Article

Rapid and Accurate Estimation of Activation Free Energy in Hydrogen Atom Transfer-Based C−H Activation Reactions: From Empirical Model to Artificial Neural Networks

Siqi Ma, Shipeng Wang, Jiawei Cao, and Fengjiao Liu*

Read Online

ACCESS | Metrics & More | Article Recommendations | SI Supporting Information

ABSTRACT: A well-performing machine learning (ML) model is obtained by using proper descriptors and artificial neural network (ANN) algorithms, which can quickly and accurately predict activation free energy in hydrogen atom transfer (HAT)-based sp^3 C−H activation. Density functional theory calculations (UωB97X-D) are used to establish the reaction system data sets of methoxyl ($CH_3O\cdot$), trifluoroethoxyl ($CF_3CH_2O\cdot$), *tert*-butoxyl (tBuO·), and cumyloxyl (CumO·) radicals. The simplified Roberts' equation proposed in our recent study works here [R^2 = 0.84, mean absolute error (MAE) = 0.85 kcal/mol]. Its performance is comparable with univariate Mulliken-type electronegativity (χ) with the ANN model. The ANN model with bond dissociation free energy, χ, α-unsaturation, and Nolan buried volume (%V_{buried}) successively improves R^2 and MAE to 0.93 and 0.54 kcal/mol, respectively. It reproduces the test sets of trichloroethoxyl ($CCl_3CH_2O\cdot$) with R^2 = 0.87 and MAE = 0.89 kcal/mol and accurately predicts the relative experimental barrier of the HAT reactions with CumO· and the site selectivity of $CH_3O\cdot$.
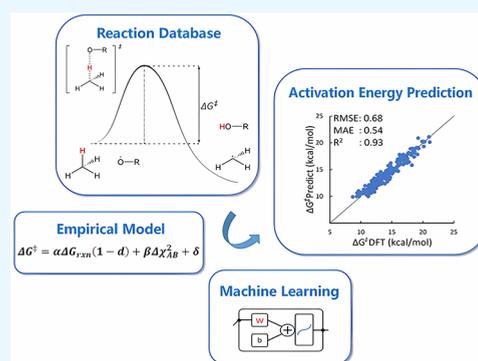
INTRODUCTION

With the increasing demand of sustainable development, use of the sp^3 C−H bond function in chemical synthesis has received great attention because it can provide a practical solution to upgrade abundant hydrocarbon raw materials into valuable products.[1−11] The hydrogen atom transfer (HAT)-based method provides an effective strategy for hydrocarbon activation and has been widely used.[2,12−20] However, the prediction of reactivity is still a challenge. The reactivity is closely related to the C−H bond strength and the choice of H-extracting radicals.[21−32] Understanding the reaction mechanisms and predicting the reactivity should assist a more reasonable and effective design of C−H activation reactions.[33−38]

The activation free energy is a key parameter for the mechanism, reaction rate, and selectivity of chemical reactions. The prediction of the activation free energy helps to achieve a comprehensive understanding of chemical reactions and to construct chemical reaction diagrams, thus accelerating catalyst design. However, the activation free energy estimation is time-consuming, both experimentally and computationally, leading to a bottleneck effect on the rapid catalyst design and characterization.

In experimental work, the reaction rate constant is determined by running multiple experiments at different temperatures, and then the activation free energy can be obtained by application of the Eyring equation. In computational science, quantum chemistry methods can evaluate the

activation free energy by identifying the transition states along a given reaction path. It is inevitable to spend a lot of computational cost to deal with complex chemical processes involving many reactions.[39,40] In this situation, it is very valuable to achieve rapid and accurate prediction of activation free energies in C−H activation reactions.

As an empirical model, the Bell−Evans−Polanyi (BEP) correlation is widely used to quickly estimate the activation energy.[41] It correlates the activation energy with reaction energy through $\Delta E^\ddagger = \gamma \Delta E_{r \times n} + \xi$. Tedder found the existence of a BEP correlation between the rate constant and the C−H bond strength for H-abstraction reactions from alkanes by various radicals ($CH_3\cdot$, $CF_3\cdot$, Br·, etc.).[42] Mayer et al. showed the generality of BEP correlations in the HAT of CrO_2Cl_2 and MnO_4^-.[43] Shaik et al. used density functional theory (DFT) to study the BEP relationship of enzyme cytochrome P450 oxidizing a series of C−H bonds and established a prediction model.[44] However, as the organic system becomes more and more complex, people cannot observe the simple linear relationship between the activation energy and the bond energy.[45−48] In addition to bond energy

Figure 1. C−H bonds studied in this work.

or reaction energy, Roberts pointed out that polar effects, steric effects, and unsaturation should be considered when discussing various HAT reactions in the liquid and gas phases. In addition, his study has emphasized and quantified the polar effects.[49] Tedder proposed that the polar effects in the HAT of $CH_3\cdot$ and $CF_3\cdot$ will cause a deviation from the BEP correlation.[42] Our previous study used DFT calculations to explore the selectivity of dimethyldioxirane (DMDO) in the C−H oxidation of various compounds, and it revealed a different BEP correlation between the C−H bonds in saturated and unsaturated compounds.[46] This bimodal BEP correlation was also found in HAT from sp^3 C−H bonds to the cumyloxyl radical by Bietti et al. All these precedents urged us to think about the influence of nonlinear characteristics and factors other than thermodynamics in HAT.

Multiple linear regression (MLR) is one strategy to consider the effect of several variables on activation energy. Roberts and Steel proposed an improved form of the Evans−Polanyi equation (Roberts' equation) involving reaction thermodynamics, radical electronegativity differences, and conjugate delocalization of the unpaired electron and structural factors, which work well for a variety of HAT reactions with different H-extracting radicals.[49] Our recent work simplified Roberts' equation with reaction energy, radical electronegativity differences, and unsaturation and led to good results for the HAT reactions of 26 sp^3 C−H bonds by alkoxyl radicals, with R^2 = 0.89 and mean absolute error (MAE) and root mean square error (RMSE) values of 0.9 and 1.1 kcal/mol, respectively (compared with the DFT calculations).[50]

On the other hand, the significant advancement of machine learning (ML) technology provides a new strategy for solving various chemical problems. The artificial neural network (ANN), as one of the most popular ML methods, is a data-driven adaptive method. Due to its parallel structure and the ability to simulate arbitrary functions, it can be regarded as a kind of multiple nonlinear regression that can handle complex nonlinear problems.[51] By establishing the potential correlations from the data, the ANN can solve problems that involve unknown relationships. Nowadays, the ANN is actively used in various chemical fields, such as quantum chemistry, virtual screening of molecular materials, and synthetic pathways of chemical substances.[52,53] While our study was ongoing, Hong

et al. have provided an ML approach for reactivity prediction in photoredox-mediated HAT catalysis.[54]

In the present study, we aim to (1) apply the ANN model to the HAT-based sp^3 C−H activation to achieve the goal of rapid and accurate prediction of the activation free energy, (2) assess the performance of the simplified Roberts' equation proposed in our recent study,[50] and (3) compare the performance between MLR and ANN with the same physical variables. The substrate C−H bonds have been expanded (Figure 1). We added a series of hydrogen donor substrates to build the model on the basis of our previous work[46] [including allylic, benzylic, formylic, and α-heteroatom (O, N, S) group]. The activated C−H bond is highlighted in red. In addition to methoxyl ($CH_3O\cdot$), trifluoroethoxyl ($CF_3CH_2O\cdot$), and tert-butoxy ($tBuO\cdot$), cumyloxyl ($CumO\cdot$) was also used as the H-extracting radical. Alkoxyl radicals are easily available and can be highly efficient for the abstraction of H atoms. They are powerful tools for the catalytic activation of hydrocarbons under mild reaction conditions. Zuo et al. achieved the C−H functionalization of short-chain alkanes (methane, ethane, etc.) utilizing cerium salts and alcohols under mild reaction conditions with $CH_3O\cdot$, $CCl_3CH_2O\cdot$, and $CF_3CH_2O\cdot$ being used as the H-extracting radicals.[55] Bietti carried out systematic kinetic studies of HAT reactions by $CumO\cdot$.[48] These experimental data are helpful to verify the theoretical model.

