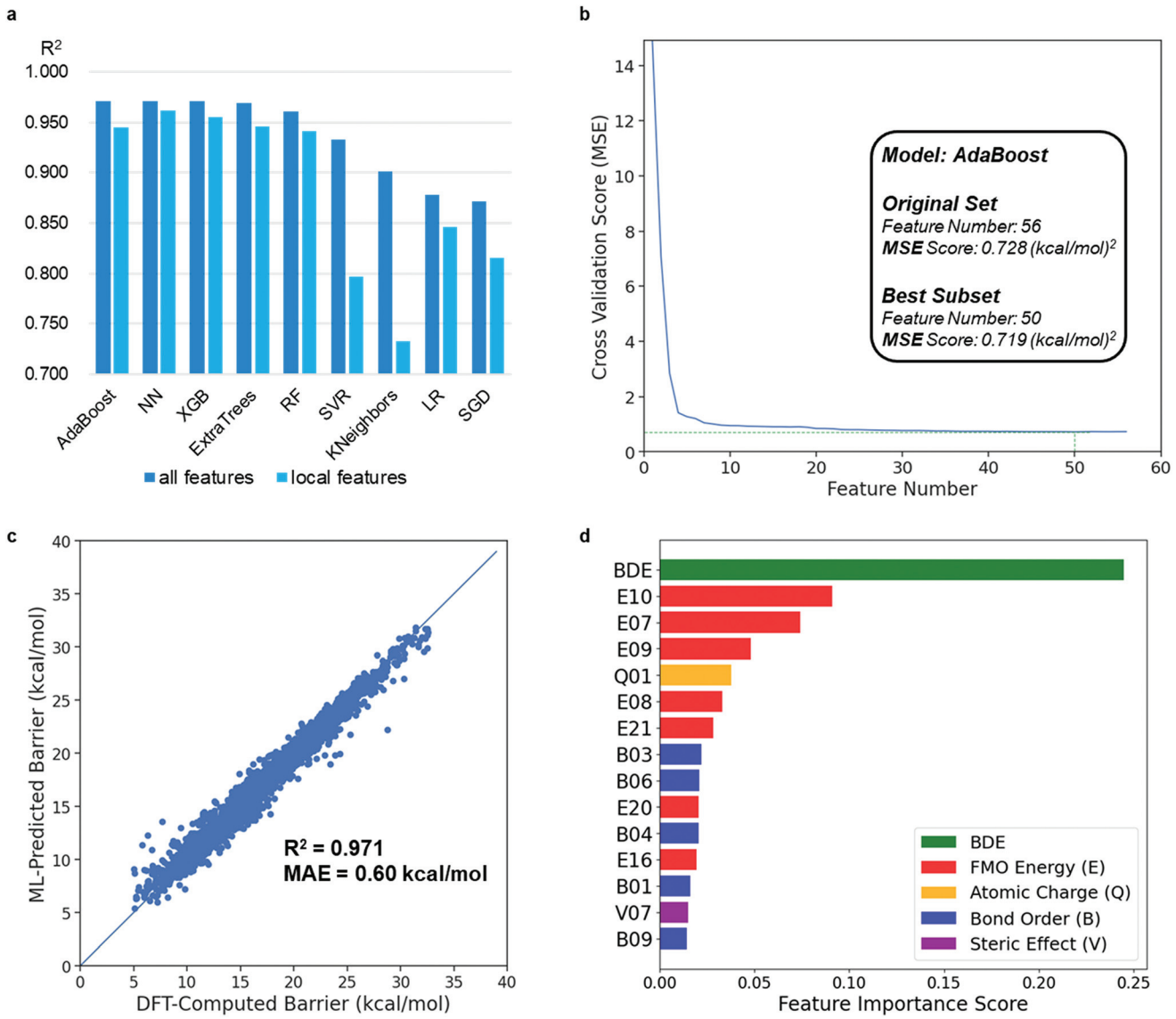


Fig. 2 Evaluation and performance of the machine learning model for HAT barrier prediction. (a) Benchmark of the combinations of machine learning algorithms and molecular descriptors for HAT barrier prediction. NN refers to the neural network. XGB refers to eXtreme Gradient Boosting. RF refers to random forest. SVR refers to support vector regression. LR refers to linear regression. SGD refers to stochastic gradient descent. Local features refer to all the physical organic descriptors except the FMO-related features. (b) Feature selection of the AdaBoost model. The performance of the AdaBoost model can be improved through feature selection. The best subset of molecular descriptors includes 50 of the original 56 physical organic descriptors, which decreases the MSE from 0.728 to 0.719 (kcal^2 mol^{-2}. (c) The regression performance of the finalized AdaBoost model using five-fold cross-validation. (d) Ranking of the feature importance scores in the finalized AdaBoost model. Among the five categories of physical organic descriptors, BDE is significantly more important than the other features.

including 12 descriptors related to the atomic charge, 1 descriptor related to the bond dissociation energy (BDE), 12 descriptors related to the steric effect described by Nolan's percent buried volume model[43] and 9 descriptors based on the Wiberg bond index.[44] In addition, the global electronic properties of the reactants were described by 22 descriptors based on frontier molecular orbital (FMO) energy. These domain knowledge-based descriptors were developed by organic chemists and extensively applied in the rationalization and prediction of organic reactions. Such merits allowed the PhysOrg descriptors to have powerful encoding capabilities and can help build a predictive machine learning model with only dozens of input parameters and potentially improved generalization ability. The generation of the listed PhysOrg descriptors were based on DFT electronic structure calculations of the ground state reactant, which generally take minutes to hours depending on the size of the C–H substrate. It should also be emphasized that the machine learning prediction of the physical organic properties of small molecules has made significant progress over the recent few years. Powerful models for pK_a,[45–48] BDE,[49–51] FMO energy,[52,53] atomic charge[54,55] and dipole[56] prediction were developed, which provide a useful alternative source for the PhysOrg descriptors in model applications.

