General QSPR Protocol for Atomic/Inter-atomic Properties Predictions: Fragments based Graph Convolutional Neural Network (F-GCN)

Peng Gao,[†] Jie Zhang,[*,‡,¶] Hongbo Qiu,[§] and Shuaifei Zhao[‖]

†*School of Chemistry and Molecular Bioscience, University of Wollongong, NSW 2500, Australia*

‡*Centre of Chemistry and Chemical Biology, Bioland Laboratory (Guangzhou Regenerative Medicine and Health-Guangdong Laboratory), Guangzhou 53000, China*

¶*School of Chemical Engineering, East China University of Science and Technology, Shanghai 200237, China*

§*Department of Chemical Engineering, Monash University, Clayton, VIC 3800, Australia*

‖*Deakin University, Geelong, Institute for Frontier Materials, VIC 3216, Australia*

E-mail: j.chang@ecust.edu.cn

Abstract

In this study, a general quantitative structure-property relationship (QSPR) protocol, fragments based graph convolutional neural network (F-GCN), was developed for atomic and inter-atomic properties predictions. We applied this novel artificial intelligence (AI) tool in NMR chemical shifts and bond dissociation energies (BDEs) predictions. The predicted results were comparable to experimental measurement, while the computational cost was substantially reduced, with respect to pure density functional theory (DFT) calculations. The two important features of F-GCN can be summarised as: first, it could utilise different levels of molecular fragments centered at the target chemical bonds for atomic and inter-atomic information extraction; second, the designed architecture is also open to include additional descriptors for more accurate solution of chemical environment, making itself more efficient for local properties descriptions. And during our test, the averaged prediction error of ^{1}H NMR chemical shifts can be as small as 0.32 ppm; and the error of C−H BDEs estimations, is 2.7 kcal/mol. Moreover, we further demonstrated the applicability of this developed F-GCN model via several challenging structural assignments. The success of the F-GCN in atomic and inter-atomic predictions also indicates an essential improvement of computational chemistry with the assistance of AI tools.

Introduction

Over the past decade, artificial intelligence (AI) has become an essential component of human life, due to its capability of performing challenging tasks. And in recent years, increasingly more AI tools have been applied to improve the efficiency of physical and chemical research; and due to them, many of the complex or computationally expensive work in atomic simulations, kinetics, photocatalysis, adsorptions, etc, were largely facilitated.[1–15] In the current stage, with the development of high-performance graph convolutional neural networks (GCNs), accurate yet efficient predictions of molecular properties have become feasible.[16–18]

The behind reason lies in the fact that, via accurate solution of molecular graphs, the predicted properties can be accurately mapped with the chemical structures. Moreover, with the assistance of extra quantum mechanics (QM) descriptors, which are correlated with the atomic chemical environment, some kinds of neural networks are able to perform well on local properties predictions.[19,20] However, for molecules, the most important chemical insights are usually reflected by the fluctuations of atomic and inter-atomic properties; unfortunately, efficient ways of precisely solving the local environment within molecular graphs still remain to be a challenging research gap for computational scientists. Until now, the largest hurdle of extending the routine GCN based approaches to this kind of predictions lies in the fact that, inside molecular graphs, it is difficult for neural network itself to actively focus on the atomic information extraction, especially under the situation that no related descriptors can be utilised; and thus the target atomic environment or bonding connection cannot be differentiated effectively. To overcome this, more advanced GCN architectures are highly needed.

In the past, many computational chemists also tend to apply quantum mechanics (QM) methods for atomic and inter-atomic properties calculations, like NMR chemical shifts, bond dissociation energies (BDE), vibrational frequencies, etc.[21–31] And decent levels of QM theory could provide comparably accurate results with respect to experimental measurement. Unfortunately, their applications in real practice are usually limited by their high computational costs. Thus, these kinds of calculations can't be performed in large scales. In this study, we try to propose a novel GCN architecture, fragments based graph neural network (F-GCN), which is able to conduct accurate atomic and inter-atomic properties predictions with substantially less cost.

The essential feature of this proposed neural network lies in the fact that, beyond efficient solution of the molecular graph, it could also utilise the generated fragments for refined solution of chemical environment at atomic level. To demonstrate its performance, we tested it on NMR chemical shifts and BDEs predictions, and the obtained accuracy is at experimental

level. To note, both the NMR chemical shift and BDE are important reference values for challenging structural assignments, as almost all the chemical reactions are related with the change of chemical environment, as well as the formation or breaking of chemical bonds.[32–38] Accurate yet affordable predictions of these two important metrics can bring forth surprising convenience for chemical researchers.[39]

Some researchers also tried to apply message passing related neural networks for these kinds of predictions, and their models usually require extra large amount of data for training, most of which were generated via DFT calculations.[40,41] The 'intra-accuracy' can be fully guaranteed due to the abundant supply of training data. However, in real practice, the actual applicability of these models are usually limited in two aspects: first, the prediction results usually don't match well with experimental measurements, due to the sourced errors of DFT; secondly, it is difficult for such kinds of models to perform well on few-shot learning cases. To overcome these hurdles, the proposed F-GCN was developed on experimental data, and its architecture was designed to be applicable on small data sets. Moreover, its original framework was also developed to be flexible to include more decent descriptors, like DFT calculated items, to further enhance its performance on challenging assignments.

The architecture of F-GCN is open and applicable for various kinds of researchers; and we believe, beyond NMR chemical shifts and BDEs, it can also be extended to other atomic or inter-atomic properties. The work presented in this study actually indicates an essential progress of computational chemistry with the assistance of advanced AI tools.

Methodology

The architecture of F-GCN

Graph based neural networks have become increasingly popular for accurate predictions of molecular properties.[16–18] It could efficiently build the correlation between the molecular structures and the target property. However, to further extend such a novel tool to atomic and inter-atomic properties predictions, some essential modifications of the original architecture are needed.

