A Fast, Low-Cost and Simple Method for Predicting Atomic/Inter-Atomic Properties by Combining a Low Dimensional Deep Learning Model with a Fragment Based Graph Convolutional Network

Peng Gao [1,2,†], Zonghang Liu [3,†], Jie Zhang [4,*], Jia-Ao Wang [5,6,*] and Graeme Henkelman [5,*]

1   School of Chemistry and Molecular Bioscience, University of Wollongong, Wollongong, NSW 2500, Australia
2   Molecular Horizons, University of Wollongong, Wollongong, NSW 2522, Australia
3   School of Chemistry and Chemical Engineering, University of Jinan, Jinan 250022, China
4   School of Chemical Engineering, East China University of Science and Technology, Shanghai 200237, China
5   Department of Chemistry and the Oden Institute for Computational Engineering and Sciences,
    The University of Texas at Austin, Austin, TX 78712, USA
6   AI for Science Group, Bytedance Inc., 151 W 42nd St, New York, NY 10036, USA
*   Correspondence: j.chang@ecust.edu.cn (J.Z.); wangjiaao0720@utexas.edu (J.-A.W.);
    henkelman@utexas.edu (G.H.)
†   These authors contributed equally to this work.

Abstract: Calculations with high accuracy for atomic and inter-atomic properties, such as nuclear magnetic resonance (NMR) spectroscopy and bond dissociation energies (BDEs) are valuable for pharmaceutical molecule structural analysis, drug exploration, and screening. It is important that these calculations should include relativistic effects, which are computationally expensive to treat. Non-relativistic calculations are less expensive but their results are less accurate. In this study, we present a computational framework for predicting atomic and inter-atomic properties by using machine-learning in a non-relativistic but accurate and computationally inexpensive framework. The accurate atomic and inter-atomic properties are obtained with a low dimensional deep neural network (DNN) embedded in a fragment-based graph convolutional neural network (F-GCN). The F-GCN acts as an atomic fingerprint generator that converts the atomistic local environments into data for the DNN, which improves the learning ability, resulting in accurate results as compared to experiments. Using this framework, the $^{13}C/^{1}H$ NMR chemical shifts of Nevirapine and phenol O–H BDEs are predicted to be in good agreement with experimental measurement.

Keywords: quantum mechanics; neural network; NMR; bond dissociation energy; machine-learning

1. Introduction

Accurate descriptions of inter-atomic information are increasingly important for chemical research [1–7]. For example, bond disassociation energies (BDEs) describe the energy difference that is caused by a change of bonding environment. Such metrics are relevant to reaction kinetics and can provide important chemical insights. In addition, the magnitudes of NMR chemical shifts can accurately reflect the atomic environment, and NMR measurements are important for structure determination. Accurate predictions of such metrics with affordable computation cost will be helpful for experimental researchers [8]. Over the past decades, density functional theory (DFT) has become a standard tool used by computational chemists [9–13]. Wave function based methods can be more accurate, but usually require a higher computational cost, and their use tends to be limited to small molecules. One way to improve the accuracy of DFT without significantly increasing the computational cost is to use data-driven methods.

Recently, artificial intelligence (AI) tools have been successfully applied in chemical and physical research, and through them, various kinds of expensive calculations, including atomic simulations, adsorptions, photocatalysis, etc, can be done with substantially reduced cost [14–30]. Moreover, with the development of graph convolutional neural networks (GCNs), accurate and efficient predictions of molecular and atomic properties have also become feasible [31–35]. Using molecular graphs, structural information can be systematically mapped to predicted properties within the framework of GCNs. However, using molecular graphs alone, it is difficult for GCNs to extract atomic and inter-atomic information; the reason behind this lies in the fact that the chemical environment cannot be effectively differentiated at the atomic level. To improve structure-function predictions, novel architectures can be exploited.

To successfully extract the local information in molecules, the fragment-based graph convolutional network (F-GCN) that is able to utilise multiple-level fragmentary graphs to accurately solve the chemical environment at the atomic level, will be helpful [36]. Unlike message-passing based approaches [37–39], the F-GCN doesn't require large amounts of molecular information. For more accurate predictions that require the assistance of extra chemical knowledge, however, our original architecture needs to be revised. In this study, we combine our original F-GCN with a low dimensional DNN. The added DNN is able to reduce the risk of over-fitting during incorporation of numerical descriptors; such a combination architecture is suitable for few-shot learning, as it is more focused on atomic or inter-atomic sites within molecular graphs. To further improve the prediction accuracy and efficiency of this tool, we will make revisions of the original architecture and include some selective descriptors that are numerically correlated with the target property. Specifically, there is a numerical correlation between the experimental NMR chemical shifts and DFT calculated isotropic shielding constants [9–11,33]; the experimental BDEs can be approximated from QM computations. First, we employed a low dimensional DNN to approximate target properties with respect to the added QM descriptors; then, a high dimensional F-GCN was applied to refine the results of the DNN within the multiple-level molecular graphs. We find that the errors of the DFT calculations due to complicated chemical environments can be mitigated with the assistance of our molecular graph based calculation.