Evaluation of the candidate machine learning algorithms yielded the best model for the target HAT reactivity prediction. Based on five-fold cross-validation and random splitting, the regression performances of nine widely applied algorithms were tested (Fig. 2a). Tree-based algorithms, AdaBoost,[57] eXtreme Gradient Boosting (XGB),[58] ExtraTrees[59] and random forest (RF)[60] generally show better performances than the other algorithm frameworks. This is consistent with the number of machine learning studies in organic chemistry using PhysOrg descriptors,[21,22,24,30,32,35,38] which could be related to the nature of the relationship between PhysOrg descriptors and the target label (reactivity and selectivity) as well as the size of the dataset (generally hundreds to thousands). The AdaBoost model gave the best regression performance with an R^2 of 0.971, which is chosen for subsequent feature selection and model application. The inclusion of FMO-based global features improved the regression performances of all models (Fig. 2a). This suggests that the HAT reactivity is synergistically controlled by the local chemical properties of the reacting position and the global FMO properties of the C–H substrate and radical.

Key features of the established machine learning model

Based on the AdaBoost model, feature selection was applied to further lower the dimensions of PhysOrg descriptor space and improve the generalization ability of the machine learning (ML) model. 50 of the original set of 56 PhysOrg descriptors were identified as the best subset using recursive feature elimination with cross-validation,[61] which decreased the mean standard error (MSE) from 0.728 to 0.719 kcal^2 mol^{-2}. The regression performance of the finalized AdaBoost model using five-fold cross validation is shown in Fig. 2c. This model is able to predict the HAT barrier with an R^2 of 0.971 and an MAE of 0.60 kcal mol^{-1} as compared to the DFT-computed values, which can support the HAT reactivity prediction with the desired accuracy. Feature ranking of the 50 selected PhysOrg descriptors in the subset highlights the deciding role of BDE for the model performance (Fig. 2d). This discovery is consistent with the previous theoretical frameworks for HAT prediction: both the valence bond theory[18] approach and the Marcus cross relation approach[19] used the BDE-related properties as the key parameters. Although BDE is important for the ML prediction of the HAT barrier, we want to emphasize that the kinetics–thermodynamics correlation of HAT reaction is not a simple analytic function. For the 2926 DFT-computed HAT barriers and the corresponding reaction free energies,

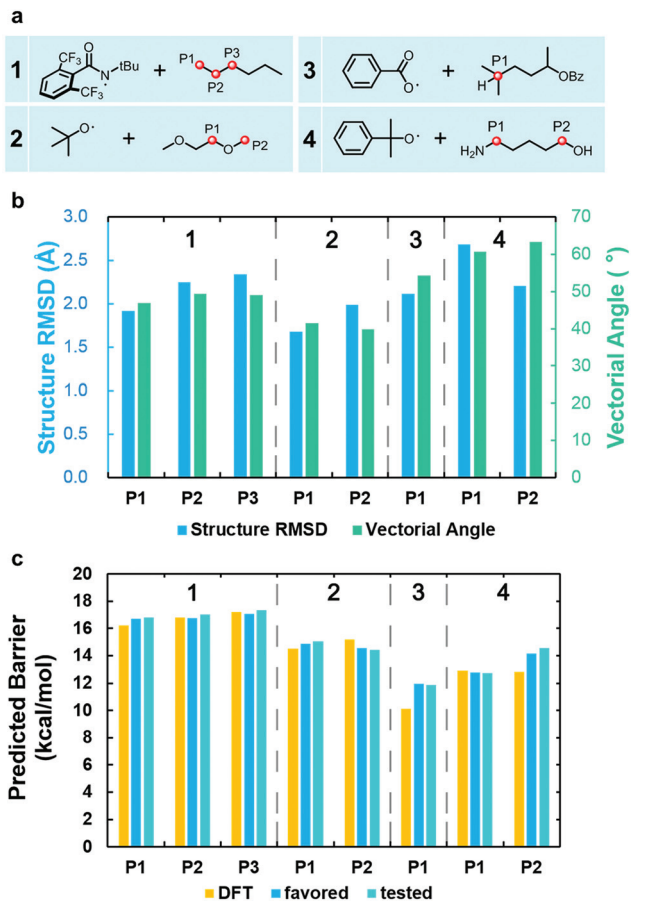


Fig. 3 Performance tests of ML prediction on structurally flexible C–H substrates. (a) The selected HAT reactions with structurally flexible C–H substrates. The studied reacting positions are labeled in red. (b) Comparison between the most favorable conformer and the tested conformer in terms of their structure and descriptor. The structural change is reflected using the RMSD of the coordinates of the most favorable conformer and the tested conformer. The descriptor change is reflected using the vectorial angle of the two sets of physical organic descriptors, generated using the most favorable conformer and the tested conformer. (c) Predicted barriers using the most favorable conformer and the tested conformer, as compared to the DFT-computed values. Despite the noticeable changes in the structure and descriptor, consistent barrier predictions were provided.