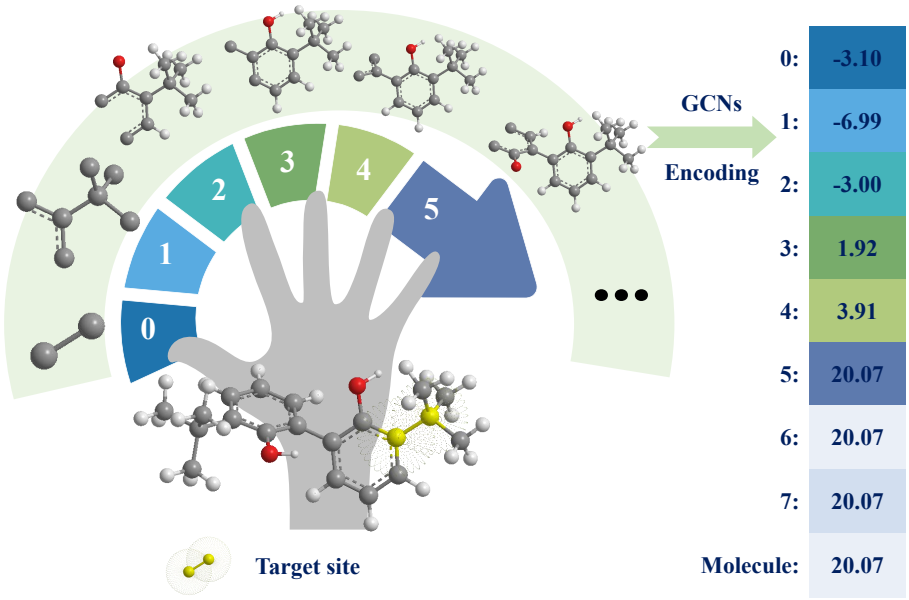

Figure 1: The generation and encoding of the fragmentary graphs.

In this study, we developed a novel architecture to include multi-level molecular fragments for chemical environment solution at atomic level. The generation of fragmentary graphs was described in Figure 1, starting from the target site, a single molecule can be transformed into fragments along its bonding structure at different levels; and then all these fragments can be further encoded by independent GCNs.[42,43] That is to say, the target bond can be marked more pertinently. The specific schematic of this proposed F-GCN is illustrated in

Figure 2.

Before input to GCN, all the molecules or fragments were transformed into nodes and edges via the modified *TencentAlchemyDataset* within the DGL library.[44,45] Each graph is represented by nodes and edges. The features involving the nodes are associated with atomic descriptors; while the features of edges are corresponding to the bonds or connections among atom pairs. Within the framework of F-GCN, fully connected molecular graph is applied for molecules representation. That is to say, inside any molecule or fragment, all the bonding connections among every two atoms will be recorded by distance tensors at the radial basis function (RBF) layer. Moreover, within the whole framework, several continuous-filter convolutions layers were also applied to further optimise the inter-atomic evolution, thus atoms' chemical environment can be better solved. The $\boldsymbol{i}$th atom's representation at $\boldsymbol{l+1}$ layer can be expressed using the equation below:

$$a_i^{l+1} = \sum_{j=0}^{N} a_j^l \circ \omega^l(d_{ij}) \tag{1}$$

where, ω^l indicates the filter-generation, which actually maps the atoms' representations to the corresponding filter bank; and $\circ$ means element-wise multiplication. To reasonably guarantee the optimisation efficiency and control the evolution of the filter values, a Gaussian function, g_k, is applied, the form of which is presented below:

$$g_k(d_{ij}) = exp(-\alpha(d_{ij} - \mu_k)^2) \tag{2}$$

where, μ_k represents a certain value of cutoff, and d_{ij} indicates the distance between the $\boldsymbol{i}$th and $\boldsymbol{j}$th atom. In this study, the value of hyper parameter α was set to 0.1.[16]

To make the developed F-GCN better focus on local environment solution, in the readout stage (R-S), a summation of all fragmentary contributions, along with the routine descriptors that are generated by RDKit,[46] will be further conducted by a dense neural network unit (more details can be seen in Figure 2). It is also worth noting that, within our frame,

some extra yet related chemical knowledge, like QM calculated descriptors,[19–22,47] can also be incorporated freely to further augment the performance of F-GCN.

Once the predicted value P is obtained, a function of squared loss with respect to experimental measurement P' is used to represent the prediction accuracy.

$$L(P, P') = (P - P')^2 \tag{3}$$

Actually, this proposed architecture is flexible for various kinds of applications. More technical details can be found on https://github.com/jeah-z/BDE-FGCN.

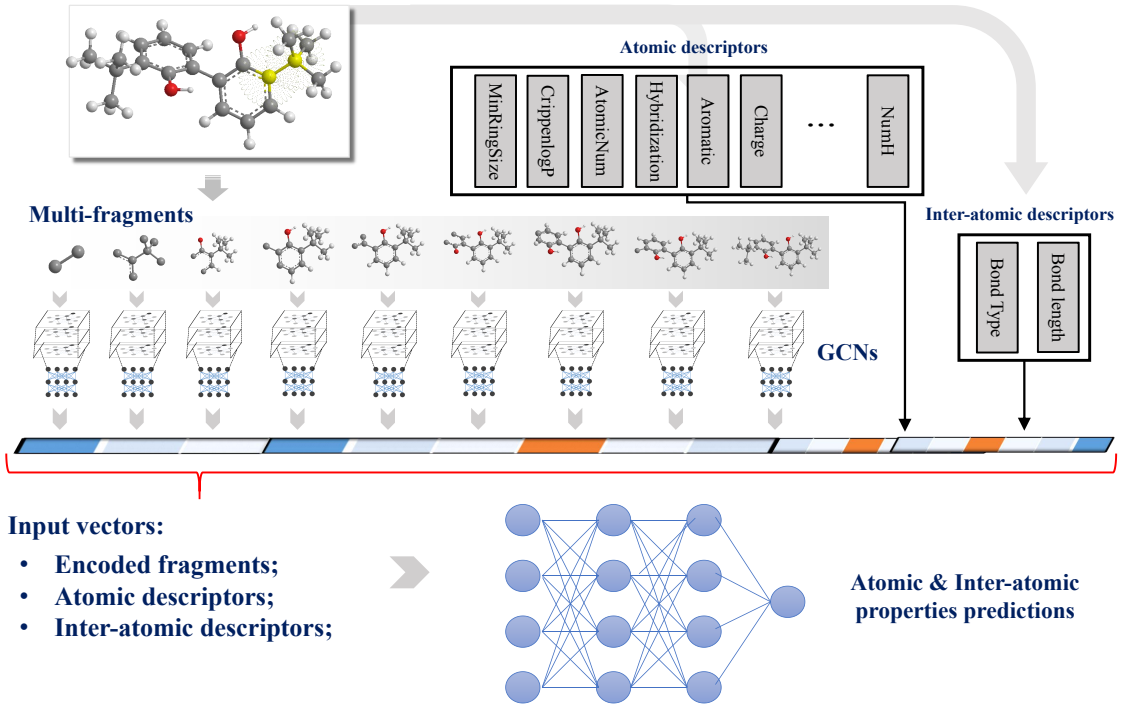

Figure 2: Illustration of the F-GCN's workflow, designed for experimental BDEs predictions.