2. Computational Details

2.1. Structure of the F-GCN

In our previous study, we reported a F-GCN architecture [36], which utilises multiple-level molecular fragments to describe atomic environments. The workflow is as follows: starting from a target site, fragmentary graphs at different levels are generated systematically; then all fragments are described by independent GCNs for information extraction. The scheme of our fragment generation is shown in Figure 1; the overall workflow can be seen in Figure 2.

Once the multiple-level fragments are generated, the GCN is able to utilise them for information extraction. Within the DGL library, the input molecular and fragmentary graphs were transformed into nodes and edges that correspond to atoms and bonds, respectively [40,41]. At the radial basis function (RBF) layer, the bonding data are recorded in a distance tensor. Additionally, continuous-filter convolution layers were applied to describe the atomic environment. The evolution of the ith atom at the $k+1$ layer can be expressed as:

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

where, ω^k represents the filter-generation, and $\circ$ indicates element-wise multiplication. To control the overall optimisation accuracy, a Gaussian function, g_k, is applied as

$$g_l(D_{ij}) = \exp(-\alpha(D_{ij} - \beta_l)^2) \tag{2}$$

in which, β_l is the magnitude of cutoff, and D_{ij} represents the connection between the ith and j^{th} atom. We applied a default value of $\alpha = 0.1$ in this study [31].
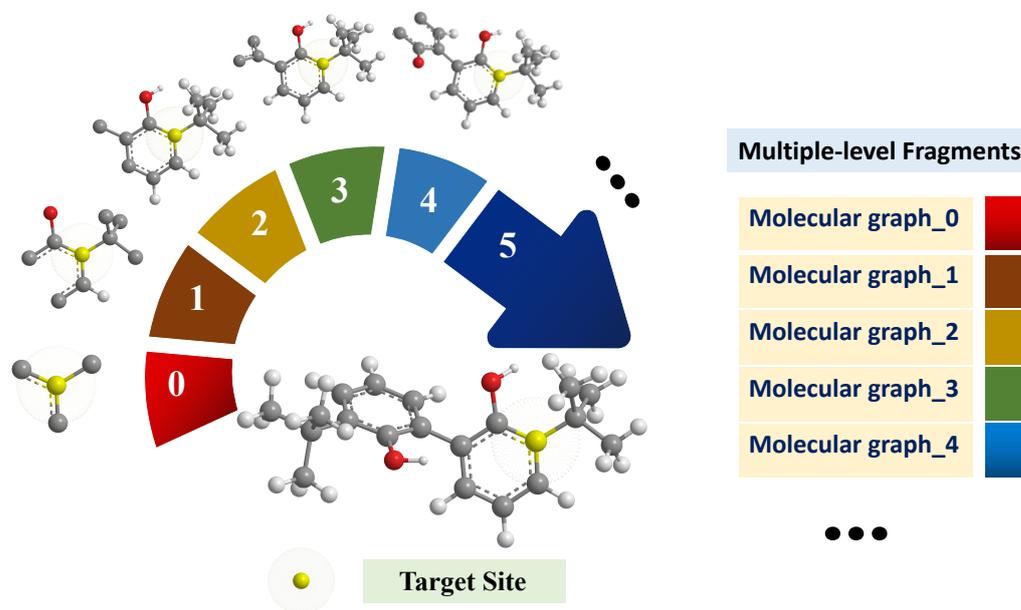

Figure 1. The generation of multiple-level molecular fragments within the framework of the F-GCN.

Moreover, for any chemical site, a loss function that takes an experimental value (Pr') as reference is applied to control the accuracy of the predicted property (Pr).

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

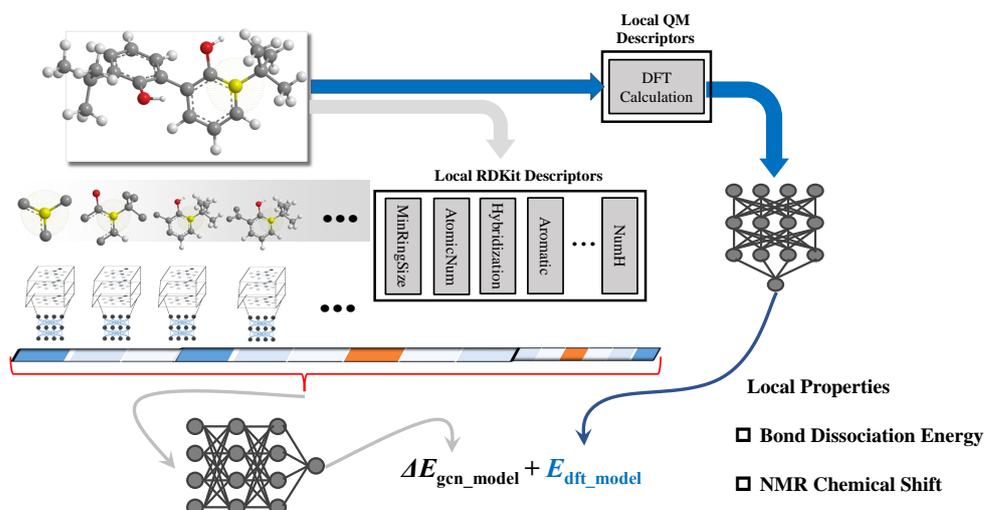To include extra chemical knowledge for more accurate predictions, modifications of this architecture are required.