only a general trend can be identified for each category of the radical or C–H substrate (Fig. S5†). In addition to BDE, feature ranking also suggested that the FMO energy-derived descriptors are important for the HAT barrier prediction (Fig. 2d). Therefore, the PhysOrg descriptors allowed certain organic interpretation of the machine learning model that can be compared with the existing knowledge of HAT transformation. The BDE determines the overall trend of the HAT barrier, and the FMO interaction acts as an additional perturbation. The steric effect plays a limited role in the HAT kinetics due to the distant partial radical components in the HAT transition state.

The utilization of PhsyOrg descriptors allowed the trained machine learning model to capture the structure–activity relationship in a "organic chemist" fashion. This enabled the machine learning model to predict the consistent HAT barrier from different conformations of the C–H substrate, not necessarily the global minimum conformer. We tested the ML predictions with four structurally flexible C–H substrates from the reported HAT reactions (Fig. 3a). For each C–H substrate, the favorable conformer after an extensive conformational search and a high-energy conformer were located and used for the generation of PhysOrg descriptors. Noticeable differences in structure and PhysOrg descriptors exist in each case, as reflected in the structural RMSD and the vectorial angle of the generated PhysOrg descriptor sets (Fig. 3b). The details of the tested conformers are included in the ESI (Fig. S7†). Despite the change in both structure and PhysOrg descriptors, the trained ML model provided consistent and accurate barrier predictions for each substrate (Fig. 3c). The maximum differ-

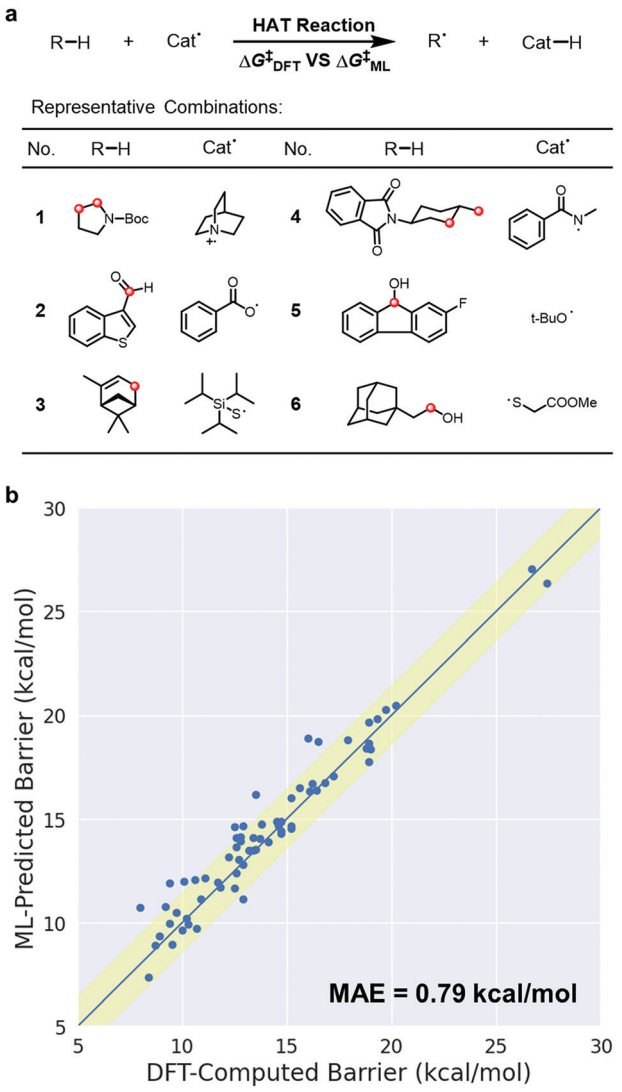


Fig. 4 Test set performance of ML prediction on HAT reactions that are not present in the designed chemical space. (a) Representative combinations of the selected 67 HAT reactions that are not present in the designed chemical space and not "seen" by the trained ML model. Studied C–H positions are labeled in red. (b) Comparison between the predicted HAT barrier and the DFT-computed values. The yellow region indicates the region of one order of magnitude deviation. 55 of the 67 tested cases fall in this region.