Figure 2. Scheme of alkoxyl radical-mediated HAT activation with methane.

■ METHOD

Database. We established data sets for the alkane sp^3 C−H activation by various alkoxyl radicals: $CH_3O\cdot$, $CF_3CH_2O\cdot$, $tBuO\cdot$, and $CumO\cdot$ (Tables S1−S4). The data sets were used to build the model. The C−Hs listed in Figure 1 were studied. The substrates and H-extracting radicals were selected to

Figure 3. HAT reaction space representation using vectors.

Table 1. Main Parameters of ANN Models

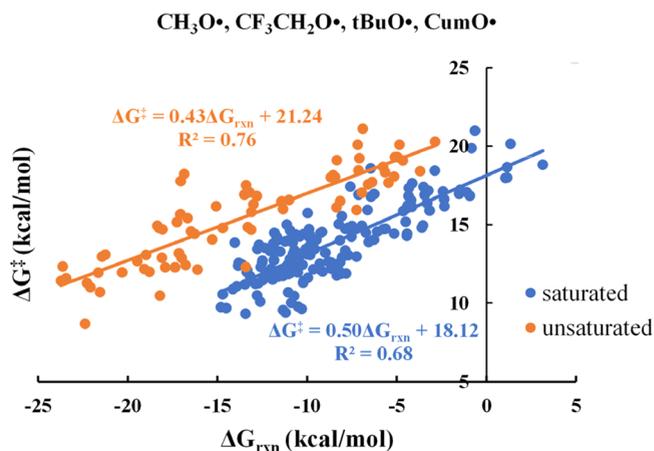| ANN	
hidden layer sizes	6
hidden unit	3
activation	tan*h*
tolerance	0.001
training function	Levenberg−Marquardt

Figure 5. Activation free energy for HAT as a function of ΔG_{rxn}.

consider polar effects, steric effects, and α-unsaturation effects. Figure 2 shows an example of the HAT step by the alkoxyl radical.[55] All calculations were carried out with Gaussian 16.[56] The geometries of minima and transition states were optimized using UωB97X-D[57] with the 6-31G (d)[58] basis set in the gas phase. Vibrational frequency analyses confirmed the nature of the structures as either local energy minima or first-order saddle points (transition states). Single point energies with a more extensive basis set were obtained with UωB97X-D/6-311++G (d,p)[59] on the optimized geometries. The solvent effect of CH_3CN on the reaction was estimated by using the SMD model.[60]

Figure 3 shows an example of the data sets. The "sub" and "rad" qualifiers represent the descriptors of the substrate and

H-extracting radical, respectively. The bond dissociation free energies (BDFEs) of the substrate C−H and alkoxyl radical O−H were both taken into consideration. The polar effect was
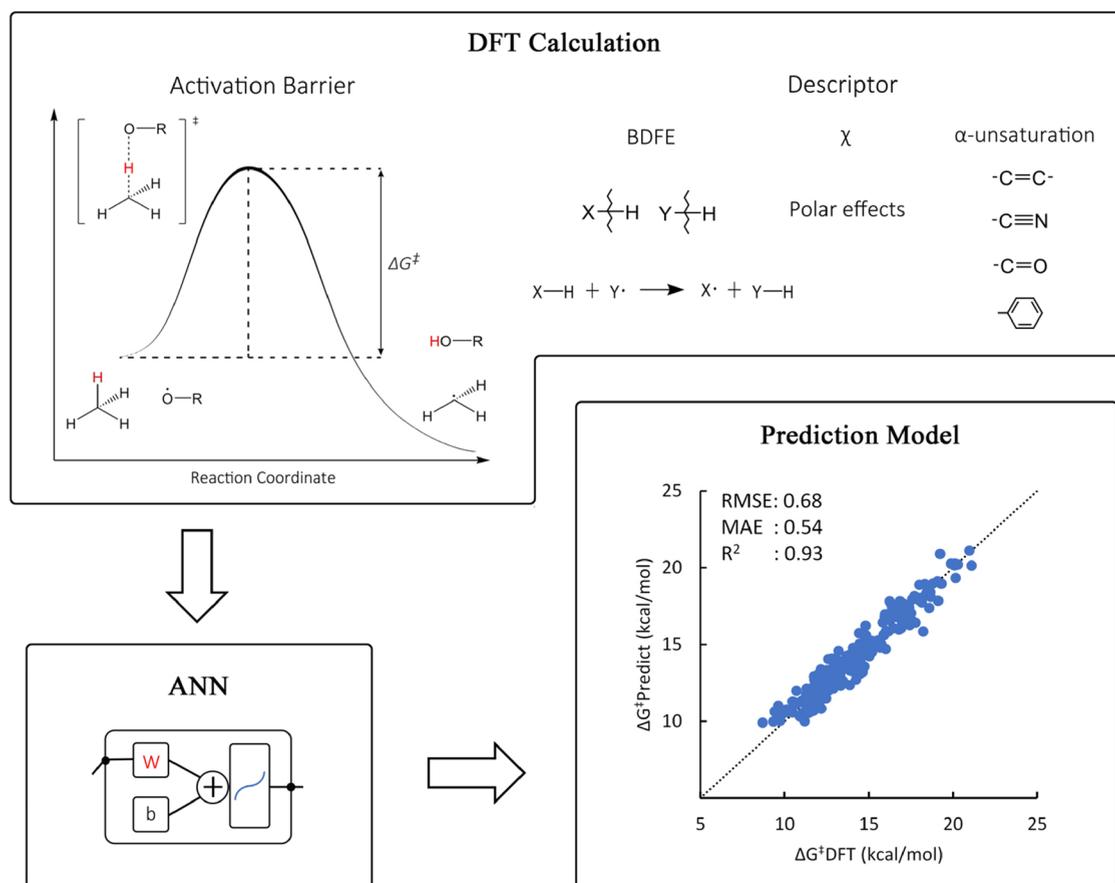
Figure 4. Process of the ANN model predicting activation free energy.
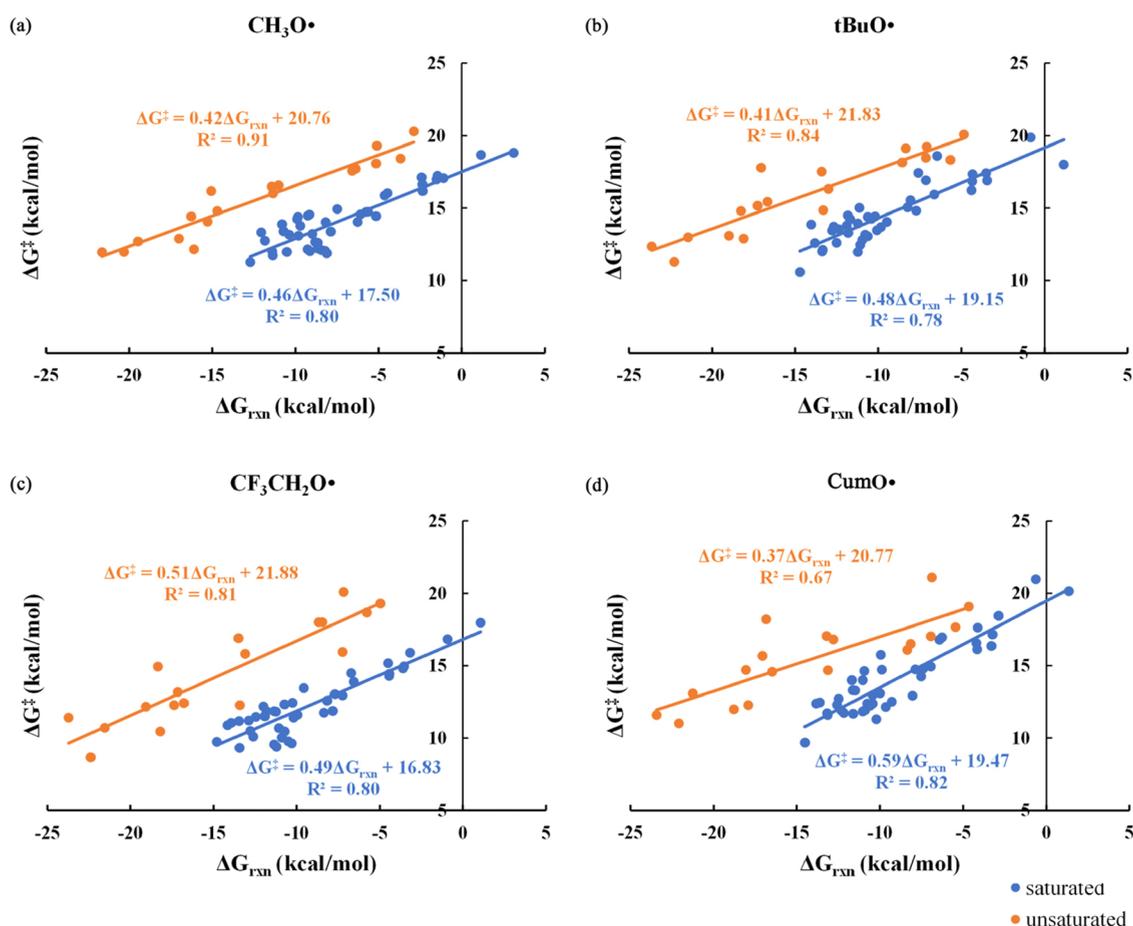
Figure 6. Activation free energy for the HAT reaction as a function of ΔG_{rxn} for the different alkoxyl radicals: (a) $CH_3O\cdot$, (b) $tBuO\cdot$, (c) $CF_3CH_2O\cdot$, and (d) $CumO\cdot$.