Figure 2. The he workflow of the proposed F-GCN, designed for inter-atomic and atomic property prediction with inclusion of assisting descriptors.

2.2. Utilisation of QM Descriptors by a Low Dimensional DNN

To first approximate atomic/inter-atomic properties with respect to QM calculated descriptors, an independent DNN is employed, which is based on a polynomial fit to process the added descriptors, T_{QM},

$$T = a_0 + a_1 T_{QM} + a_2 (T_{QM})^2 + a_3 (T_{QM})^3 + a_4 (T_{QM})^4. \tag{4}$$

If the included descriptors that numerically correlate with the target values are directly input into the GCN, there is a risk of over-fitting, especially when using few-shot learning, as the QM calculated values cannot be fully utilised; this can be attributed to the original architecture of graph based neural networks, as the actual prediction accuracy is highly associated with the sampling probability P_{model}

$$\theta_{model} = \text{argmax}(\theta) \sum_{k=1} P_{model}(x^k, y^k; \theta). \tag{5}$$

With numerical corrections by the independent DNN, however, the QM calculated results can be approximately calibrated with experimental values, and the difference between these two values can be input into the next GCN stage. The accuracy of such a methodology had been well verified by previous study [33]. Then, the range of the target values' distribution becomes narrower (see the ^{1}H NMR chemical shifts in Figure 3) where the risk of over-fitting is reduced, as shown in Figure 4, and the performance of the GCN is improved.

In this study, we focused on few-shot learning cases, and made a trial to combine QM calculations with molecular graphs for experimental NMR chemical shifts and BDEs predictions. Within molecules, for the ith nucleus, v_i, the NMR frequency can be expressed as

$$v_i = (\gamma_i/2\pi)B_1 = (\gamma_i/2\pi)B_0(1 - \sigma_i) \tag{6}$$

where γ_i represents the gyromagnetic ratio of the ith nucleus, which is approximately a constant value, and σ_i is the isotropic shielding constant, which is calculated with the DFT/GIAO approach [12]. The accuracy of the QM calculations depends on many factors [11] including B_1, the strength of the induced field which is proportional to the strength of the uniformed external field along z-axis, and the value of the chemical shift, δ_i (in ppm), which can be obtained as

$$\delta_i = 10^6(v_i - v_0)/v_0 \tag{7}$$

where v_0 is the resonance frequency of the referenced nucleus [42–44].

There is a quasi-linear relationship between the experimental chemical shifts (δ_i) and the QM calculated isotropic shielding constants (σ_i) [11,45]. In previous studies, scaling factors have been applied to approximate NMR chemical shifts [11,45,46], using

$$\delta_i = -(\text{intercept} - \sigma_i)/\text{slope}. \tag{8}$$

It is worth noting, however, that for complex bonding environments, the approximate chemical shifts may deviate nonlinearly from the experimental values [11,12]. Within the F-GCN framework the structural information extracted can efficiently correct the approximate results of the DNN. First, we tested this refined F-GCN on ^{13}C and ^{1}H NMR chemical shift predictions with inclusion of the DFT calculated isotropic shielding constants [11]. To obtain such a descriptor, we first conducted geometry optimisation of the molecules to locate the minimum of the potential energy surface. The isotropic shielding constants were calculated with the GIAO approach; in this step, the SMD implicit solvent model [47] was adopted. In this study, we employed the M062X/6-31+G(d,p) level of theory for geometry optimisation, and mPW1PW91/6-311+G(2d,p) for the NMR GIAO calculations. The geometries with lowest energy were adopted. All the calculations were performed within the Gaussian 09 software package [48].