Results and discussion

The overall performance of the F-GCN model in NMR chemical shifts predictions

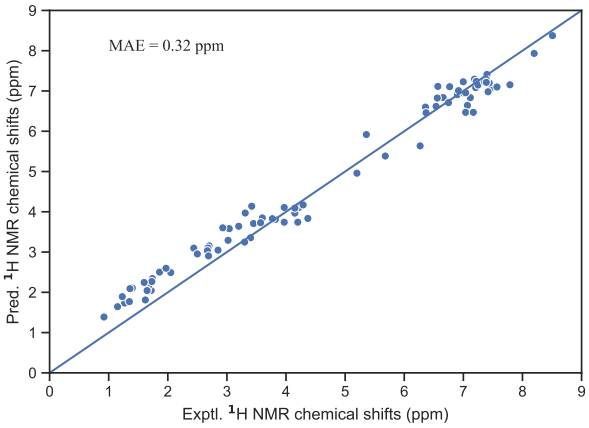

Figure 3: Comparison between the predicted and experimental ^{1}H NMR chemical shifts. There are 997 ^{1}H chemical shifts in the original data set; all the experimental data were taken from Refs.[48,49]

We applied F-GCN for predictions of ^{1}H NMR chemical shifts, the results were presented in Figure 3. The prediction accuracy of F-GCN is comparable to other GCN based approaches.[41] It was worth noting that, compared to other architectures, F-GCN is developed for few-shot learning, thus it doesn't require huge amount of data for model training. Moreover, in our previous studies, we focused on combining QM descriptors with structural information,[19,20] to conduct this kind of predictions, and higher accuracy was obtained; however, running QM calculations is usually time expensive. With the introduction of F-GCN, accurate predictions can be realised without the assistance of QM calculations, thus computational cost can be substantially reduced, indicating an initial yet important step for the application of AI tools in computational chemistry. The essential reason for the success of F-GCN in atomic properties predictions can be attributed to its special architecture, which starts processing molecules from target site, and then extends to the overall structure

through multiple-level fragments. All these fragmentary graphs were encoded by independent neural networks, thus the properties of the local site can be well emphasised within the framework of F-GCN. Then with the assistance of atomic descriptors extracted by RDKit, prediction accuracy can be further improved. We hope such a promising architecture can be conveniently applied by more computational and experimental researchers to predict other challenging properties in the future.

The overall performance of the F-GCN model in BDE predictions

We also tested the developed F-GCN via BDE predictions on various kinds of chemical bonds; and its overall performance was presented in Figure 4 and Table 1. The mean absolute errors (MAEs) for these listed bonds are all within a reasonable range; and the plot of error distribution is close to a Gaussian function, indicating the systematic errors had been well controlled with this proposed architecture. In real practice, F-GCN can be applied as a useful quantitative structure–property relationships (QSPR) model to detect a molecule's weakest bond, and further identify the possible reactive site for experimental researchers. Furthermore, to conduct a systematic benchmark with another popular approach on BDE predictions, we also loaded the ALFABET model, trained by John *et al* on 290,664 DFT calculated BDEs of single bonds.[40] Our proposed F-GCN model shows a higher accuracy and applicability on 2800s experimental BDEs predictions; and it was also proved to be superior in few-shot learning. The essential reason lies in the fact that there do exist sourced deviations between DFT calculations and experimental values, and such kinds of errors can not be easily eliminated via purely improving the DFT methods. That said using DFT calculated data for training is not sufficient for GCN based approaches to reach the accuracy of experimental level; and moreover, some sourced errors of DFT calculations, like the ones caused by relativistic effects,[50] may be introduced to the model. Therefore, to make the developed model experimentally meaningful, we propose that applying experimental data for training is more suitable in the case of BDEs predictions.

However, we also need to point out that, for some compounds in the test set, the larger deviations between their predicted and experimental BDEs are mainly caused by the lower molecular diversity of the training set. And, to overcome this, further enriching the original data set via covering more kinds of molecular structures will be helpful.

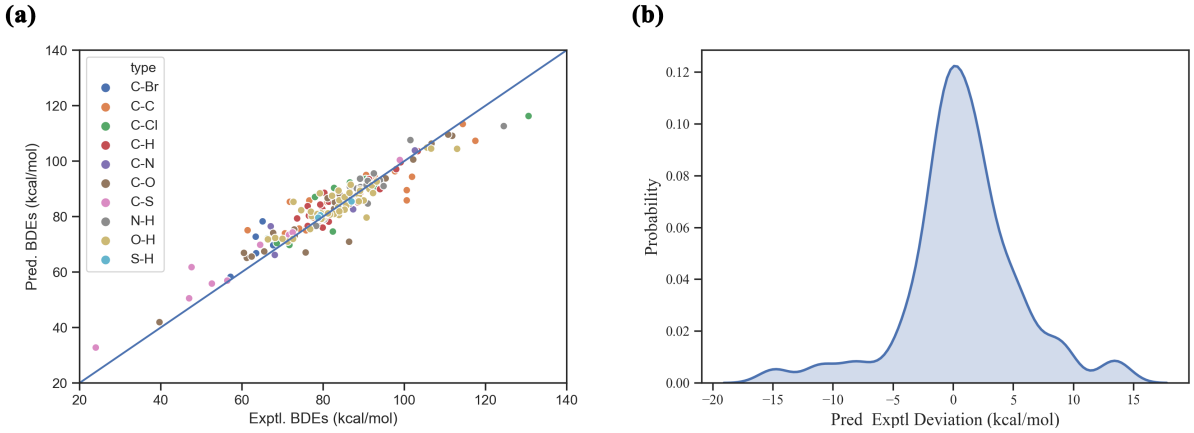

Figure 4: (a). Comparison between the predicted and experimental BDEs of the test set; (b). Distributions of errors between the predicted and experimental BDEs.