quantified by the Mulliken-type electronegativity. The Mulliken-type electronegativity of the X· radical (χ_x) is defined in eq 1, where IE and EA are the X· vertical ionization energy and vertical electron affinity, respectively.

$$\chi_x = (IE_x + EA_x)/2 \qquad (1)$$

$\Delta\chi_{AB}$ $(\Delta\chi_{AB} = \chi_A - \chi_B, A-H + B\cdot \rightarrow A\cdot + BH)$ reflects the electronegativity difference between the substrate and H-extracting radical. The unsaturation effect refers to whether an unsaturated group adjacent to the investigated C−H is present. The "0" and "1" qualifiers are used for "saturated" C−H and "unsaturated" C−H bonds, respectively. The Nolan buried volume $(\%V_{buried})^{61}$ is used to consider the steric effect (see Section S8 for the detailed calculation).

ANN Model. Figure 4 is an overview of the ANN model. Activation barriers and descriptors were all obtained by DFT methods. The ANN model was trained on the basis of appropriate descriptors.

The back-propagation ANN model[62,63] was applied to predict the activation free energy. Through K-fold cross-validation $(k = 3)$,[64] the topology of the ANN model was finally determined with three hidden neurons and three hidden layers (see Table 1). The activation function was a $\tanh$ function $(\tan h(x) = \frac{2}{1 + e^{-2x}} - 1)$ that allows the network to map any nonlinear process. The second-order optimization method was an implementation of the Levenberg−Marquardt (LM) algorithm.[65−67]

RESULTS AND DISCUSSION

Tables S1−S4 list the activation free energy $(\Delta G^{\ddagger})$, the reaction free energy (ΔG_{rxn}), the BDFE, the Mulliken electronegativity (χ), and the Nolan buried volume $(\%V_{buried})$ for HAT reactions promoted by $CH_3O\cdot$, $CF_3CH_2O\cdot$, $tBuO\cdot$, and $CumO\cdot$, respectively. The substrate C−H BDFE ranges from 71.2 to 96.0 kcal/mol. The $\Delta G^{\ddagger}$ values range from 11.3 to 20.3 kcal/mol for $CH_3O\cdot$, range from 10.6 to 20.1 kcal/mol for $tBuO\cdot$, range from 9.7 to 21.1 kcal/mol for $CumO\cdot$, and range from 8.7 to 20.1 kcal/mol for $CF_3CH_2O\cdot$. Overall, $CF_3CH_2O\cdot$ has a greater reactivity than $CH_3O\cdot$, $tBuO\cdot$, and $CumO\cdot$.

Empirical Model. In HAT, the BEP correlation can be expressed as shown in eq 2, where $\Delta G^{\ddagger}$ is the activation free energy, ΔG_{rxn} is the reaction free energy, $BDFE_{sub}$ and $BDFE_{rad}$ are the BDFE of the substrate C−H and alkoxyl radical O−H, respectively, and γ and ξ are obtained from linear regression analysis.

$$\Delta G^{\ddagger} = \gamma\Delta G_{rxn} + \xi = \gamma\left(BDFE_{sub} - BDFE_{rad}\right) + \xi \qquad (2)$$

Figure 5 shows the linear relationship of $\Delta G^{\ddagger}$ vs ΔG_{rxn} for the reaction data sets of 60 sp^3 C−Hs (Figure 1) with four alkoxyl radicals ($CH_3O\cdot$, $CF_3CH_2O\cdot$, $tBuO\cdot$, and $CumO\cdot$). The traditional BEP correlation obviously does not work here. Although the correlation is rough, the "saturated" and "unsaturated" C−Hs tend to be divided into two categories (Figure 5) as found in our previous study.[46] The effects of the