We also applied the proposed F-GCN for the experimental BDEs predictions with inclusion of DFT calculated bonding energies for C−H, O−H and C−C bonds [49]; and in this part, all the DFT calculations were performed at the M062X/def2-TZVP level, with the formula AB ⟶ A· + B·.
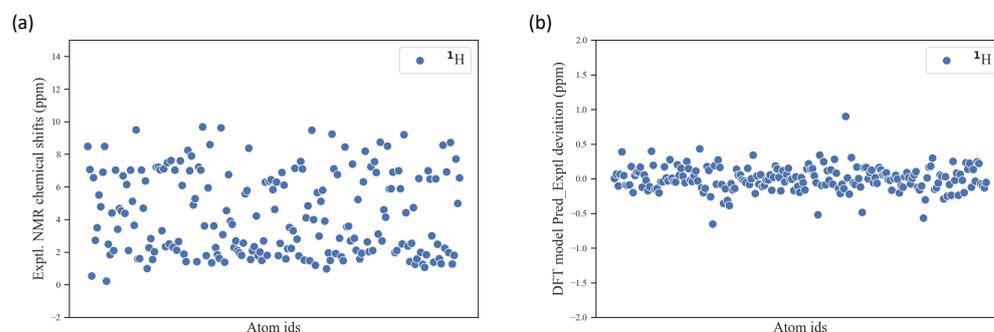
(a) (b)

Figure 3. (**a**) Distribution of experimental NMR ^{13}C chemical shifts. (**b**) Distribution of the difference between the predicted and experimental ^{1}H NMR chemical shifts.
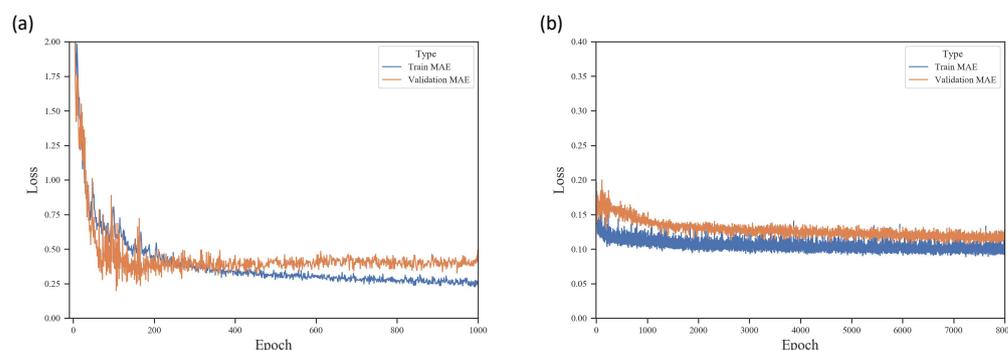
(a) (b)

Figure 4. Convergence plot of the (**a**) F-GCN model on experimental NMR ^{1}H chemical shifts predictions and (**b**) difference NMR ^{1}H chemical shifts predictions.

3. Results and Discussion

3.1. Performance of the QM Augmented F-GCN in NMR Chemical Shift Predictions

The prediction results for ^{13}C and ^{1}H NMR chemical shifts are presented in Figure 5; for the original data set, the ratio of the number of training to test molecules is set to 9:1. We can see that the architecture of the F-GCN is able to calibrate the DFT calculation results with respect to experimental values via inclusion of a low dimensional DNN network to numerically process the DFT calculated information. The reason behind this lies in the fact that the magnitudes of the chemical shifts are largely determined by the atomic environment; for complicated bonding environments, the heavily asymmetric distribution of electron density may make the DFT calculated isotropic shielding constants deviate non-linearly from the experimental NMR chemical shifts, leading to numerical errors. Such errors cannot be simply overcome by merely improving the level of QM theory. This combination F-GCN has natural advantages of accurately describing the chemical environment at the atomic level, and thus can effectively correct the DFT errors. At the same time, different from other models that merely rely on RDKit generated descriptors to extract structural information [33], the architecture of F-GCN is more efficient, due to the fact that it directly uses multiple-level fragments for information extraction. Additionally, with the inclusion of DFT calculated isotropic shielding constants that are processed by the mentioned DNN model, the augmented F-GCN becomes suitable for few-shot learning with a lower risk of over-fitting.
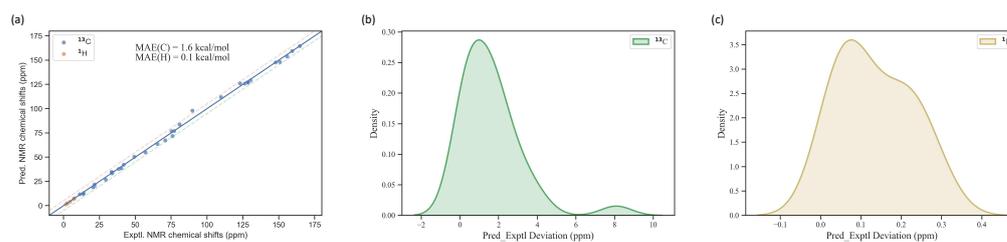
Figure 5. (**a**) Comparison between the predicted and experimental NMR chemical shifts for 459 ^{13}C and 213 ^{1}H chemical shifts contained in our original data set. (**b**) Distributions of errors between the predicted and experimental ^{13}C NMR chemical shifts. (**c**) Distributions of errors between the predicted and experimental ^{1}H NMR chemical shifts.