Table 1: The mean absolute errors (MAEs, in kcal/mol) for different kinds of predicted BDEs with respect to experimental values.

Bonds[a)]	F-GCN[b)]	ALFABET[c)]
C-C	4.45	5.85
C-H	2.71	3.07
O-H	2.47	4.13
C-X (X=N,O)	3.58	4.33
N-H	3.01	4.03
C-S	4.16	N/A
C-X (X=F,Cl,Br)	5.87	N/A

[a)] The experimental BDE data were obtained from iBond 2.0 databank.[51] In total, 2400s bonds were applied to develop the F-GCN model, among them 90% were used as the training set, and the remaining 10% as the test set; [b)] The MAEs of the developed F-GCN model on the test set; [c)] The MAEs of the ALFABET model by John *et al*,[40] on the test set;

To further demonstrate the high capability of F-GCN in processing the fragmentary graphs, we presented the encoding results of a test set (see Figure 5, and more details of the applied molecules can be found in supporting information). From Figure 5a, we can see

that F-GCN can successfully encode all the test bonds with distinguishable values among different levels of fragments. And moreover, in Figure 5b, we took C−C bonds for instance, and found that F-GCN also showed high performance in recognising the specific chemical environment of C−C bonds in various kinds of molecules. Therefore, we could reasonably conclude that F-GCN is a powerful tool for bonds differentiation, and will also be useful for atomic and inter-atomic properties predictions.

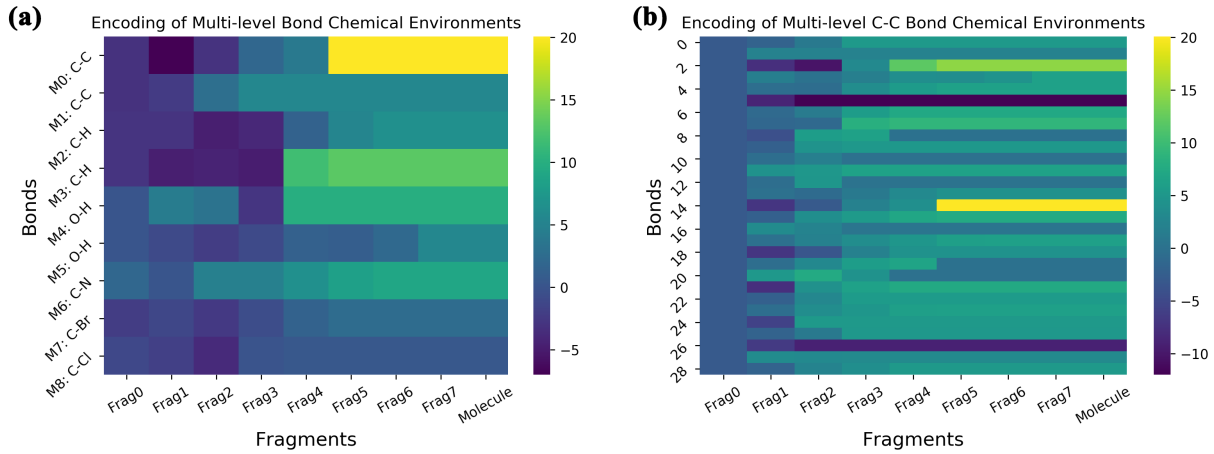

Figure 5: (a). Encoding of various kinds of chemical bonds by F-GCN at multi-level fragments; (b). Encoding of different C−C bonds by F-GCN at multi-level fragments.

Accurate estimations of phenol O−H BDEs

Phenol based inhibitors have been proved useful for the retardation of polymer oxidation; and theO−H bonds tend to be attacked by peroxyl radicals, thus the corresponding dissociation energy is an important index to characterise the performance of the inhibitor.[52] Unfortunately, there are only a few reference O−H BDE values for synthesis researchers to use, and the experimental measurement of this kind of BDEs via kinetic analysis is usually time-consuming. With F-GCN, accurate estimations of phenol O−H BDEs can be essentially facilitated and fastened. In Figure 6 and 7, we presented the predicted and experimental O−H BDEs for several classic phenol compounds; and the errors are all within a small range, indicating that F-GCN is a reliable tool for this kind of characterisations. To note, for O−H

or other hydrogen involved bonds, F-GCN is especially accurate; and the chemical environment of this kinds of bonds is relatively simpler, due to the fact that, hydrogen atom can only be bound with one atom.