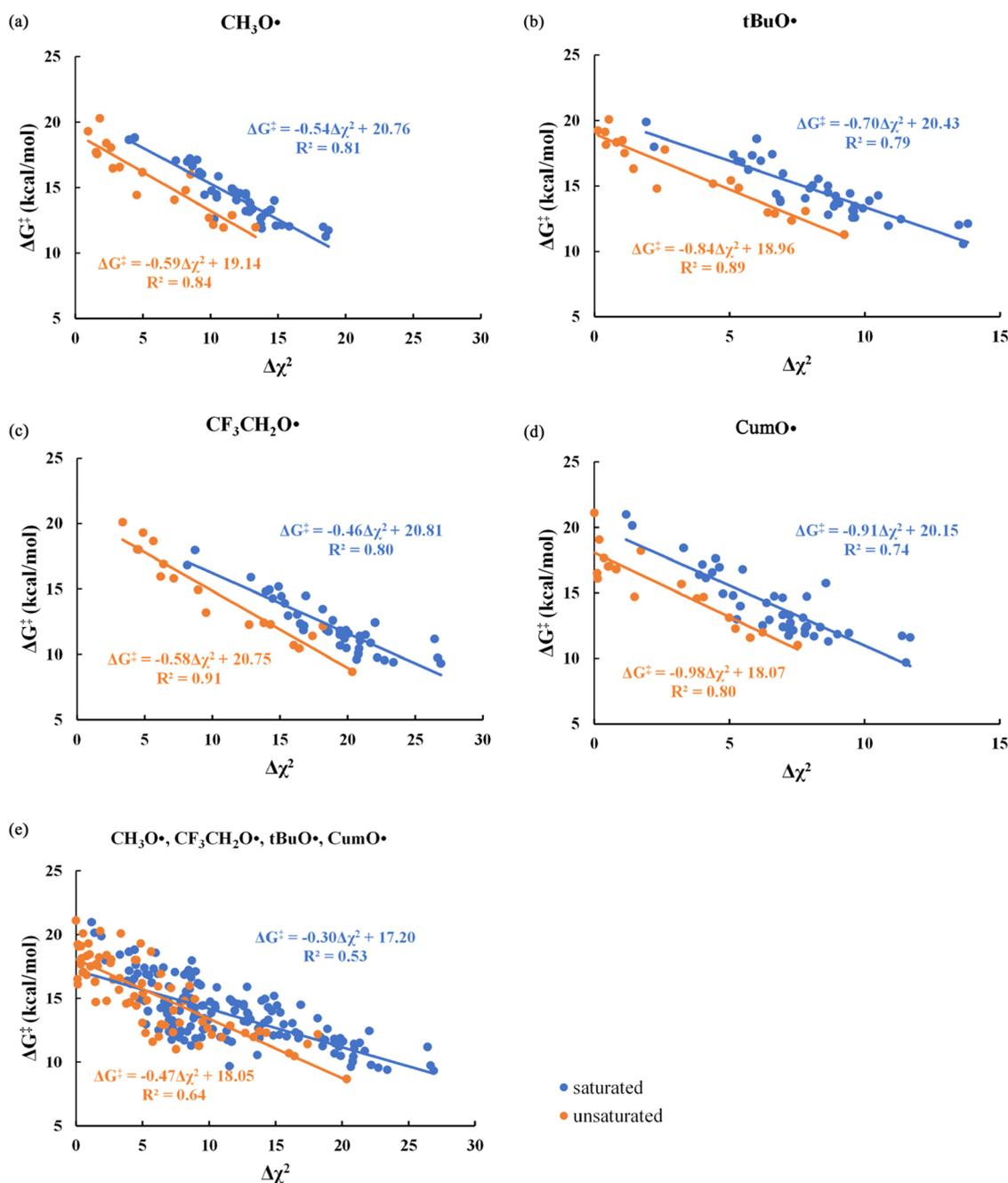**Figure 7.** Activation free energy for the HAT reaction as a function of $\Delta\chi^2$ for different alkoxyl radicals: (a) $CH_3O\cdot$, (b) $tBuO\cdot$, (c) $CF_3CH_2O\cdot$, (d) $CumO\cdot$, and (e) all four alkoxyl radicals ($CH_3O\cdot$, $CF_3CH_2O\cdot$, $tBuO\cdot$, and $CumO\cdot$).

unsaturation and of the radical nature are responsible for the inadequacy of the BEP correlation.

Figure 6a−d shows the scatter diagram of $\Delta G^{\ddagger}$ vs ΔG_{rxn} for each alkoxyl radical. In comparison with $CH_3O\cdot$, the higher electronegativity $CF_3CH_2O\cdot$ radical is more reactive and yields a worse linear correlation. The steric effect also influences the BEP correlation, as is evident from a comparison of the differences between the $\Delta G^{\ddagger}$ vs ΔG_{rxn} scatter plots for $CH_3O\cdot$, $tBuO\cdot$, and $CumO\cdot$. Due to the influence of different H-extracting radicals, the BEP correlation for the whole series is poor, as shown in Figure 5.

Figure 7 shows the scatter diagram of $\Delta G^{\ddagger}$ vs $\Delta\chi^2$ ($\Delta\chi = \chi_{\_sub} - \chi_{\_rad}$). This correlation is better than that of $\Delta G^{\ddagger}$ vs ΔG_{rxn}, and the unsaturation effect is less marked, to the point

that the "saturated" and "unsaturated" C−Hs can be classified into one category.

The results of Figure 7 imply that $\Delta\chi^2$ plays a more important role than ΔG_{rxn} in predicting the activation free energies with univariate linear regression. The importance of $\Delta\chi^2$ is further proved by the random forest (RF) algorithm.[68] RF can provide a measure of the feature importance based on the mean decrease in impurity (MDI), and the impurity is calculated by the split criterion of the decision trees (entropy).[68] Figure 8 shows the feature importance of the unsaturation effect, $\Delta\chi^2$, and ΔG_{rxn} analyzed by the RF algorithm. As shown in Figure 8, $\Delta\chi^2$ plays the most important role, followed by ΔG_{rxn} and the least important is the effect of unsaturation. It is worth mentioning, however, that the feature

Feature Importances

Figure 8. Feature importance of each descriptor analyzed by RF.

importance analysis based on MDI is biased to high cardinality features (typically numerical features, e.g., $\Delta\chi^2$) and probably underestimates low cardinality features (binary features, e.g., the unsaturation effect).[69] Therefore, the impact of descriptors needs to be further analyzed. Though $\Delta\chi^2$ shows a good correlation with $\Delta G^{\ddagger}$, this is insufficient to accurately predict the activation free energy, especially using a simple linear relationship.

Then, the simplified Roberts' relationship developed in our recent work[50] was applied (eq 3).

$$\Delta G^{\ddagger} = \alpha \Delta G_{r\times n}(1 - d) + \beta \Delta\chi^2 + \delta \tag{3}$$

This expression contains the reaction energy ($\Delta G_{r\times n}$, $\Delta G_{r\times n}$ = $BDFE_{sub} - BDFE_{rad}$), the unsaturation term (d), and the Mulliken-type electronegativity difference ($\Delta\chi$, $\Delta\chi = \chi_{sub} - \chi_{rad}$) between the substrate and H-extracting radical. From an MLR analysis of our data sets, the coefficients for this simplified Roberts' relationship are shown in eq 4.

$$\Delta G^{\ddagger}_{predict} = 0.42\Delta G_{r\times n}(1 - d) - 0.20\Delta\chi^2 + 19.71 \tag{4}$$

The correlation between $\Delta G^{\ddagger}_{DFT}$ and the $\Delta G^{\ddagger}_{Predict}$ obtained by this multivariate linear relationship, shown in Figure 9, is good ($R^2 = 0.84$, MAE = 0.85, and RMSE = 1.05).

$$\Delta G^{\ddagger} = 0.41\Delta G^{\ddagger}_{r\times n}(1-d) - 0.20\Delta\chi^2 + 19.70$$

Figure 9. DFT-computed vs predicted barriers using eq 4 for the reaction data sets of 60 sp³ C−Hs (Figure 1) with four alkoxyl radicals (CH₃O·, CF₃CH2O·, tBuO·, and CumO·).