3.2. Performance of the QM Augmented F-GCN in BDE Predictions

To improve the performance of F-GCN on experimental inter-atomic properties predictions, inclusion of DFT calculated bonding energy descriptors may be helpful [33]. Thus, we revised the architecture of F-GCN to include both molecular graphs and QM calculated bond dissociation energy, and tested it for predictions at experimental level. The performance is summarised in Figure 6. We can clearly see that the prediction error is largely reduced, compared to that produced by the F-GCN model [36], due to the fact that F-GCN can describe the inter-atomic environment within molecular graphs through the calibration of DFT calculated BDEs. In summary, we can reasonably conclude that the inclusion of QM descriptors is helpful to enhance the overall performance of graph convolutional network. However, it is still worth noting that running a large number of DFT calculations may be computationally expensive, to handle this, more general quantitative structure-property relationship (QSPR) protocols still remain to be explored. Additionally, in the cases of complicated bonding structures, the calculated bonding energy may significantly deviate from the experimental value due to the deficiency of QM calculations, thus the performance of graph based approaches may also be negatively influenced.
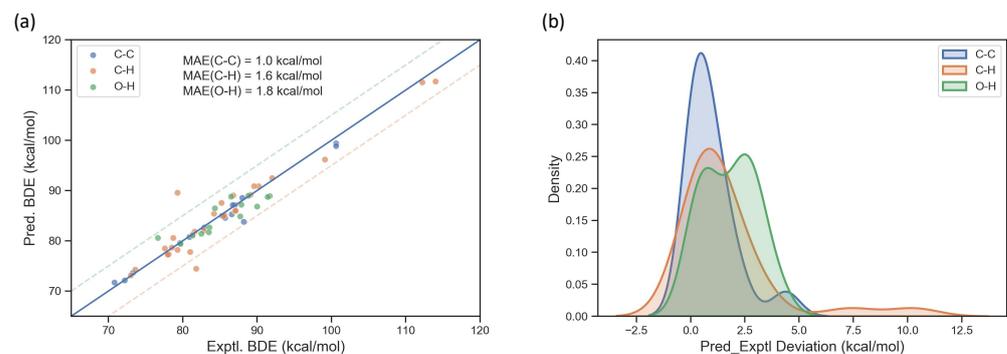

Figure 6. (**a**) Comparison between the predicted and experimental BDEs for the 217 C−C, 375 C−H and 141 O−H bonds contained in our test set. (**b**) Distribution of errors between the predicted and experimental BDEs.

It is understandable that the performance of the F-GCN in atomic/inter-atomic properties predictions is largely dependent on molecular coverage of the original data sets, as there always exists an imperfect mapping between the accuracy of DFT calculations and solution degree of molecular graphs; the prediction precision is positively correlated with the actual consistency between these two items. That is to say, to systematically overcome the hurdles of graph based approaches and make the proposed architecture more flexible for specific applications, in one aspect, it is helpful to enlarge the training data set selectively to enhance molecular diversity, thus their differentiation capability among various molecular graphs can be subsequently enhanced. In another aspect, identification of correlated yet affordable QM descriptors, is also of great significance, the ones that contain important

chemical knowledge can positively influence the prediction accuracy within the framework of GCN.

3.3. Nevirapine Structure Elucidation by the QM Augmented F-GCN Architecture

Nevirapine is an important inhibitor for HIV reverse transcriptase, its structure is shown in Figure 7. Unfortunately, there exist multiple possible structures, and for experimental researchers, accurate structure elucidation still remains to be a challenging goal. In addition, due to the lower solubility of this compound, NMR spectra are difficult to obtain. Thus, a good consistence in NMR chemical shifts between computational predictions and experimental measurement will be of great significance for studies of this compound. We applied our proposed F-GCN for ^{13}C/^{1}H chemical shifts predictions for this complex structure. The results are summarised in Tables 1 and 2; we can clearly see that the predicted results match well with the experimental values, indicating the applicability of this QM augmented F-GCN in challenging structure assignments.
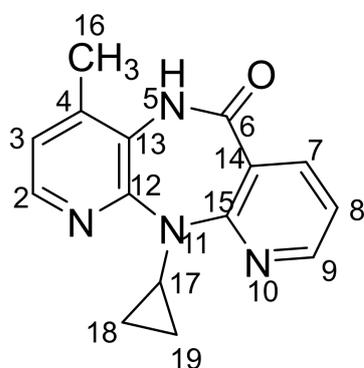

Figure 7. The structure of Nevirapine.

Table 1. Predicted and experimental ^{13}C NMR chemical shifts (in ppm) of Nevirapine.