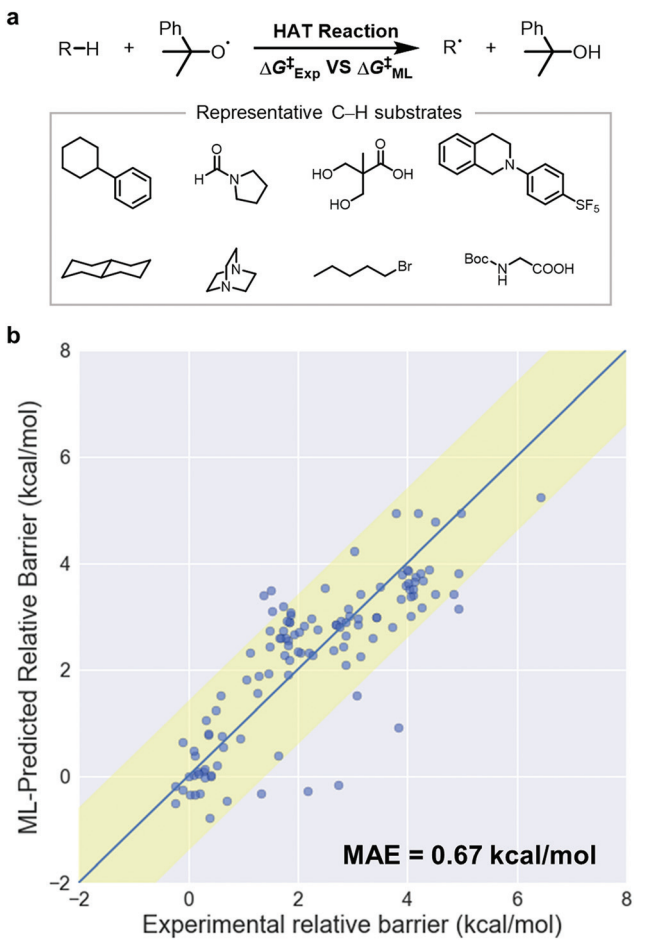


Fig. 5 Comparison between ML-predicted and experimental relative reactivities of HAT reactions with the cumyloxyl radical. (a) Representative substrates of the selected C–H substrates. (b) Comparison between the ML-predicted and experimental relative barriers using 1-(*t*-butyl)piperidine as the reference. The yellow region indicates the region of one order of magnitude deviation. 106 of the 117 tested cases fall in this region.

ence of the predicted HAT barrier for the eight positions of the four C–H substrates is only 0.4 kcal mol^{-1} (P2 of substrate 4, Fig. 3c). Considering the potentially large conformational space of organic molecules, this merit of the PhysOrg-based ML model is a useful advantage in practical applications.

Application of the machine learning model in test sets

The generalization ability of the ML model was further tested on a group of HAT reactions involved in the experimental advances[1–3,16] of C–H functionalization. This set of 67 HAT reactions include the molecular scaffolds and substituents of C–H substrates that were not present in the designed sample space. Six representative cases are shown in Fig. 4a, and additional details are presented in the ESI (Table S14 and Fig. S8†). The ML model provided satisfactory predictions on this test set of "unseen" HAT reactions (Fig. 4b). Comparing the ML-predicted and DFT-computed barriers, the MAE is 0.79 kcal mol^{-1}, and 55 of the 67 tested cases fall in the region of one order of magnitude deviation (highlighted in yellow, Fig. 4b). For substrates containing multiple C–H positions, our model has satisfactory ability to predict the correct position (Fig. S9†). This validation corroborated the predictive ability of the ML model on the HAT designs of target photoredox-mediated C–H functionalization.

In addition to the above test set of DFT-computed barriers, the ML predictions were also compared to the experimental kinetics data. Based on Bietti's systematic studies on the HAT kinetics,[17] HAT reactions involving 117 substrates and the cumyloxyl radical were selected, and the ML-predicted relative reactivities using our model were compared to Bietti's experimental data (Fig. 5). The rate constants and the corresponding relative barriers are provided in the ESI (Table S15 and Fig. S10†). The MAE is 0.67 kcal mol^{-1}, and 106 of the 117 tested substrates fall in the region of one order of magnitude deviation (highlighted in yellow, Fig. 5b). Our ML model is able to provide reliable predictions of the relative HAT reactivities as compared to the experimental values, despite the fact that the model training does not involve any experimental data. These results further validated the application of the ML model in the experimental designs as well as the strategy of mechanism-based computational statistics in the establishment of the predictive ML model for organic reactions.

In light of the success of ML prediction in the HAT reactivity, we are curious whether this model can predict the HAT selectivity for challenging targets such as drug candidates or natural products. Two molecules, (+)-camptothecin and menthol, were tested in this scenario (Fig. 6). (+)-Camptothecin contains four possible C–H positions for HAT reaction, but its selectivity prediction is not a challenging case considering the reactive ester-substituted benzylic position 1 (Fig. 6a). Empirical knowledge on photoredox-mediated C–H functionalization would lead to the prediction of HAT at position 1. Our ML model gave a satisfactory selectivity prediction favoring position 1 for the quinuclidine radical cation and *t*-butanethiyl radical, which is consistent with the DFT-computed results. However, for the sulphate radical anion, the ML model still predicts that position 1 is the major position for HAT reaction, while the DFT calculations suggested that positions 1 and 2 have comparable reactivities. These results indicated that the trained ML model by reactivity data tends to give similar selectivity prediction for the three radicals in the HAT reaction of (+)-camptothecin, which leads to difficulty in differentiating the subtle difference between the three radicals in the selectivity prediction task. For the case of menthol, this molecule contains four similar tertiary C–H positions, and the