ANN Model. The ANN model was trained on the reaction data sets of 60 sp³ C−Hs (Figure 1) with four alkoxyl radicals (CH₃O·, CF₃CH₂O·, tBuO·, and CumO·) by using different descriptors. Figure 10a,b shows the activation free energies obtained by the ANN model (with only the BDFE and χ as descriptors, respectively) versus the DFT method. The BDFE can hardly be used to predict the activation free energy, as shown by $R^2 = 0.62$, MAE = 1.28, and RMSE = 1.62. The χ still works better than BDFE when used as the unique descriptor in the ANN model. This is reflected by $R^2 = 0.85$, MAE = 0.79, and RMSE = 1.03. It is worth noting that the performance of the single χ descriptor with the ANN model is comparable to the MLR of eq 4. Subsequently, we tried to add the descriptors. The combination of BDFE_sub, BDFE_rad, χ_sub, χ_rad, and α-unsaturation successively makes the R^2 improve to 0.92, accompanied by MAE = 0.60 and RMSE = 0.75, as shown in Figure 10c. After adding the $\%V_{buried}$, the ANN model performs best ($R^2 = 0.93$, MAE = 0.54, and RMSE = 0.68), see Figure 10d.

In order to further evaluate the predictive ability of the model, we used the reaction data of CCl₃CH₂O· as the test set (Table S5). The R^2 and MAE values obtained from the ANN model without $\%V_{buried}$ are 0.80 and 1.10 kcal/mol, showing that the ANN model trained on the data set composed of CH₃O·, CF₃CH₂O·, tBuO·, and CumO· can map the influence of the polarity changes of CCl₃CH₂O· (Figure 11a). By adding the $\%V_{buried}$ term, the R^2 and MAE values improve to 0.87 and 0.79 kcal/mol, respectively, showing the importance of the steric effect (Figure 11b).

Furthermore, we predicted the experimental results by using the trained model. According to Bietti's research on the HAT reaction of CumO·, we collected another 45 sp³ C−Hs from his study and used the ANN model trained on the DFT computational data to make predictions. To eliminate the final prediction bias caused by the errors of the DFT calculation and the experiment, the relative activation free energy was used here. Rate constants and relative activation free energies are provided in the Supporting Information (Table S6). Figure 12 compares the activation free energies predicted by the ANN model with the relative activation free energies of Bietti's experiments. The prediction results show that the ANN model containing $\%V_{buried}$ can better map the influence of the CumO· steric effects and make more accurate predictions ($R^2 = 0.70$, MAE = 0.65, and RMSE = 0.80).

We also tried to predict the results of the CH₃O· selectivity from Zuo's reports (Table S7).[70] The BDFE, χ, α-unsaturation, and $\%V_{buried}$ descriptors were used. Figure 13 includes the relative ratio of site selectivity and the activation free energies obtained by the ANN prediction and by the DFT calculation. The numbers marked in red indicate that the prediction error is over 2 kcal/mol, and the blue backgrounds indicate the main activation site predicted by the ANN model. Among the six substrates not included in the training set, 2,3-dimethylbutane has been widely used as a standard substrate for evaluating selectivity in C−H bond functionalization. The α-tertiary carbon can be formed with good selectivity (ratio 97:3) when CH₃O· was used as a HAT reagent. Comparing with the DFT-calculated and ANN-predicted activation free energies, the reaction site is the same. The tertiary C−H bond of adamantane can be functionalized and predicted as well. N-Hexane has three different types of C−H bonds in terms of steric hindrance and bond strength. In this case, CH₃O· has high selectivity for the weaker methylene C−H bonds, and the
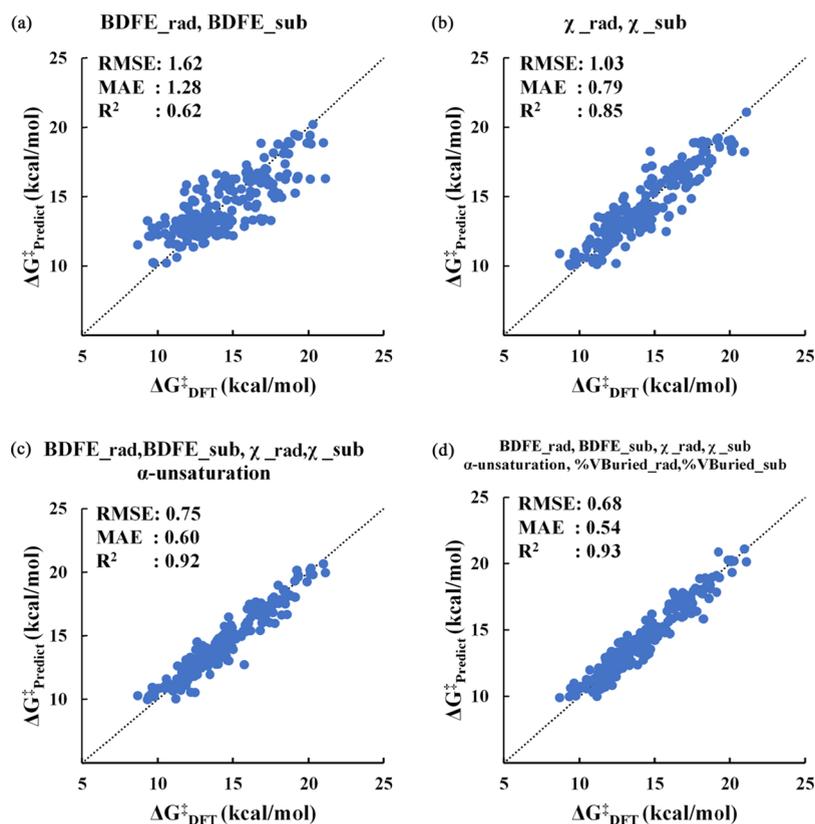
Figure 10. DFT-computed barriers vs ANN-predicted barriers for the reaction data sets of 60 sp^3 C−Hs (Figure 1) with four alkoxyl radicals (CH$_3$O·, CF$_3$CH$_2$O·, tBuO·, and CumO·) by different descriptors (a) BDFE_sub and BDFE_rad, (b) χ_sub and χ_rad, and (c) BDFE_sub, BDFE_rad, χ_sub, χ_rad, and α-unsaturation. (d) BDFE_sub, BDFE_rad, χ_sub, χ_rad, α-unsaturation, %Vburied_sub, and %Vburied_rad. The "sub" and "rad" qualifiers represent the descriptors of the substrate and H-extracting radical, respectively.
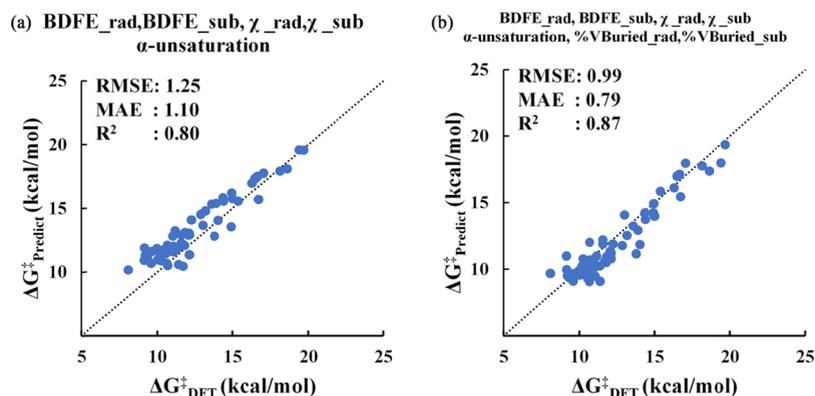
Figure 11. DFT-computed vs ANN-predicted barriers for the test set of 60 sp^3 C−Hs (Figure 1) with CCl$_3$CH$_2$O· by different descriptors: (a) BDFE_sub, BDFE_rad, χ_sub, χ_rad, and α-unsaturation. (b) BDFE_sub, BDFE_rad, χ_sub, χ_rad, α-unsaturation, %Vburied_sub, and % Vburied_rad.

predicted activation free energy at this site is also the lowest. In 2,4-dimethylpentane, CH$_3$O· is used to obtain methine functionalization. Although the activation free energy predicted by the ANN model is lower, the selectivity is consistent with the experiment.