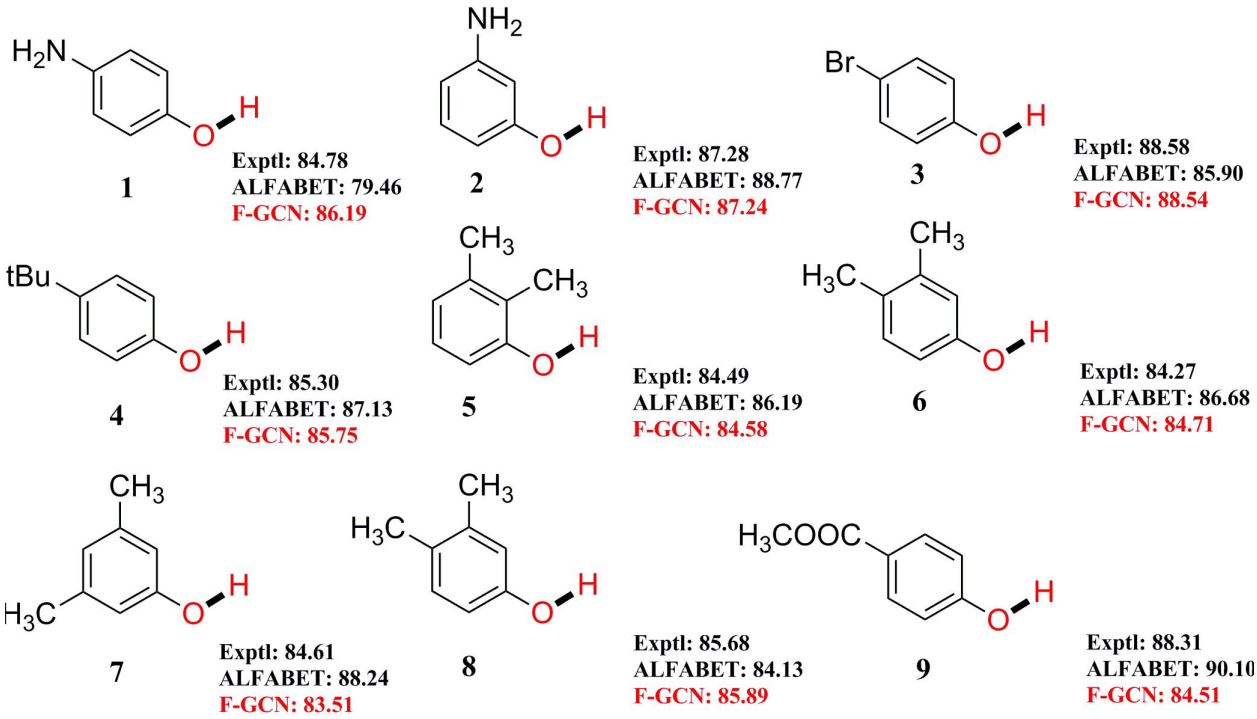

Figure 6: The structures of the selected molecules with phenol O−H bonds.

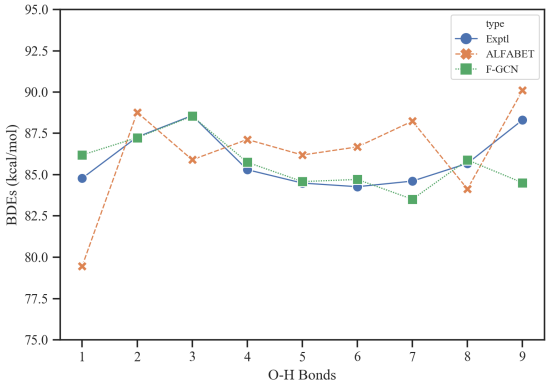

Figure 7: Comparison between the predicted and experimental BDEs (in kcal/mol) of the selected phenol O−H bonds listed in Figure 6.

Assisting in stereoselectivity analysis of functional C−H bonds

To further demonstrate the applicability of F-GCN, we tested its performance on a series of organic compounds with more complex structures, which are used for methodology development of site-selective C−H functionalizations.[53–57] The specific reaction was described in Figure 8.

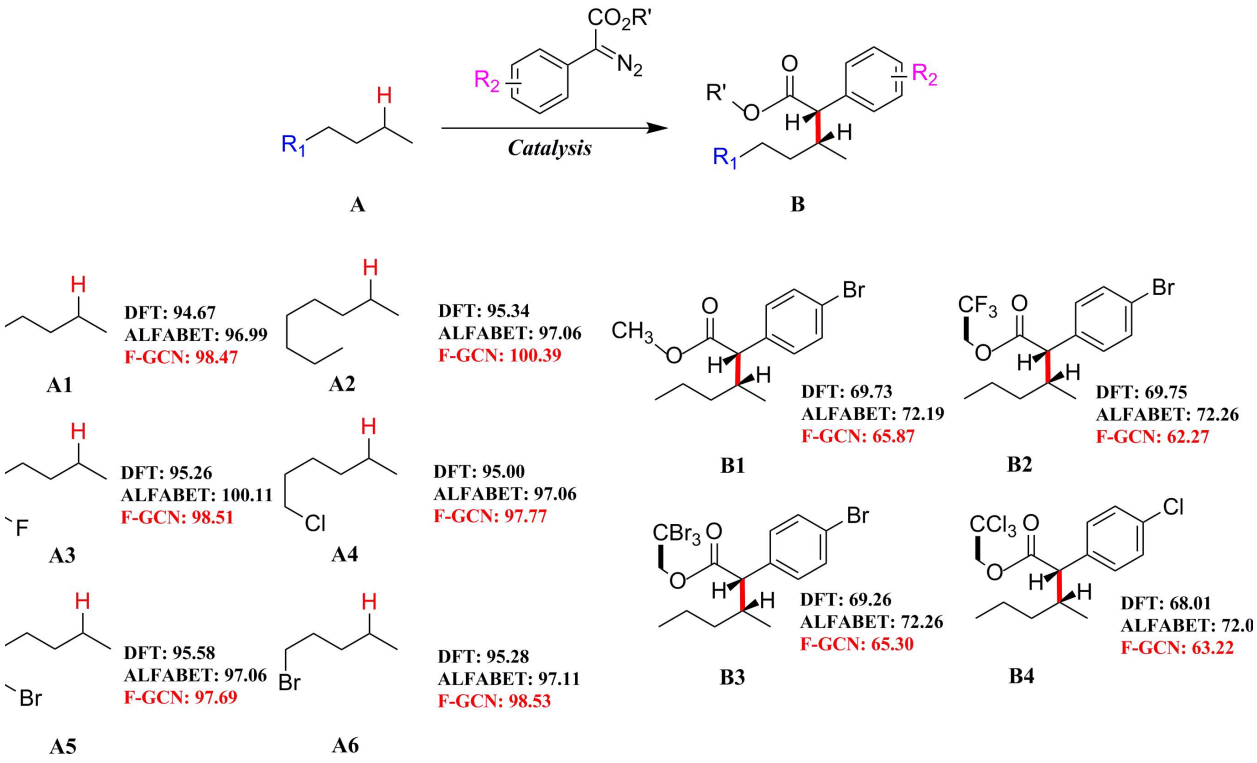

Figure 8: The structures of the selected molecules with functional C−H and C−C bonds.