Position [a]	Exptl.	Pred. [b]	Error
2	140.36	143.33	2.97
3	120.35	122.41	2.06
4	139.52	137.42	2.10
6	169.06	165.45	3.61
7	144.47	138.38	6.09
8	118.99	119.93	0.94
9	152.15	155.36	3.21
12	154.17	152.68	1.49
13	124.97	126.29	1.32
14	122.13	120.39	1.74
15	160.73	159.95	0.78
16	17.86	20.39	2.53
17	29.65	32.49	2.84
18	8.88	9.07	0.19
19	9.15	9.45	0.30

[a] Positions for the carbon atoms of interest. [b] The predicted ^{13}C NMR chemical shifts via the trained QM augmented F-GCN architecture.

Table 2. Predicted and experimental ^{1}H NMR chemical shifts (in ppm) of Nevirapine.

Position [a]	Exptl.	Pred. [b]	Error
2	8.08	7.80	0.28
3	7.07	6.71	0.36
7	8.02	7.93	0.09
8	7.20	6.77	0.43
9	8.51	8.25	0.26
16	2.34	2.25	0.09
17	3.62	3.60	0.02
18	0.35	0.51	0.16
19	0.88	0.87	0.01

[a] Positions for the hydrogen atoms of interest. [b] The predicted ^{1}H NMR chemical shifts via the trained QM augmented F-GCN architecture.

3.4. Calculations of Phenol O−H BDEs by the QM Augmented F-GCN Architecture

Phenol inhibitors can efficiently retard the oxidation of polymers; for this class of compounds, the phenol O−H bonds are usually first attacked by the peroxyl based radicals. That is to say, the O−H BDE can serve as a demonstrative metric to characterise the performance of phenol inhibitors [50]. Currently, due to the operational complexity and time required for experimental measurements that are based on kinetic analysis, the number and types of referable O−H BDE values are limited and remain to be enlarged. Within the framework of F-GCN, by inclusion of QM descriptors, accurate calculations of phenol O−H BDEs can be realised. In Figure 8, we present the experimental and predicted O−H BDEs for a series of phenol compounds. It is notable that with F-GCN, the prediction results were all at experimental level, demonstrating its promising prospect for this kind of predictions. In addition, the DFT calculation results were found to be largely calibrated with the assistance of molecular graphs.
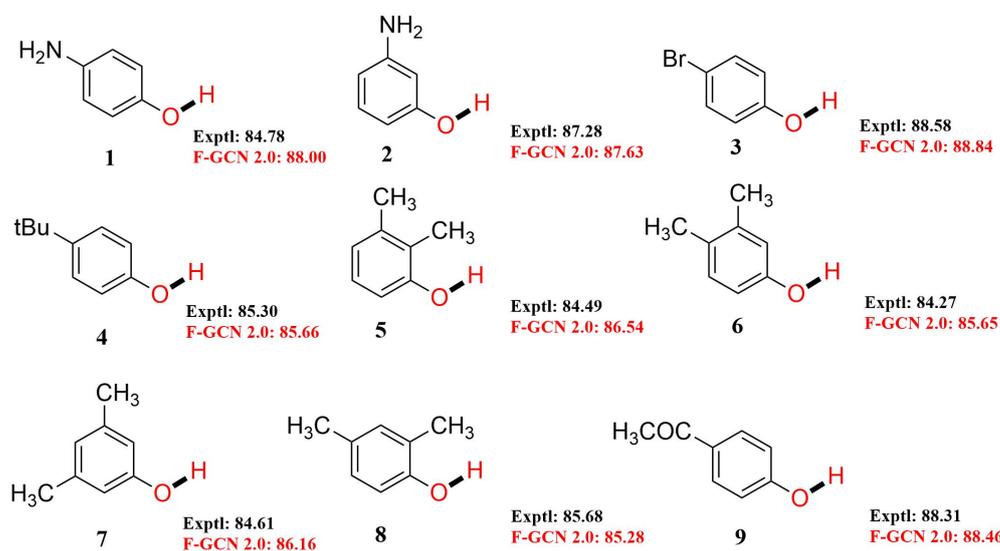

Figure 8. Comparison between the experimental and predicted O−H BDEs (in kcal/mol) of the selected phenol compounds.

4. Conclusions

To sum up, through modification of the original architecture of F-GCN, the prediction performance in both atomic and inter-atomic properties are enhanced. The inclusion of QM descriptors within a separated neural network calibrate the prediction results, making it suitable for few-shot learning cases; the proposed F-GCN is shown to be more powerful for chemical environment description at the atomic level. The prediction results of NMR

chemical shifts and BDEs are comparable to experimental measurement. Moreover, the proposed architecture is flexible to include other useful descriptors, thus can be applicable for various kinds of challenging structural assignments. The success of F-GCN indicates a promising direction for incorporating advanced AI technologies into physical and chemical research; however, purely QM calculations are time expensive, and cannot be conducted in large scales, thus reasonable identification of alternative yet affordable calculations will be expected. In the future, we expect that more scientific insights can be provided with assistance of novel yet functional data-driven approaches.