■ CONCLUSIONS

We have strived to achieve a rapid and accurate estimation of activation free energies in HAT-based C−H activation reactions with both an empirical method and the ANN model. First, we established a data set of 300 HAT reactions on the basis of DFT calculations. By simply analyzing the data set, we found that unsaturation effects are responsible for the poor performance of the BEP relationship, while the correlation between $\Delta G^{\ddagger}$ and $\Delta \chi^2$ is better than that between $\Delta G^{\ddagger}$ and $\Delta G_{r \times n}$. The simplified Roberts' equation proposed in our recent study also works here. Then, we used an ML method to establish a reactivity and selectivity prediction model based on appropriate descriptors. As a unique descriptor used in the ANN model, χ works better than the BDFE. Its performance is comparable with that of the simplified Roberts' equation. The introduction of %V_{buried} can improve the generalization ability between different H-extracting radical and make more accurate predictions, which shows the
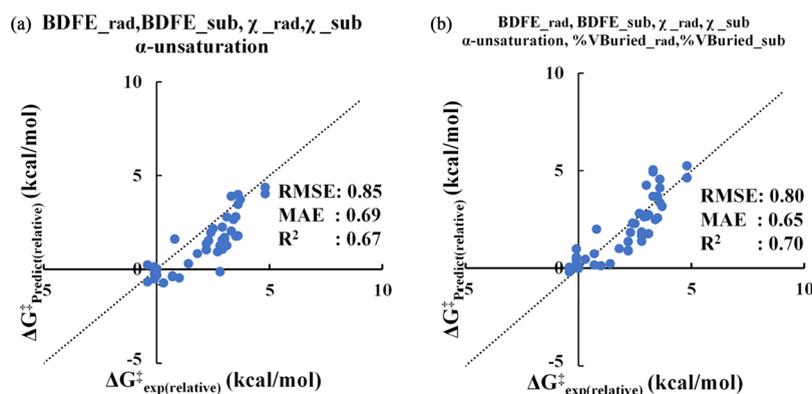
Figure 12. Experiment vs ANN-predicted barriers for the test set of 45 additional sp^3 C–Hs with CumO· by different descriptors (a) BDFE_sub, BDFE_rad, χ_sub, χ_rad, and α-unsaturation. (b) BDFE_sub, BDFE_rad, χ_sub, χ_rad, α-unsaturation, %Vburied_sub, and %Vburied_rad.

Site	Exp ratio	DFT	ANN
1	97	12.47	12.36
2	3	17.28	15.70

Site	Exp ratio	DFT	ANN
1	22	14.97	14.51
2	78	13.01	13.65

Site	Exp ratio	DFT	ANN
1	47	14.34	13.79
2	48	14.05	14.28
3	5	17.56	16.62

Site	Exp ratio	DFT	ANN
1	14	15.71	13.05
2	77	12.10	11.72
3	9	16.87	15.98

Site	Exp ratio	DFT	ANN
1	-	17.23	16.40
2	-	15.62	17.06
3	1	13.35	14.37

Site	Exp ratio	DFT	ANN
1	1	10.49	11.54
2	-	12.84	13.87
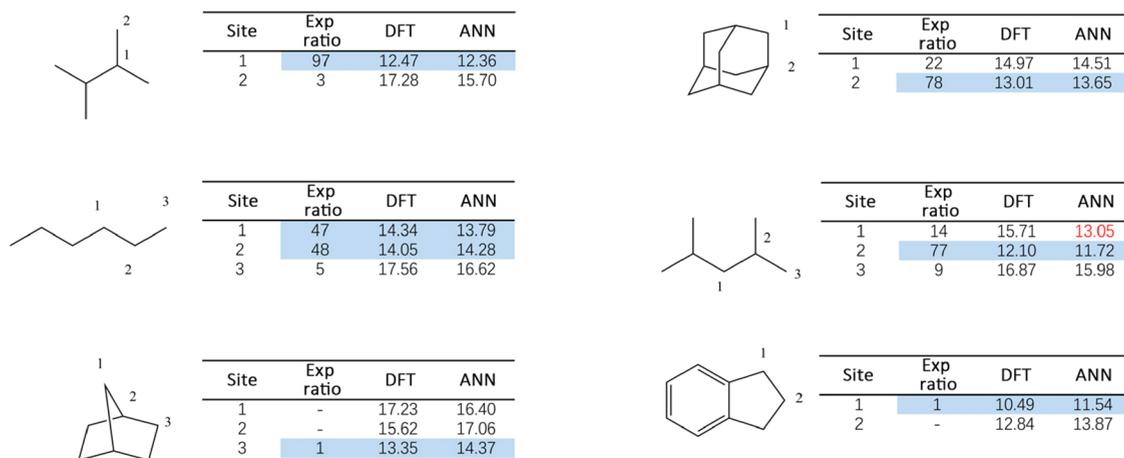
Figure 13. Evaluation of the ability to predict HAT reaction sites of CH$_3$O·.

importance of steric effects. The combination of BDFE, χ, α-unsaturation, and %V$_{buried}$ successively makes the R^2, MAE, and RMSE improve to 0.93, 0.54, and 0.68, respectively. The ANN model reproduces the experimental CumO· relative activation free energies and CH$_3$O· selectivities with good accuracy.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.2c03252.

Bietti's and Zuo's experimental results and the computational results of activation free energies and descriptors for HAT (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Fengjiao Liu − *School of Chemistry and Chemical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China; Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90095, United States;* ◉ orcid.org/0000-0002-2039-4786; Email: fjliu@sues.edu.cn

Authors

Siqi Ma − *School of Chemistry and Chemical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China*

Shipeng Wang − *School of Chemistry and Chemical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China*

Jiawei Cao − *School of Chemistry and Chemical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acsomega.2c03252

Notes

The authors declare no competing financial interest.

The code for the ANN model, the csv file of data sets, and the cartesian coordinates of DFT-computed structures are freely available on GitHub.[71] The instructions on how to use the model are also provided.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Ortiz de Montellano, P. R.; De Voss, J. J. Oxidizing species in the mechanism of cytochrome P450. *Nat. Prod. Rep.* **2002**, *19*, 477−493.

(2) Davies, H. M. L.; Beckwith, R. E. J. Catalytic Enantioselective C−H Activation by Means of Metal-Carbenoid-Induced C−H Insertion. *Chem. Rev.* **2003**, *103*, 2861−2904.

(3) Crabtree, R. H. Introduction to Selective Functionalization of C−H Bonds. *Chem. Rev.* **2010**, *110*, 575−575.