Accurate predictions of the BDEs of the target C−H and C−C bonds with affordable cost will undoubtedly bring valuable chemical insights for synthesis researchers. We applied the developed model to predict the BDEs of the C−H and C−C bonds, the results were shown in Figure 8 and 9. We can see that the prediction accuracy of F-GCN is close to the level of DFT calculation, and the corresponding computational cost is essentially reduced, indicating the high reliability and convenience of our proposed approach for this kind of structural assignments. With such a useful tool, the development of high throughput screening (HTS) technology can be further prompted, as accurate estimations of various bonding energies can

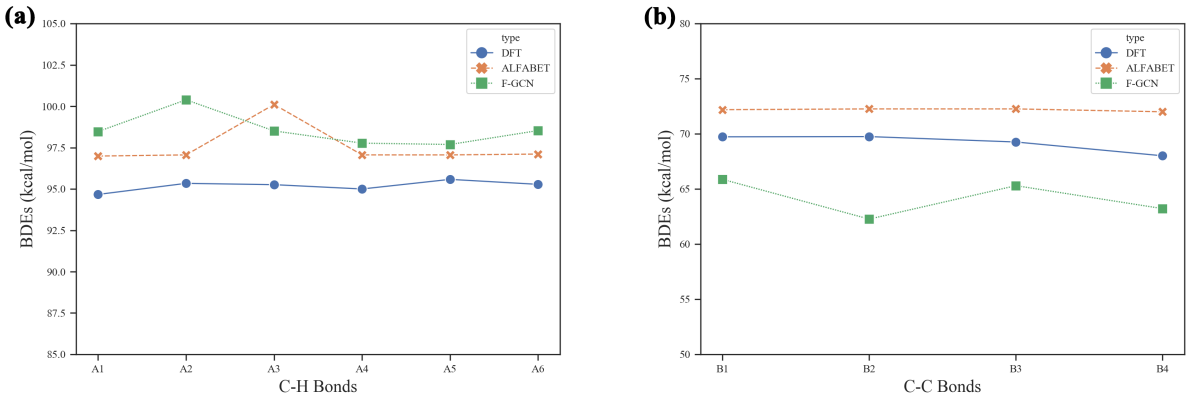

Figure 9: Comparison between the predicted and DFT calculated BDEs (in kcal/mol) of the selected C−H and C−C bonds listed in Figure 8; all the DFT calculations were performed at M062X/def2-TZVP within G09.

be realised in batches; and therefore, the actual feasibility of synthesis routes can be well assessed beforehand. We hope in the future, more stereoselectivity analysis work could be significantly facilitated with the assistance of F-GCN.

Conclusion

To sum up, with the proposed F-GCN, accurate predictions of atomic and inter-atomic properties, like NMR chemical shifts, BDEs, etc, become available. Moreover, its performance can be further enhanced via inclusion of more advanced descriptors that are helpful to refine the solution of atomic environment. The applicability of this novel F-GCN was also tested by independent structural assignments. The work provided in this study actually indicates a promising prospect for artificial intelligence (AI) technologies in chemical research. However, to note, further optimisation of the original architecture of F-GCN still remains to be a goal; and in the near future, we hope to expand this proposed approach to other areas, and provide more valuable chemical insights.

CRediT Author Contribution Statement

P.G. and J.Z. outlined the whole project and designed the algorithm. P.G. and J.Z. carried out the calculations. J.Z. worked on the architecture development, and P.G. worked on the data analysis. H.Q. and S.Z collected the experimental data. P.G. and J.Z. contributed to the writing of the manuscript.

Conflicts of interest

There is no conflict of interest.

Acknowledgement

We thank Dr. Haibo Yu for providing critical feedback and useful suggestions to our work. We thank the NCI system, supported by the Australian Government (Project id: v15) to provide computational resource. We thank the Australian Government, which offered P.G an Australian International Postgraduate Award scholarship to complete his PhD study.

Supporting Information Available

More technical details can be found in supporting information and GitHub page: https://github.com/jeah-z/BDE-FGCN.

References

(1) Behler, J. Perspective: Machine learning potentials for atomistic simulations. *The Journal of Chemical Physics* **2016**, *145*, 170901.

(2) Behler, J. First Principles Neural Network Potentials for Reactive Simulations of Large

Molecular and Condensed Systems. *Angewandte Chemie International Edition* **2017**, *56*, 12828–12840.

(3) Wang, J.; Olsson, S.; Wehmeyer, C.; Pérez, A.; Charron, N. E.; de Fabritiis, G.; Noé, F.; Clementi, C. Machine Learning of Coarse-Grained Molecular Dynamics Force Fields. *ACS Central Science* **2019**, *5*, 755–767.

(4) Botu, V.; Batra, R.; Chapman, J.; Ramprasad, R. Machine Learning Force Fields: Construction, Validation, and Outlook. *The Journal of Physical Chemistry C* **2017**, *121*, 511–522.

(5) Meldgaard, S. A.; Kolsbjerg, E. L.; Hammer, B. Machine learning enhanced global optimization by clustering local environments to enable bundled atomic energies. *The Journal of Chemical Physics* **2018**, *149*, 134104.

(6) Ouyang, R.; Xie, Y.; Jiang, D.-e. Global minimization of gold clusters by combining neural network potentials and the basin-hopping method. *Nanoscale* **2015**, *7*, 14817–14821.

(7) Sørensen, K. H.; Jørgensen, M. S.; Bruix, A.; Hammer, B. Accelerating atomic structure search with cluster regularization. *The Journal of Chemical Physics* **2018**, *148*, 241734.

(8) Wexler, R. B.; Martirez, J. M. P.; Rappe, A. M. Chemical Pressure-Driven Enhancement of the Hydrogen Evolving Activity of Ni_2P from Nonmetal Surface Doping Interpreted via Machine Learning. *Journal of the American Chemical Society* **2018**, *140*, 4678–4683.

(9) Mansouri Tehrani, A.; Oliynyk, A. O.; Parry, M.; Rizvi, Z.; Couper, S.; Lin, F.; Miyagi, L.; Sparks, T. D.; Brgoch, J. Machine Learning Directed Search for Ultraincompressible, Superhard Materials. *Journal of the American Chemical Society* **2018**, *140*, 9844–9853.

(10) Panapitiya, G.; Avendaño-Franco, G.; Ren, P.; Wen, X.; Li, Y.; Lewis, J. P. Machine-Learning Prediction of CO Adsorption in Thiolated, Ag-Alloyed Au Nanoclusters. *Journal of the American Chemical Society* **2018**, *140*, 17508–17514.

(11) Rupp, M.; Ramakrishnan, R.; von Lilienfeld, O. A. Machine Learning for Quantum Mechanical Properties of Atoms in Molecules. *The Journal of Physical Chemistry Letters* **2015**, *6*, 3309–3313.