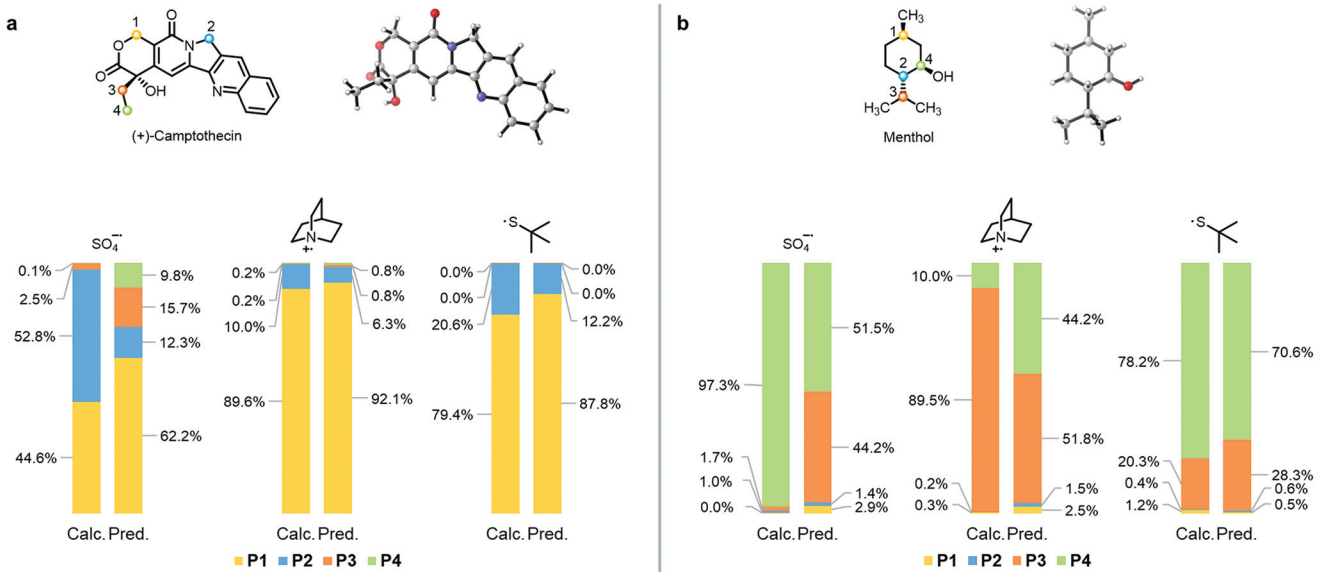


Fig. 6 DFT-computed and ML-predicted selectivities of HAT reaction of (a) (+)-camptothecin and (b) menthol with the sulphate radical anion, quinuclidine radical cation, and *t*-butanethiyl radicals. The color of the selectivity bar refers to the color-labeled C–H positions.

selectivity prediction is challenging based on the empirical experience (Fig. 6b). The ML selectivity prediction works well with the *t*-butanethiyl radical, and consistent results were found comparing the ML-predicted selectivity with the DFT-computed result. For the sulphate radical anion and quinuclidine radical cation, although the ML model is able to identify that positions 3 and 4 are the major positions for HAT reaction, the predicted degree of selectivity has a noticeable difference as compared to the DFT-computation result. Therefore, the ML model trained for reactivity prediction may not meet the accuracy requirement for selectivity prediction in challenging tasks. Modification of this reactivity-prediction model for selectivity-prediction through transfer learning is currently under investigation in our laboratory.

Conclusions

In summary, we developed a machine learning model that can accurately predict the barrier of HAT reaction in the photoredox-mediated C–H functionalization. Based on DFT-computed reactivity statistics of 2926 HAT transformations, the trained AdaBoost model is able to provide an efficient and reliable reactivity prediction with a mean absolute error of 0.60 kcal mol^{-1} using the physical organic descriptors of the ground state C–H substrate and radical. Feature ranking revealed the controlling factors of bond dissociation energy and frontier molecular orbital interactions for the barrier prediction. Due to the molecular encoding *via* physical organic descriptors, this model does not necessarily rely on the global minimum conformer for structurally flexible C–H substrates. The generalization ability of the ML model was further validated by both the DFT-computed barriers of "unseen" scaffolds and substituents as well as the experimental kinetics data of HAT reactions with the cumyloxyl radical. This work demonstrated the machine learning approach using physical organic descriptors and designed chemical space for the reactivity prediction of an organic transformation, which offers a readily available method for the HAT design in the photoredox-mediated C–H functionalization.

Author contributions

X. H. conceived the project and designed the computational study. L.-C. Y. performed the computations and the machine learning studies. X. L. and S.-Q. Z. are involved in the script writing for computational data generation and project discussion. All authors participated in writing the manuscript.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

Financial support from the National Natural Science Foundation of China (21702182 and 21873081), the Fundamental Research Funds for the Central Universities (2020XZZX002-02), the State Key Laboratory of Clean Energy Utilization (ZJUCEU2020007), and the Center of Chemistry for Frontier Technologies is gratefully acknowledged. Calculations were performed on the high-performance computing system at the Department of Chemistry, Zhejiang University. We thank Prof. Jie Wu and Qing-Yao Li of the Department of Chemistry, National University of Singapore for the helpful discussions on HAT reactivity.