(4) Gutekunst, W. R.; Baran, P. S. C−H functionalization logic in total synthesis. *Chem. Soc. Rev.* **2011**, *40*, 1976−1991.

(5) Yamaguchi, J.; Yamaguchi, A. D.; Itami, K. C−H Bond Functionalization: Emerging Synthetic Tools for Natural Products and Pharmaceuticals. *Angew. Chem., Int. Ed.* **2012**, *51*, 8960−9009.

(6) Doyle, M. P. G.; Goldberg, K. I. C−H Functionalization. *Acc. Chem. Res.* **2012**, *45*, 777−777.

(7) Wencel-Delord, J.; Glorius, F. C−H bond activation enables the rapid construction and late-stage diversification of functional molecules. *Nat. Chem.* **2013**, *44* (), DOI: 10.1002/chin.201330250.

(8) Hartwig, J. F. Evolution of C−H bond functionalization from methane to methodology. *J. Am. Chem. Soc.* **2016**, *138*, 2−24.

(9) Qiu, Y.; Gao, S. Trends in applying C−H oxidation to the total synthesis of natural products. *Nat. Prod. Rep.* **2016**, *33*, 562−581.

(10) Mantry, L.; Maayuri, R.; Kumar, V.; Gandeepan, P. Photoredox catalysis in nickel-catalyzed C−H functionalization. *Beilstein J. Org. Chem.* **2021**, *17*, 2209−2259.

(11) Cao, S.; Hong, W.; Ye, Z.; Gong, L. Photocatalytic three-component asymmetric sulfonylation via direct C(sp3)-H functionalization. *Nat. Commun.* **2021**, *12*, 2377.

(12) Murray, R. W.; Jeyaraman, R. J. T. J. Dioxiranes: synthesis and reactions of methyldioxiranes. *J. Org. Chem.* **1985**, *50*, 2847−2853.

(13) Adam, W.; Curci, R.; Edwards, J. O. Dioxiranes: a new class of powerful oxidants. *Acc. Chem. Res.* **1989**, *22*, 205−211.

(14) Mkhalid, I. A.; Barnard, J. H.; Marder, T. B.; Murphy, J. M.; Hartwig, J. F. C−H activation for the construction of C-B bonds. *Chem. Rev.* **2010**, *110*, 890−931.

(15) Karimov, R. R.; Hartwig, J. F. Transition-Metal-Catalyzed Selective Functionalization of C(sp3)-H Bonds in Natural Products. *Angew. Chem., Int. Ed. Engl.* **2018**, *57*, 4234−4241.

(16) Li, J.; Zhang, Z.; Wu, L.; Zhang, W.; Chen, P.; Lin, Z.; Liu, G. Site-specific allylic C−H bond functionalization with a copper-bound N-centred radical. *Nature* **2019**, *574*, 516−521.

(17) Wang, Y.; Carder, H. M.; Wendlandt, A. E. Synthesis of rare sugar isomers through site-selective epimerization. *Nature* **2020**, *578*, 403−408.

(18) Vasilopoulos, A.; Krska, S. W.; Stahl, S. S. C(sp3)-H methylation enabled by peroxide photosensitization and Ni-mediated radical coupling. *Science* **2021**, *372*, 398−403.

(19) Li, H.; Yang, Y.; Jing, X.; He, C.; Duan, C. Triarylamine-based porous coordination polymers performing both hydrogen atom transfer and photoredox catalysis for regioselective α-amino C(sp3)-H arylation. *Chem. Sci.* **2021**, *12*, 8512−8520.

(20) Huang, C. Y.; Li, J.; Li, C. J. A cross-dehydrogenative C(sp3)-H heteroarylation via photo-induced catalytic chlorine radical generation. *Nat. Commun.* **2021**, *12*, 4010.

(21) Salamone, M.; Milan, M.; DiLabio, G. A.; Bietti, M. Reactions of the Cumyloxyl and Benzyloxyl Radicals with Tertiary Amides. Hydrogen Abstraction Selectivity and the Role of Specific Substrate-Radical Hydrogen Bonding. *J. Org. Chem.* **2013**, *78*, 5909−5917.

(22) Roberts, B. P. Polarity-reversal catalysis of hydrogen-atom abstraction reactions: concepts and applications in organic chemistry. *Chem. Soc. Rev.* **1999**, *30*, 25−35.

(23) Mayer, J. M. Understanding hydrogen atom transfer: from bond strengths to Marcus theory. *Acc. Chem. Res.* **2011**, *44*, 36−46.

(24) Chu, J. C. K.; Rovis, T. Complementary Strategies for Directed C(sp3)-H Functionalization: A Comparison of Transition-Metal-Catalyzed Activation, Hydrogen Atom Transfer, and Carbene/Nitrene Transfer. *Angew. Chem., Int. Ed.* **2018**, *57*, 62−101.

(25) Massimo, B. Activation and Deactivation Strategies Promoted by Medium Effects for Selective Aliphatic C−H Bond Functionalization. *Angew. Chem., Int. Ed.* **2018**, *57*, 16618−16637.

(26) Curci, R.; D'Accolti, L.; Fusco, C. A novel approach to the efficient oxygenation of hydrocarbons under mild conditions. Superior oxo transfer selectivity using dioxiranes. *Acc. Chem. Res.* **2006**, *39*, 1−9.

(27) Chen, K.; Eschenmoser, A.; Baran, P. S. Strain Release in C−H Bond Activation? *Angew. Chem., Int. Ed.* **2009**, *48*, 9705−9708.

(28) Zou, L.; Paton, R. S.; Eschenmoser, A.; Newhouse, T. R.; Baran, P. S.; Houk, K. Enhanced reactivity in dioxirane C−H oxidations via strain release: a computational and experimental study. *J. Org. Chem.* **2013**, *78*, 4037−4048.

(29) Kawamata, Y.; Yan, M.; Liu, Z.; Bao, D.; Chen, J.; Starr, J. T.; Baran, P. S. Scalable, electrochemical oxidation of unactivated C−H bonds. *J. Am. Chem. Soc.* **2017**, *139*, 7448−7451.

(30) Shaw, M. H.; Shurtleff, V. W.; Terrett, J. A.; Cuthbertson, J. D.; MacMillan, D. W. Native functionality in triple catalytic cross-coupling: sp3 C−H bonds as latent nucleophiles. *Science* **2016**, *352*, 1304−1308.

(31) Yang, Z.; Yu, P.; Houk, K. N. Molecular Dynamics of Dimethyldioxirane C−H Oxidation. *J. Am. Chem. Soc.* **2016**, *138*, 4237−4242.

(32) Curci, R.; Dinoi, A.; Rubino, M. F. Dioxirane oxidations: Taming the reactivity-selectivity principle. *Pure Appl. Chem.* **1995**, *67*, 811−822.

(33) Finn, M.; Friedline, R.; Suleman, N. K.; Wohl, C. J.; Tanko, J. M. Chemistry of the t-butoxyl radical: evidence that most hydrogen abstractions from carbon are entropy-controlled. *J. Am. Chem. Soc.* **2004**, *126*, 7578−7584.

(34) Salamone, M.; Mangiacapra, L.; Dilabio, G. A.; Bietti, M. Effect of metal ions on the reactions of the cumyloxyl radical with hydrogen atom donors. Fine control on hydrogen abstraction reactivity determined by Lewis acid-base interactions. *J. Am. Chem. Soc.* **2013**, *135*, 415−423.