References

1. Gani, T.Z.H.; Kulik, H.J. Understanding and Breaking Scaling Relations in Single-Site Catalysis: Methane to Methanol Conversion by FeIV=O. *ACS Catal.* **2018**, *8*, 975–986. [CrossRef]
2. 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. [CrossRef]
3. 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. [CrossRef]
4. 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. *J. Phys. Chem. Lett.* **2011**, *2*, 2846–2852. [CrossRef]
5. 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. *Eur. J. Med. Chem.* **2012**, *56*, 48–55. [CrossRef]
6. Blanksby, S.J.; Ellison, G.B. Bond Dissociation Energies of Organic Molecules. *Acc. Chem. Res.* **2003**, *36*, 255–263. [CrossRef] [PubMed]
7. Hartwig, J.F. Catalyst-Controlled Site-Selective Bond Activation. *Acc. Chem. Res.* **2017**, *50*, 549–555. [CrossRef] [PubMed]
8. Yao, K.; Herr, J.E.; Brown, S.N.; Parkhill, J. Intrinsic Bond Energies from a Bonds-in-Molecules Neural Network. *J. Phys. Chem. Lett.* **2017**, *8*, 2689–2694. [CrossRef] [PubMed]
9. Gao, P.; Wang, X.; Yu, H. Towards an Accurate Prediction of Nitrogen Chemical Shifts by Density Functional Theory and Gauge-Including Atomic Orbital. *Adv. Theory Simul.* **2019**, *2*, 1800148. [CrossRef]
10. 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. [CrossRef]
11. Lodewyk, M.W.; Siebert, M.R.; Tantillo, D.J. Computational Prediction of ^{1}H and ^{13}C Chemical Shifts: A Useful Tool for Natural Product, Mechanistic, and Synthetic Organic Chemistry. *Chem. Rev.* **2012**, *112*, 1839–1862. [CrossRef]
12. Ditchfield, R. Self-consistent perturbation theory of diamagnetism. *Mol. Phys.* **1974**, *27*, 789–807. [CrossRef]
13. 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. *Int. J. Quantum Chem.* **2020**, *121*, e26482. [CrossRef]
14. Behler, J. Perspective: Machine learning potentials for atomistic simulations. *J. Chem. Phys.* **2016**, *145*, 170901. [CrossRef]
15. Behler, J. First Principles Neural Network Potentials for Reactive Simulations of Large Molecular and Condensed Systems. *Angew. Chem. Int. Ed.* **2017**, *56*, 12828–12840. [CrossRef] [PubMed]

16. 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 Cent. Sci.* **2019**, *5*, 755–767. [CrossRef]

17. Botu, V.; Batra, R.; Chapman, J.; Ramprasad, R. Machine Learning Force Fields: Construction, Validation, and Outlook. *J. Phys. Chem. C* **2017**, *121*, 511–522. [CrossRef]

18. Meldgaard, S.A.; Kolsbjerg, E.L.; Hammer, B. Machine learning enhanced global optimization by clustering local environments to enable bundled atomic energies. *J. Chem. Phys.* **2018**, *149*, 134104. [CrossRef]

19. 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. [CrossRef] [PubMed]

20. Sørensen, K.H.; Jørgensen, M.S.; Bruix, A.; Hammer, B. Accelerating atomic structure search with cluster regularization. *J. Chem. Phys.* **2018**, *148*, 241734. [CrossRef]

21. Wexler, R.B.; Martirez, J.M.P.; Rappe, A.M. Chemical Pressure-Driven Enhancement of the Hydrogen Evolving Activity of Ni2P from Nonmetal Surface Doping Interpreted via Machine Learning. *J. Am. Chem. Soc.* **2018**, *140*, 4678–4683. [CrossRef] [PubMed]

22. 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. *J. Am. Chem. Soc.* **2018**, *140*, 9844–9853. [CrossRef] [PubMed]

23. 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. *J. Am. Chem. Soc.* **2018**, *140*, 17508–17514. [CrossRef] [PubMed]

24. Rupp, M.; Ramakrishnan, R.; von Lilienfeld, O.A. Machine Learning for Quantum Mechanical Properties of Atoms in Molecules. *J. Phys. Chem. Lett.* **2015**, *6*, 3309–3313. [CrossRef]

25. 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. *J. Am. Chem. Soc.* **2019**, *141*, 9063–9071. [CrossRef]

26. 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. [CrossRef]

