



Cite this: *Digital Discovery*, 2024, 3, 1714

# Universal neural network potentials as descriptors: towards scalable chemical property prediction using quantum and classical computers

Tomoya Shiota,<sup>a</sup> Kenji Ishihara<sup>b</sup> and Wataru Mizukami<sup>a</sup>

Accurate prediction of diverse chemical properties is crucial for advancing molecular design and materials discovery. Here we present a versatile approach that uses the intermediate information of a universal neural network potential as a general-purpose descriptor for chemical property prediction. Our method is based on the insight that by training a sophisticated neural network architecture for universal force fields, it learns transferable representations of atomic environments. We show that transfer learning with graph neural network potentials such as M3GNet and MACE achieves accuracy comparable to state-of-the-art methods for predicting the NMR chemical shifts by using quantum machine learning as well as a standard classical regression model, despite the compactness of its descriptors. In particular, the MACE descriptor demonstrates the highest accuracy to date on the <sup>13</sup>C NMR chemical shift benchmarks for drug molecules. This work provides an efficient way to accurately predict properties, potentially accelerating the discovery of new molecules and materials.

Received 9th April 2024

Accepted 12th July 2024

DOI: 10.1039/d4dd00098f

rsc.li/digitaldiscovery

## 1 Introduction

As evidenced by the enumeration of 166.4B possible organic molecules containing up to 17 heavy elements, such as C, N, O, S, and halogens (excluding hydrogen), the expansion of the chemical space is astronomical with the increase in types and numbers of elements.<sup>1,2</sup> This vast landscape has given rise to multidisciplinary approaches to combining experimental and computational chemistry for the discovery of new chemical substances and materials in a wide range of fields, including material, catalysis, and drug design.<sup>1–6</sup> Although quantum chemistry and first-principles calculations offer accurate descriptions of chemical substances, their high computational demands make an exhaustive exploration of the chemical space impractical.<sup>6–12</sup> However, machine- and deep-learning techniques are overcoming these limitations to enable a more extensive exploration.<sup>4,6,10,13–27</sup>

With machine learning, physics-inspired descriptors that characterize the chemical space have been developed and serve as the cornerstone for building efficient and highly accurate models.<sup>21,23,28–40</sup> Smooth overlap of atomic positions (SOAP),<sup>14,18,21,28,31,32,41</sup> Faber–Christensen–Huang–Lilienfeld (FCHL),<sup>29,30,33,34,38</sup> and similar descriptors offer atom-level descriptions within molecular or material environments based

on physical insights and are effective in regressing chemical quantities, such as interatomic potentials (IAP) and nuclear magnetic resonance (NMR) chemical shifts.<sup>11,14,15,21,23,24,28–38,41–43</sup> Notably, IAPs built using descriptors and Gaussian process regression (GPR)<sup>14</sup> have been termed Gaussian approximation potentials (GAP) and have found success in the exploration of the chemical space of molecules and materials.<sup>14,21,37</sup> Both kernel ridge regression (KRR) and GPR have been employed to improve the accuracy of NMR chemical shift prediction.<sup>29,30,41–44</sup> However, the dimensionality of the descriptors becomes a barrier to generalization and high accuracy as the molecular or material composition becomes more diverse owing to the addition of different types of elements.<sup>19,35,39,45</sup>

Recently, deep-learning models based on graph neural networks (GNNs) have been proposed to describe chemical spaces using graph representations.<sup>9,10,17–20,24,25,45–64</sup> In most GNN-based IAPs, atoms within a molecular or material environment are represented as nodes, and their local connectivity as edges in a graph. The graph is then convolved to embed atom-specific information within each node, and further processed using multilayer perceptrons (MLP) to predict target observables. In molecular and materials simulation and modeling, the consideration of symmetry is extremely important. It is desirable for GNNs to be invariant or equivariant to symmetry operations such as translation, rotation, and reflection for the models to make physically meaningful predictions. GNNs that possess these properties are referred to as invariant GNNs or equivariant GNNs. The universal GNN-based IAPs proposed thus far have been designed to satisfy these symmetries. Recently, E(3) or SE(3) equivariant GNN-based IAPs (e.g.,