(12) Bai, Y.; Wilbraham, L.; Slater, B. J.; Zwijnenburg, M. A.; Sprick, R. S.; Cooper, A. I. Accelerated Discovery of Organic Polymer Photocatalysts for Hydrogen Evolution from Water through the Integration of Experiment and Theory. *Journal of the American Chemical Society* **2019**, *141*, 9063–9071.

(13) Ahneman, D. T.; Estrada, J. G.; Lin, S.; Dreher, S. D.; Doyle, A. G. Predicting reaction performance in C–N cross-coupling using machine learning. *Science* **2018**, *360*, 186–190.

(14) Mater, A. C.; Coote, M. L. Deep Learning in Chemistry. *Journal of Chemical Information and Modeling* **2019**, *59*, 2545–2559.

(15) Faber, F. A.; Hutchison, L.; Huang, B.; Gilmer, J.; Schoenholz, S. S.; Dahl, G. E.; Vinyals, O.; Kearnes, S.; Riley, P. F.; von Lilienfeld, O. A. Prediction Errors of Molecular Machine Learning Models Lower than Hybrid DFT Error. *Journal of Chemical Theory and Computation* **2017**, *13*, 5255–5264.

(16) Schütt, K. T.; Sauceda, H. E.; Kindermans, P.-J.; Tkatchenko, A.; Müller, K.-R. SchNet – A deep learning architecture for molecules and materials. *The Journal of Chemical Physics* **2018**, *148*, 241722.

(17) Lu, C.; Liu, Q.; Wang, C.; Huang, Z.; Lin, P.; He, L. Molecular Property Prediction: A Multilevel Quantum Interactions Modeling Perspective. *arXiv* **2019**, 1906.11081.

(18) Gao, P.; Zhang, J.; Sun, Y.; Yu, J. Accurate predictions of aqueous solubility of drug molecules via the multilevel graph convolutional network (MGCN) and SchNet architectures. *Phys. Chem. Chem. Phys.* **2020**, *22*, 23766–23772.

(19) Gao, P.; Zhang, J.; Peng, Q.; Zhang, J.; Glezakou, V.-A. General Protocol for the Accurate Prediction of Molecular $^{13}C/^{1}H$ NMR Chemical Shifts via Machine Learning Augmented DFT. *Journal of Chemical Information and Modeling* **2020**, *60*, 3746–3754.

(20) Gao, P.; Zhang, J.; Sun, Y.; Yu, J. Toward Accurate Predictions of Atomic Properties via Quantum Mechanics Descriptors Augmented Graph Convolutional Neural Network: Application of This Novel Approach in NMR Chemical Shifts Predictions. *The Journal of Physical Chemistry Letters* **2020**, *11*, 9812–9818.

(21) Gao, P.; Wang, X.; Yu, H. Towards an Accurate Prediction of Nitrogen Chemical Shifts by Density Functional Theory and Gauge-Including Atomic Orbital. *Advanced Theory and Simulations* **2019**, *2*, 1800148.

(22) Gao, P.; Wang, X.; Huang, Z.; Yu, H. ^{11}B NMR Chemical Shift Predictions via Density Functional Theory and Gauge-Including Atomic Orbital Approach: Applications to Structural Elucidations of Boron-Containing Molecules. *ACS Omega* **2019**, *4*, 12385–12392.

(23) Harris, N. J.; Lammertsma, K. Ab Initio Density Functional Computations of Conformations and Bond Dissociation Energies for Hexahydro-1,3,5-trinitro-1,3,5-triazine. *Journal of the American Chemical Society* **1997**, *119*, 6583–6589.

(24) Feng, Y.; Liu, L.; Wang, J.-T.; Huang, H.; Guo, Q.-X. Assessment of Experimental Bond Dissociation Energies Using Composite ab Initio Methods and Evaluation of the Performances of Density Functional Methods in the Calculation of Bond Dissociation Energies. *Journal of Chemical Information and Computer Sciences* **2003**, *43*, 2005–2013.

(25) Zhao, Y.; Truhlar, D. G. How Well Can New-Generation Density Functionals Describe the Energetics of Bond-Dissociation Reactions Producing Radicals? *The Journal of Physical Chemistry A* **2008**, *112*, 1095–1099.

(26) Izgorodina, E. I.; Brittain, D. R. B.; Hodgson, J. L.; Krenske, E. H.; Lin, C. Y.; Namazian, M.; Coote, M. L. Should Contemporary Density Functional Theory Methods Be Used to Study the Thermodynamics of Radical Reactions? *The Journal of Physical Chemistry A* **2007**, *111*, 10754–10768.

(27) Goerigk, L.; Grimme, S. A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions. *Phys. Chem. Chem. Phys.* **2011**, *13*, 6670–6688.

(28) Neese, F.; Schwabe, T.; Kossmann, S.; Schirmer, B.; Grimme, S. Assessment of Orbital-Optimized, Spin-Component Scaled Second-Order Many-Body Perturbation Theory for Thermochemistry and Kinetics. *Journal of Chemical Theory and Computation* **2009**, *5*, 3060–3073.

(29) Goerigk, L.; Hansen, A.; Bauer, C.; Ehrlich, S.; Najibi, A.; Grimme, S. A look at the density functional theory zoo with the advanced GMTKN55 database for general main group thermochemistry, kinetics and noncovalent interactions. *Phys. Chem. Chem. Phys.* **2017**, *19*, 32184–32215.

(30) 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.

(31) Goerigk, L.; Grimme, S. Efficient and Accurate Double-Hybrid-Meta-GGA Density Functionals—Evaluation with the Extended GMTKN30 Database for General Main Group Thermochemistry, Kinetics, and Noncovalent Interactions. *Journal of Chemical Theory and Computation* **2011**, *7*, 291–309.

(32) Gani, T. Z. H.; Kulik, H. J. Understanding and Breaking Scaling Relations in Single-Site Catalysis: Methane to Methanol Conversion by FeIVO. *ACS Catalysis* **2018**, *8*, 975–986.

(33) Lin, C. Y.; Marque, S. R. A.; Matyjaszewski, K.; Coote, M. L. Linear-Free Energy Relationships for Modeling Structure–Reactivity Trends in Controlled Radical Polymerization. *Macromolecules* **2011**, *44*, 7568–7583.