(35) Salamone, M.; Bietti, M. 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.

(36) Salamone, M.; Carboni, G.; Mangiacapra, L.; Bietti, M. Binding to Redox-Inactive Alkali and Alkaline Earth Metal Ions Strongly Deactivates the C−H Bonds of Tertiary Amides toward Hydrogen Atom Transfer to Reactive Oxygen Centered Radicals. *J. Org. Chem.* **2015**, *80*, 9214−9223.

(37) Salamone, M.; Ortega, V. B.; Bietti, M. Enhanced Reactivity in Hydrogen Atom Transfer from Tertiary Sites of Cyclohexanes and Decalins via Strain Release: Equatorial C−H Activation vs Axial C−H Deactivation. *J. Org. Chem.* **2015**, *80*, 4710−4715.

(38) Brückl, T.; Baxter, R. D.; Ishihara, Y.; Baran, P. S. Innate and Guided C−H Functionalization Logic. *Acc. Chem. Res.* **2012**, *45*, 826−839.

(39) Henkelman, G.; Jónsson, H. Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *J. Chem. Phys.* **2000**, *113*, 9978−9985.

(40) Henkelman, G.; Uberuaga, B. P.; Jónsson, H. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys.* **2000**, *113*, 9901−9904.

(41) Evans, M. G.; Polanyi, M. Inertia and driving force of chemical reactions. *Trans. Faraday Soc.* **1938**, *34*, 11.

(42) Tedder, J. M. Which Factors Determine the Reactivity and Regioselectivity of Free Radical Substitution and Addition Reactions? *Angew. Chem., Int. Ed. Engl.* **1982**, *21*, 401.

(43) Mayer, J. M. Hydrogen Atom Abstraction by Metal-Oxo Complexes: Understanding the Analogy with Organic Radical Reactions. *Acc. Chem. Res.* **1998**, *31*, 441−450.

(44) de Visser, S. P.; Kumar, D.; Cohen, S.; Shacham, R.; Shaik, S. A Predictive Pattern of Computed Barriers for C−H Hydroxylation by Compound I of Cytochrome P450. *J. Am. Chem. Soc.* **2004**, *126*, 8362−8363.

(45) Yang, J. D.; Ji, P.; Xue, X. S.; Cheng, J. P. Recent Advances and Advisable Applications of Bond Energetics in Organic Chemistry. *J. Am. Chem. Soc.* **2018**, *140*, 8611−8623.

(46) Liu, F.; Yang, Z.; Yu, Y.; Mei, Y.; Houk, K. N. Bimodal Evans-Polanyi Relationships in Dioxirane Oxidations of sp3 C−H: Non-Perfect Synchronization in Generation of Delocalized Radical Intermediates. *J. Am. Chem. Soc.* **2017**, *139*, 16650−16656.

(47) Bernasconi, C. F. The principle of imperfect synchronization: I. Ionization of carbon acids. *Tetrahedron* **1985**, *41*, 3219−3234.

(48) Salamone, M.; Galeotti, M.; Romero-Montalvo, E.; van Santen, J. A.; Groff, B. D.; Mayer, J. M.; DiLabio, G. A.; Bietti, M. Bimodal Evans−Polanyi Relationships in Hydrogen Atom Transfer from C (sp3)−H Bonds to the Cumyloxyl Radical. A Combined Time-Resolved Kinetic and Computational Study. *J. Am. Chem. Soc.* **2021**, *143*, 11759−11776.

(49) Roberts, B. P.; Steel, A. J. An extended form of the Evans−Polanyi equation: a simple empirical relationship for the prediction of activation energies for hydrogen-atom transfer reactions. *J. Chem. Soc., Perkin Trans. 2* **1994**, *10*, 2155−2162.

(50) Liu, F.; Ma, S.; Lu, Z.; Nangia, A.; Duan, M.; Yu, Y.; Xu, G.; Mei, Y.; Bietti, M.; Houk, K. N. Hydrogen Abstraction by Alkoxyl Radicals: Computational Studies of Thermodynamic and Polarity Effects on Reactivities and Selectivities. *J. Am. Chem. Soc.* **2022**, *144*, 6802−6812.

(51) Govindaraju, R. Artificial Neural Networks in Hydrology. II: Hydrologic Applications. *J. Hydrol. Eng.* **2000**, *5* (2), 124−137.

(52) Keisuke, T.; Itsuki, M. Rapid estimation of activation energy in heterogeneous catalytic reactions via machine learning. *J. Comput. Chem.* **2018**, *39*, 2405.

(53) Granda, J. M.; Donina, L.; Dragone, V.; Long, D. L.; Cronin, L. Controlling an organic synthesis robot with machine learning to search for new reactivity. *Nature* **2018**, *559*, 377−381.

(54) Yang, L.; Li, X.; Zhang, S.; Hong, X. Machine learning prediction of hydrogen atom transfer reactivity in photoredox-mediated C−H functionalization. *Org. Chem. Front.* **2021**, *8*, 6187−6195.

(55) Hu, A.; Guo, J. J.; Pan, H.; Zuo, Z. Selective functionalization of methane, ethane, and higher alkanes by cerium photocatalysis. *Science* **2018**, *361*, 668−672.

(56) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16 Rev. B.01*; Gaussian, Inc.: Wallingford, CT, 2016.

(57) Chai, J.-D.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615−6620.

(58) Hariharan, P. C.; Pople, J. A. The influence of polarization functions on molecular orbital hydrogenation energies. *Theor. Chim. Acta* **1973**, *28*, 213−222.

(59) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A. Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. *J. Chem. Phys.* **1982**, *77*, 3654.

(60) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J. Phys. Chem. B* **2009**, *113*, 6378−6396.

(61) Clavier, H.; Nolan, S. P. Percent buried volume for phosphine and N-heterocyclic carbene ligands: steric properties in organo-metallic chemistry. *Chem. Commun.* **2010**, *46*, 841−861.

(62) Rumelhart, D. E.; Hinton, G. E.; Williams, R. J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533−536.

(63) Govindaraju, R. S. Artificial neural networks in hydrology. I: Preliminary concepts. *J. Hydrol. Eng.* **2000**, *5*, 115−123.

(64) Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *International Joint Conference on Artificial Intelligence*; Montreal, Canada, 1995; pp 1137−1145.

(65) Levenberg, K. A method for the solution of certain non-linear problems in least squares. *Q. Appl. Math.* **1944**, *2*, 164−168.

(66) Marquardt, D. W. An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. Ind. Appl. Math.* **1963**, *11*, 431−441.

(67) Williams, R. J.; Zipser, D. A learning algorithm for continually running fully recurrent neural networks. *Neural Comput.* **1989**, *1*, 270−280.

(68) Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5−32.

(69) Li, X.; Wang, Y.; Basu, S.; Kumbier, K.; Yu, B. A debiased MDI feature importance measure for random forests. *ArXiv* **2019**.

(70) An, Q.; Wang, Z.; Chen, Y.; Wang, X.; Zhang, K.; Pan, H.; Liu, W.; Zuo, Z. Cerium-Catalyzed C−H Functionalizations of Alkanes Utilizing Alcohols as Hydrogen Atom Transfer Agents. *J. Am. Chem. Soc.* **2020**, *142*, 6216−6226.

(71) Datasets Models, **2022**. https://github.com/ScottMagit/Neural_Network (accessed May, 2022).