Notes and references

1 L. Capaldo and D. Ravelli, Hydrogen Atom Transfer (HAT): A Versatile Strategy for Substrate Activation in Photocatalyzed Organic Synthesis, *Eur. J. Org. Chem.*, 2017, 2056–2071.
2 L. Capaldo, L. L. Quadri and D. Ravelli, Photocatalytic hydrogen atom transfer: the philosopher's stone for late-stage functionalization?, *Green Chem.*, 2020, **22**, 3376–3396.
3 S. Protti, M. Fagnoni and D. Ravelli, Photocatalytic C—H Activation by Hydrogen-Atom Transfer in Synthesis, *ChemCatChem*, 2015, **7**, 1516–1523.
4 J. Twilton, C. Le, P. Zhang, M. H. Shaw, R. W. Evans and D. W. C. MacMillan, The merger of transition metal and photocatalysis, *Nat. Rev. Chem.*, 2017, **1**, 0052.
5 K. Qvortrup, D. A. Rankic and D. W. C. MacMillan, A General Strategy for Organocatalytic Activation of C–H Bonds via Photoredox Catalysis: Direct Arylation of Benzylic Ethers, *J. Am. Chem. Soc.*, 2014, **136**, 626–629.
6 D. Hager and D. W. C. MacMillan, Activation of C–H Bonds via the Merger of Photoredox and Organocatalysis: A Coupling of Benzylic Ethers with Schiff Bases, *J. Am. Chem. Soc.*, 2014, **136**, 16986–16989.
7 J. D. Cuthbertson and D. W. C. MacMillan, The direct arylation of allylic sp^3 C–H bonds via organic and photoredox catalysis, *Nature*, 2015, **519**, 74–77.
8 S. Mukherjee, T. Patra and F. Glorius, Cooperative Catalysis: A Strategy to Synthesize Trifluoromethyl-thioesters from Aldehydes, *ACS Catal.*, 2018, **8**, 5842–5846.
9 S. Mukherjee, R. A. Garza-Sanchez, A. Tlahuext-Aca and F. Glorius, Alkynylation of $C_{sp}{}^2$(O)–H Bonds Enabled by Photoredox-MediatedHydrogen-Atom Transfer, *Angew. Chem., Int. Ed.*, 2017, **56**, 14723–14726.
10 X. Zhang and D. W. C. MacMillan, Direct Aldehyde C–H Arylation and Alkylation via the Combination of Nickel, Hydrogen Atom Transfer, and Photoredox Catalysis, *J. Am. Chem. Soc.*, 2017, **139**, 11353–11356.
11 S. Mukherjee, B. Maji, A. Tlahuext-Aca and F. Glorius, Visible-Light-Promoted Activation of Unactivated C(sp^3)–H Bonds and Their Selective Trifluoromethylthiolation, *J. Am. Chem. Soc.*, 2016, **138**, 16200–16203.