<sup>a</sup>Graduate School of Engineering Science, Osaka University, 1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan. E-mail: shiota.tomoya.ss@gmail.com; mizukami.wataru.qiqb@osaka-u.ac.jp

<sup>b</sup>Center for Quantum Information and Quantum Biology, Osaka University, 1-2 Machikaneyama, Toyonaka 560-8531, Japan



Allegro,<sup>61</sup> GNoME,<sup>65</sup> MACE<sup>62–64</sup>) have demonstrated superior performance compared to E(3) invariant GNN-based IAPs (e.g., MEGNet,<sup>52</sup> M3GNet<sup>10</sup>).<sup>66,67</sup>

Similarly, GNN-based models have been developed to predict NMR chemical shifts.<sup>46,47,49,50,59,68</sup> DFT-level calculations of NMR chemical shifts for  $^1\text{H}$  and  $^{13}\text{C}$  have demonstrated the ability to predict within a target accuracy range of 1–2% relative to the possible ranges of approximately 10 ppm and 200 ppm, respectively.<sup>69,70</sup> Therefore, the uncertainty in machine learning models using DFT-level datasets is this level of precision, with the target accuracy of 0.2 ppm for  $^1\text{H}$  and 2 ppm for  $^{13}\text{C}$ .<sup>30</sup> For example, Yanfei Guan *et al.* achieved the target accuracy of 0.16 ppm for  $^1\text{H}$  and 1.26 ppm for  $^{13}\text{C}$  by training the SchNet architecture<sup>51</sup> on molecular NMR chemical shifts (CASCADE).<sup>46</sup>

However, the scalability remains an issue due to the increasing optimization costs of GNN and MLP parameters when the size of datasets increase. Han *et al.* addressed this issue by constraining the nodes in a GNN to heavy elements only, thereby rendering the construction of scalable GNN-based NMR chemical shift models feasible while achieving a state-of-the-art prediction accuracy comparable to that of CASCADE.<sup>68</sup> Furthermore, NMR chemical shifts of various nuclei beyond hydrogen and carbon have become crucial for understanding systems involving a wide range of elements, such as proteins and solids.<sup>71–77</sup> Consequently, efforts are being made to develop machine learning models for NMR chemical shifts of nuclei such as  $^{15}\text{N}$ ,  $^{17}\text{O}$ , and  $^{19}\text{F}$ .<sup>73–76</sup> These elements exhibit wide chemical shift ranges, with about 600, 2500, 500 ppm for  $^{15}\text{N}$ ,  $^{17}\text{O}$ , and  $^{19}\text{F}$ , respectively. The target accuracy for these nuclei is set at 25 ppm for  $^{15}\text{N}$  and 5 ppm for  $^{19}\text{F}$  as well as  $^1\text{H}$  and  $^{13}\text{C}$ .<sup>71–77</sup>

Notably, both descriptor-based and GNN-based methods face challenges. The former faces increased learning costs as the composition becomes more complex, and the latter faces increasing parameter optimization costs with larger training datasets. To address these issues simultaneously, we focused on the potential utility of the outputs from pre-trained GNN-based IAPs as descriptors. We considered these outputs GNN transfer learning (GNN-TL) descriptors and built machine-learning models for predicting chemical properties. Note that there are

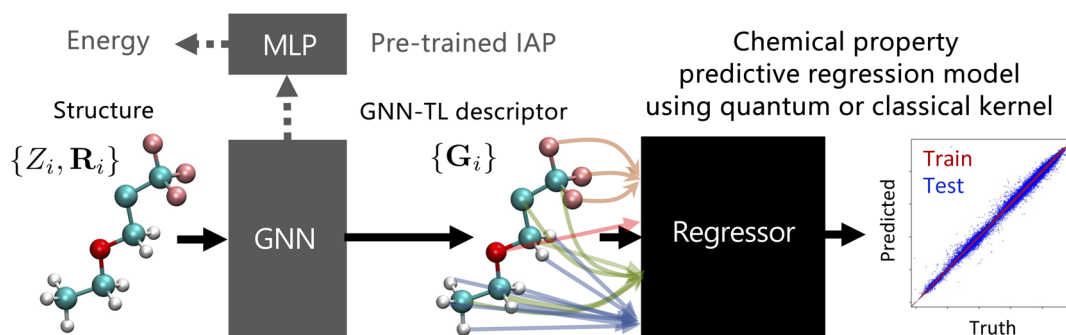
existing studies attempting to apply pre-trained GNN potentials to other tasks, particularly to generative modeling.<sup>78–81</sup>