(34) Bian, C.; Wang, S.; Liu, Y.; Jing, X. Thermal stability of phenolic resin: new insights based on bond dissociation energy and reactivity of functional groups. *RSC Adv.* **2016**, *6*, 55007–55016.

(35) Kim, S.; Chmely, S. C.; Nimlos, M. R.; Bomble, Y. J.; Foust, T. D.; Paton, R. S.; Beckham, G. T. Computational Study of Bond Dissociation Enthalpies for a Large Range of Native and Modified Lignins. *The Journal of Physical Chemistry Letters* **2011**, *2*, 2846–2852.

(36) Drew, K. L.; Reynisson, J. The impact of carbon–hydrogen bond dissociation energies on the prediction of the cytochrome P450 mediated major metabolic site of drug-like compounds. *European Journal of Medicinal Chemistry* **2012**, *56*, 48–55.

(37) Blanksby, S. J.; Ellison, G. B. Bond Dissociation Energies of Organic Molecules. *Accounts of Chemical Research* **2003**, *36*, 255–263.

(38) Hartwig, J. F. Catalyst-Controlled Site-Selective Bond Activation. *Accounts of Chemical Research* **2017**, *50*, 549–555.

(39) Yao, K.; Herr, J. E.; Brown, S. N.; Parkhill, J. Intrinsic Bond Energies from a Bonds-in-Molecules Neural Network. *The Journal of Physical Chemistry Letters* **2017**, *8*, 2689–2694.

(40) St. John, P. C.; Guan, Y.; Kim, Y.; Kim, S.; Paton, R. S. Prediction of organic homolytic bond dissociation enthalpies at near chemical accuracy with sub-second computational cost. *Nat Commun* **2020**, *11*, 2328.

(41) Kwon, Y.; Lee, D.; Choi, Y.-S.; Kang, M.; Kang, S. Neural Message Passing for NMR Chemical Shift Prediction. *Journal of Chemical Information and Modeling* **2020**, *60*, 2024–2030.

(42) Cereto-Massagué, A.; Ojeda, M. J.; Valls, C.; Mulero, M.; Garcia-Vallvé, S.; Pujadas, G. Molecular fingerprint similarity search in virtual screening. *Methods* **2015**, *71*, 58–63, Virtual Screening.

(43) Myint, K.-Z.; Wang, L.; Tong, Q.; Xie, X.-Q. Molecular Fingerprint-Based Artificial Neural Networks QSAR for Ligand Biological Activity Predictions. *Molecular Pharmaceutics* **2012**, *9*, 2912–2923.

(44) Wang, M.; Zheng, D.; Ye, Z.; Gan, Q.; Li, M.; Song, X.; Zhou, J.; Ma, C.; Yu, L.; Gai, Y.; Xiao, T.; He, T.; Karypis, G.; Li, J.; Zhang, Z. Deep Graph Library: A Graph-Centric, Highly-Performant Package for Graph Neural Networks. 2019.

(45) Chen, G.; Chen, P.; Hsieh, C.-Y.; Lee, C.-K.; Liao, B.; Liao, R.; Liu, W.; Qiu, J.; Sun, Q.; Tang, J.; Zemel, R.; Zhang, S. Alchemy: A Quantum Chemistry Dataset for Benchmarking AI Models. *arXiv preprint arXiv:1906.09427* **2019**,

(46) Landrum, G. A. RDKit: Open-source cheminformatics software. *http://www.rdkit.org* **2018**,

(47) Gao, P.; Zhang, J.; Chen, H. A systematic benchmarking of ^{31}P and ^{19}F NMR chemical shift predictions using different DFT/GIAO methods and applying linear regression to improve the prediction accuracy. *International Journal of Quantum Chemistry* **2020**, e26482.

(48) CHESHIRE CCAT, the Chemical Shift Repository for computed NMR scaling factors, with Coupling Constants Added Too. 2017; http://cheshirenmr.info/index.htm.

(49) Structure Determination Using Spectroscopic Methods. 2017; https://www.chem.wisc.edu/areas/reich/nmr/.

(50) Latypov, S. K.; Polyancev, F. M.; Yakhvarov, D. G.; Sinyashin, O. G. Quantum chemical calculations of ^{31}P NMR chemical shifts: scopes and limitations. *Phys. Chem. Chem. Phys.* **2015**, *17*, 6976–6987.

(51) Internet Bond-energy Databank (pKa and BDE)—iBonD Home Page. 2020; http://ibond.nankai.edu.cn/.

(52) Denisov, E. A new semiempirical method of estimation of activity and bond dissociation energies of antioxidants. *Polymer Degradation and Stability* **1995**, *49*, 71–75.

(53) Liao, K.; Negretti, S.; Musaev, D. G.; Bacsa, J.; Davies, H. M. L. Site-selective and stereoselective functionalization of unactivated C–H bonds. *Nature* **2016**, *533*, 230–234.

(54) Liao, K.; Yang, Y.-F.; Li, Y.; Sanders, J. N.; Houk, K. N.; Musaev, D. G.; Davies, H. M. L. Design of catalysts for site-selective and enantioselective functionalization of non-activated primary C–H bonds. *Nature Chem* **2018**, *10*, 1048–1055.

(55) Liao, K.; Liu, W.; Niemeyer, Z. L.; Ren, Z.; Bacsa, J.; Musaev, D. G.; Sigman, M. S.; Davies, H. M. L. Site-Selective Carbene-Induced C–H Functionalization Catalyzed by Dirhodium Tetrakis(triarylcyclopropanecarboxylate) Complexes. *ACS Catalysis* **2018**, *8*, 678–682.

(56) Liao, K.; Pickel, T. C.; Boyarskikh, V.; Bacsa, J.; Musaev, D. G.; Davies, H. M. L. Site-selective and stereoselective functionalization of non-activated tertiary C–H bonds. *Nature* **2017**, *551*, 609–613.

(57) Hansen, J.; Autschbach, J.; Davies, H. M. L. Computational Study on the Selectivity of Donor/Acceptor-Substituted Rhodium Carbenoids. *The Journal of Organic Chemistry* **2009**, *74*, 6555–6563.

Graphical TOC Entry