12 M. H. Shaw, V. W. Shurtleff, J. A. Terrett, J. D. Cuthbertson and D. W. C. MacMillan, Native functionality in triple catalytic cross-coupling: sp^3 C–H bonds as latent nucleophiles, *Science*, 2016, **352**, 1304–1308.
13 C. Le, Y. Liang, R. W. Evans, X. Li and D. W. C. MacMillan, Selective sp^3 C–H alkylation via polarity-match-based cross-coupling, *Nature*, 2017, **547**, 79–83.
14 A. M. Carestia, D. Ravelli and E. J. Alexanian, Reagent-dictated site selectivity in intermolecular aliphatic C–H functionalizations using nitrogen-centered radicals, *Chem. Sci.*, 2018, **9**, 5360–5365.
15 M. Milan, M. Salamone, M. Costas and M. Bietti, The Quest for Selectivity in Hydrogen Atom Transfer Based Aliphatic C–H Bond Oxygenation, *Acc. Chem. Res.*, 2018, **51**, 1984–1995.
16 M. Bietti, Activation and Deactivation Strategies Promoted by Medium Effects for Selective Aliphatic C–H Bond Functionalization, *Angew. Chem., Int. Ed.*, 2018, **57**, 16618–16637.
17 M. Salamone and M. Bietti, Tuning Reactivity and Selectivity in Hydrogen Atom Transfer from Aliphatic C–H Bonds to Alkoxyl Radicals: Role of Structural and Medium Effects, *Acc. Chem. Res.*, 2015, **48**, 2895–2903.
18 W. Lai, C. Li, H. Chen and S. Shaik, Hydrogen-Abstraction Reactivity Patterns from A to Y: The Valence Bond Way, *Angew. Chem., Int. Ed.*, 2012, **51**, 5556–5578.
19 J. M. Mayer, Understanding Hydrogen Atom Transfer: From Bond Strengths to Marcus Theory, *Acc. Chem. Res.*, 2011, **44**, 36–46.
20 B. P. Roberts, Polarity-reversal catalysis of hydrogen-atom abstraction reactions: concepts and applications in organic chemistry, *Chem. Soc. Rev.*, 1999, **28**, 25–35.
21 D. T. Ahneman, J. G. Estrada, S. Lin, S. D. Dreher and A. G. Doyle, Predicting reaction performance in C–N cross-coupling using machine learning, *Science*, 2018, **360**, 186.
22 M. K. Nielsen, D. T. Ahneman, O. Riera and A. G. Doyle, Deoxyfluorination with Sulfonyl Fluorides: Navigating Reaction Space with Machine Learning, *J. Am. Chem. Soc.*, 2018, **140**, 5004–5008.
23 A. F. Zahrt, J. J. Henle, B. T. Rose, Y. Wang, W. T. Darrow and S. E. Denmark, Prediction of higher-selectivity catalysts by computer-driven workflow and machine learning, *Science*, 2019, **363**, eaau5631.
24 W. Beker, E. P. Gajewska, T. Badowski and B. A. Grzybowski, Prediction of Major Regio-, Site-, and Diastereoisomers in Diels–Alder Reactions by Using Machine-Learning: The Importance of Physically Meaningful Descriptors, *Angew. Chem., Int. Ed.*, 2019, **58**, 4515–4519.
25 G. Skoraczyński, P. Dittwald, B. Miasojedow, S. Szymkuć, E. P. Gajewska, B. A. Grzybowski and A. Gambin, Predicting the outcomes of organic reactions via machine learning: are current descriptors sufficient?, *Sci. Rep.*, 2017, **7**, 3582.
26 Y. Guan, C. W. Coley, H. Wu, D. Ranasinghe, E. Heid, T. J. Struble, L. Pattanaik, W. H. Green and K. F. Jensen, Regio-selectivity prediction with a machine-learned reaction representation and on-the-fly quantum mechanical descriptors, *Chem. Sci.*, 2021, **12**, 2198–2208.
27 C. W. Coley, W. Jin, L. Rogers, T. F. Jamison, T. S. Jaakkola, W. H. Green, R. Barzilay and K. F. Jensen, A graph-convolutional neural network model for the prediction of chemical reactivity, *Chem. Sci.*, 2019, **10**, 370–377.
28 F. Sandfort, F. Strieth-Kalthoff, M. Kühnemund, C. Beecks and F. Glorius, A Structure-Based Platform for Predicting Chemical Reactivity, *Chem*, 2020, **6**, 1379–1390.
29 J. P. Reid and M. S. Sigman, Holistic prediction of enantioselectivity in asymmetric catalysis, *Nature*, 2019, **571**, 343–348.
30 S. Singh, M. Pareek, A. Changotra, S. Banerjee, B. Bhaskararao, P. Balamurugan and R. B. Sunoj, A unified machine-learning protocol for asymmetric catalysis as a proof of concept demonstration using asymmetric hydrogenation, *Proc. Natl. Acad. Sci. U. S. A.*, 2020, **117**, 1339.
31 S. Banerjee, A. Sreenithya and R. B. Sunoj, Machine learning for predicting product distributions in catalytic regioselective reactions, *Phys. Chem. Chem. Phys.*, 2018, **20**, 18311–18318.
32 A. Tomberg, M. J. Johansson and P.-O. Norrby, A Predictive Tool for Electrophilic Aromatic Substitutions Using Machine Learning, *J. Org. Chem.*, 2019, **84**, 4695–4703.
33 K. Jorner, T. Brinck, P.-O. Norrby and D. Buttar, Machine learning meets mechanistic modelling for accurate prediction of experimental activation energies, *Chem. Sci.*, 2021, **12**, 1163–1175.
34 N. Rollins, S. L. Pugh, S. M. Maley, B. O. Grant, R. S. Hamilton, M. S. Teynor, R. Carlsen, J. R. Jenkins and D. H. Ess, Machine Learning Analysis of Direct Dynamics Trajectory Outcomes for Thermal Deazetization of 2,3-Diazabicyclo[2.2.1]hept-2-ene, *J. Phys. Chem. A*, 2020, **124**, 4813–4826.
35 S. M. Maley, D.-H. Kwon, N. Rollins, J. C. Stanley, O. L. Sydora, S. M. Bischof and D. H. Ess, Quantum-mechanical transition-state model combined with machine learning provides catalyst design features for selective Cr olefin oligomerization, *Chem. Sci.*, 2020, **11**, 9665–9674.
36 L. Wang, C. Zhang, R. Bai, J. Li and H. Duan, Heck reaction prediction using a transformer model based on a transfer learning strategy, *Chem. Commun.*, 2020, **56**, 9368–9371.
37 Y. Chen, B. Tian, Z. Cheng, X. Li, M. Huang, Y. Sun, S. Liu, X. Cheng, S. Li and M. Ding, Electro-Descriptors for the Performance Prediction of Electro-Organic Synthesis, *Angew. Chem.*, 2021, **60**, 4199–4207.
38 X. Li, S.-Q. Zhang, L.-C. Xu and X. Hong, Predicting Regioselectivity in Radical C–H Functionalization of Heterocycles through Machine Learning, *Angew. Chem.*, 2020, **59**, 13253–13259.
39 J. L. Jeffrey, J. A. Terrett and D. W. C. MacMillan, O–H hydrogen bonding promotes H-atom transfer from α C–H bonds for C-alkylation of alcohols, *Science*, 2015, **349**, 1532.
40 J. Jin and D. W. C. MacMillan, Direct α-Arylation of Ethers through the Combination of Photoredox-Mediated C–H