27. Mater, A.C.; Coote, M.L. Deep Learning in Chemistry. *J. Chem. Inf. Model.* **2019**, *59*, 2545–2559. [CrossRef]

28. 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. *J. Chem. Theory Comput.* **2017**, *13*, 5255–5264. [CrossRef]

29. Kendall, A.; Gal, Y. What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 30.

30. De Sousa Ribeiro, F.; Calivá, F.; Swainson, M.; Gudmundsson, K.; Leontidis, G.; Kollias, S. Deep Bayesian Self-Training. *Neural Comput. Appl.* **2020**, *32*, 4275–4291. [CrossRef]

31. Schütt, K.T.; Sauceda, H.E.; Kindermans, P.J.; Tkatchenko, A.; Müller, K.R. SchNet—A deep learning architecture for molecules and materials. *J. Chem. Phys.* **2018**, *148*, 241722. [CrossRef]

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

33. 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. *J. Chem. Inf. Model.* **2020**, *60*, 3746–3754. [CrossRef]

34. 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. [CrossRef]

35. 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. *J. Phys. Chem. Lett.* **2020**, *0*, 9812–9818. [CrossRef]

36. Gao, P.; Zhang, J.; Qiu, H.; Zhao, S. A general QSPR protocol for the prediction of atomic/inter-atomic properties: A fragment based graph convolutional neural network (F-GCN). *Phys. Chem. Chem. Phys.* **2021**, *23*, 13242–13249. [CrossRef]

37. 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. [CrossRef]

38. Kwon, Y.; Lee, D.; Choi, Y.S.; Kang, M.; Kang, S. Neural Message Passing for NMR Chemical Shift Prediction. *J. Chem. Inf. Model.* **2020**, *60*, 2024–2030. [CrossRef]

39. Gerrard, W.; Bratholm, L.A.; Packer, M.J.; Mulholland, A.J.; Glowacki, D.R.; Butts, C.P. IMPRESSION—Prediction of NMR parameters for 3-dimensional chemical structures using machine learning with near quantum chemical accuracy. *Chem. Sci.* **2020**, *11*, 508–515. [CrossRef]

40. Wang, M.; Zheng, D.; Ye, Z.; Gan, Q.; Li, M.; Song, X.; Zhou, J.; Ma, C.; Yu, L.; Gai, Y.; et al. Deep Graph Library: A Graph-Centric, Highly-Performant Package for Graph Neural Networks. *arXiv* **2019**, arXiv:cs.LG/1909.01315.

41. Chen, G.; Chen, P.; Hsieh, C.Y.; Lee, C.K.; Liao, B.; Liao, R.; Liu, W.; Qiu, J.; Sun, Q.; Tang, J.; et al. Alchemy: A Quantum Chemistry Dataset for Benchmarking AI Models. *arXiv* **2019**, arXiv:1906.09427.

42. Pople, J.A.; Bernstein, H.J.; Schneider, W.G. *High-Resolution Nuclear Magnetic Resonance*; McGraw-Hill: New York, NY, USA, 1959.

43. Becker, E. *High Resolution NMR: Theory and Chemical Applications*; Elsevier Science: Amsterdam, The Netherlands, 1999.

44. Slichter, C. *Principles of Magnetic Resonance*; Springer Series in Solid-State Sciences; Springer: Berlin/Heidelberg, Germany, 2013.

45. Lodewyk, M.W.; Soldi, C.; Jones, P.B.; Olmstead, M.M.; Rita, J.; Shaw, J.T.; Tantillo, D.J. The Correct Structure of Aquatolide—Experimental Validation of a Theoretically-Predicted Structural Revision. *J. Am. Chem. Soc.* **2012**, *134*, 18550–18553. [CrossRef]

46. Xin, D.; Sader, C.A.; Chaudhary, O.; Jones, P.J.; Wagner, K.; Tautermann, C.S.; Yang, Z.; Busacca, C.A.; Saraceno, R.A.; Fandrick, K.R.; et al. Development of a ^{13}C NMR Chemical Shift Prediction Procedure Using B3LYP/cc-pVDZ and Empirically Derived Systematic Error Correction Terms: A Computational Small Molecule Structure Elucidation Method. *J. Org. Chem.* **2017**, *82*, 5135–5145. [CrossRef]

47. 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. [CrossRef]

48. Frisch, M.J.; Trucks, G.W.; Schlegel, H.B.; Scuseria, G.E.; Robb, M.A.; Cheeseman, J.R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G.A.; et al. *Gaussian 09 Revision E.01.*; Gaussian Inc.: Wallingford, CT, USA, 2009.

49. Internet Bond-Energy Databank (pKa and BDE)—iBonD Home Page. 2020. Available online: http://ibond.nankai.edu.cn/ (accessed on 22 October 2022).

50. Denisov, E. A new semiempirical method of estimation of activity and bond dissociation energies of antioxidants. *Polym. Degrad. Stab.* **1995**, *49*, 71–75. [CrossRef]