Functionalization and the Minisci Reaction, *Angew. Chem., Int. Ed.*, 2015, **54**, 1565–1569.
41 C. Lee, W. Yang and R. G. Parr, Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1988, **37**, 785–789.
42 A. D. Becke, Density-functional thermochemistry. III. The role of exact exchange, *J. Chem. Phys.*, 1993, **98**, 5648–5652.
43 H. Clavier and S. P. Nolan, Percent buried volume for phosphine and N-heterocyclic carbene ligands: steric properties in organometallic chemistry, *Chem. Commun.*, 2010, **46**, 841–861.
44 K. B. Wiberg, Application of the pople-santry-segal CNDO method to the cyclopropylcarbinyl and cyclobutyl cation and to bicyclobutane, *Tetrahedron*, 1968, **24**, 1083–1096.
45 Q. Yang, Y. Li, J.-D. Yang, Y. Liu, L. Zhang, S. Luo and J.-P. Cheng, Holistic Prediction of the pKa in Diverse Solvents Based on a Machine-Learning Approach, *Angew. Chem.*, 2020, **59**, 19282–19291.
46 R. Roszak, W. Beker, K. Molga and B. A. Grzybowski, Rapid and Accurate Prediction of pKa Values of C–H Acids Using Graph Convolutional Neural Networks, *J. Am. Chem. Soc.*, 2019, **141**, 17142–17149.
47 K. Mansouri, N. F. Cariello, A. Korotcov, V. Tkachenko, C. M. Grulke, C. S. Sprankle, D. Allen, W. M. Casey, N. C. Kleinstreuer and A. J. Williams, Open-source QSAR models for pKa prediction using multiple machine learning approaches, *J. Cheminf.*, 2019, **11**, 60.
48 Y. Lu, S. Anand, W. Shirley, P. Gedeck, B. P. Kelley, S. Skolnik, S. Rodde, M. Nguyen, M. Lindvall and W. Jia, Prediction of pKa Using Machine Learning Methods with Rooted Topological Torsion Fingerprints: Application to Aliphatic Amines, *J. Chem. Inf. Model.*, 2019, **59**, 4706–4719.
49 P. C. St. John, Y. Guan, Y. Kim, S. Kim and R. S. Paton, Prediction of organic homolytic bond dissociation enthalpies at near chemical accuracy with sub-second computational cost, *Nat. Commun.*, 2020, **11**, 2328.
50 C. Feng, E. Sharman, S. Ye, Y. Luo and J. Jiang, A neural network protocol for predicting molecular bond energy, *Sci. China: Chem.*, 2019, **62**, 1698–1703.
51 H. Yu, Y. Wang, X. Wang, J. Zhang, S. Ye, Y. Huang, Y. Luo, E. Sharman, S. Chen and J. Jiang, Using Machine Learning to Predict the Dissociation Energy of Organic Carbonyls, *J. Phys. Chem. A*, 2020, **124**, 3844–3850.
52 F. A. Faber, L. Hutchison, B. Huang, J. Gilmer, S. S. Schoenholz, G. E. Dahl, O. Vinyals, S. Kearnes, P. F. Riley and O. A. von Lilienfeld, Prediction Errors of Molecular Machine Learning Models Lower than Hybrid DFT Error, *J. Chem. Theory Comput.*, 2017, **13**, 5255–5264.
53 G. A. Pinheiro, J. Mucelini, M. D. Soares, R. C. Prati, J. L. F. Da Silva and M. G. Quiles, Machine Learning Prediction of Nine Molecular Properties Based on the SMILES Representation of the QM9 Quantum-Chemistry Dataset, *J. Phys. Chem. A*, 2020, **124**, 9854–9866.
54 P. Bleiziffer, K. Schaller and S. Riniker, Machine Learning of Partial Charges Derived from High-Quality Quantum-Mechanical Calculations, *J. Chem. Inf. Model.*, 2018, **58**, 579–590.
55 B. Nebgen, N. Lubbers, J. S. Smith, A. E. Sifain, A. Lokhov, O. Isayev, A. E. Roitberg, K. Barros and S. Tretiak, Transferable Dynamic Molecular Charge Assignment Using Deep Neural Networks, *J. Chem. Theory Comput.*, 2018, **14**, 4687–4698.
56 A. E. Sifain, N. Lubbers, B. T. Nebgen, J. S. Smith, A. Y. Lokhov, O. Isayev, A. E. Roitberg, K. Barros and S. Tretiak, Discovering a Transferable Charge Assignment Model Using Machine Learning, *J. Phys. Chem. Lett.*, 2018, **9**, 4495–4501.
57 Y. Freund and R. E. Schapire, A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting, *J. Comput. Syst. Sci.*, 1997, **55**, 119–139.
58 T. Chen and C. Guestrin, 2016, arXiv, preprint, arXiv:1603.02754, https://arxiv.org/abs/1603.02754.
59 P. Geurts, D. Ernst and L. Wehenkel, Extremely randomized trees, *Mach. Learn.*, 2006, **63**, 3–42.
60 L. Breiman, Random Forests, *Mach. Learn.*, 2001, **45**, 5–32.
61 I. Guyon, J. Weston, S. Barnhill and V. Vapnik, Gene Selection for Cancer Classification using Support Vector Machines, *Mach. Learn.*, 2002, **46**, 389–422.