The remainder of this paper is organized as follows. Section 2 details the GNN-TL descriptor and the kernel method, implemented on both classical and quantum computers, for predicting NMR chemical shifts of  $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ ,  $^{17}\text{O}$ , and  $^{19}\text{F}$ . Section 3 presents the performance of our developed machine learning models. Section 4 discusses the benefits and applications of the GNN-TL descriptor. Finally, Section 5 concludes the paper.

## 2 Method: transfer learning using pre-trained graph neural network

In this section, we discuss the transfer learning of a pre-trained GNN-based IAP. This approach integrates the outputs from the GNN layer of the IAP as shown in Fig. 1. The architecture of a GNN-based IAP can be broadly segmented into a GNN layer and an MLP layer (gray area of Fig. 1). For the E(3) invariant GNN-based IAP, we opted for two backbones: a MEGNet pre-trained on the QM9 dataset<sup>82</sup> and a M3GNet trained on the MPF.2021.2.8 dataset, which encompasses compounds covering all 89 elements from the Materials Project.<sup>10</sup> The parameters of the GNN layer in the M3GNet IAP were optimized to predict system energy, forces, and stress tensors. Additionally, we incorporated the E(3) equivariant GNN-based IAPs, namely MACE,<sup>62–64</sup> into our study. We employed two types of pre-trained MACE IAPs: one trained on a larger dataset named MPTrj<sup>83</sup> from Materials Project, referred to as the MACE-MP0 model,<sup>64</sup> and another trained on an organic molecule dataset covering 10 types of elements including SPICE<sup>84</sup> and QMug,<sup>85</sup> termed the MACE-OFF23 model.<sup>63</sup> Each model has variations in parameter size, and for this study, we utilized the “small” and “large” versions.<sup>63,64</sup>

When fed with the atomic coordinates of a molecule with  $N$  atoms, denoted by  $\{Z_i, \mathbf{R}_i\}$ , where  $Z_i$  represents the atomic number indicating the type of each atom, and  $\mathbf{R}_i$  is the three-dimensional position vector of the  $i$ th atom, the GNN layer generates a set of vectors,  $\{\mathbf{G}_i\}$ , which mirrors the environment



**Fig. 1** Schematic diagram of our proposed graph neural network transfer learning for predicting chemical properties. The black arrows depict the flow of our transfer learning process. The gray area is a pre-trained IAP (NNP) designed for predicting the energy of the system and composed of a GNN and an MLP. The initial step in our learning procedure involves obtaining the pre-trained GNN block output a set of vectors,  $\{\mathbf{G}_i\}$ , using the atomic coordinates of a molecule with  $N$  atoms,  $\{Z_i, \mathbf{R}_i\}$  as input. Subsequently, we construct a regression model to predict the chemical properties e.g. NMR shielding constants, using this GNN output  $\{\mathbf{G}_i\}$  as a descriptor.



of the  $i$ th atom in the molecule. This is referred to as the GNN-TL descriptor. The GNN layer for both MEGNet and M3GNet outputs GNN-TL descriptors with dimensions of 32 and 64 per atom, respectively. On the other hand, MACE is a GNN architecture that predicts energy in the form of atomic cluster expansion. As in ref. 86, only the output of the 1st layer of the GNN layer, corresponding to the one-body term of the many-body expansion, is used as the GNN-TL descriptor. The dimensions of this GNN-TL descriptor are 128, 256, 96, and 224 per atom for MACE-MP0-small, MACE-MP0-large, MACE-OFF23-small, and MACE-OFF23-large, respectively.

Using GNN-TL descriptors as input, a regression model was constructed to predict NMR chemical shielding constants. For the regressor, one can choose methodologies, such as GPR, KRR, or feed-forward neural network (NNs), which are contingent on the specific task. In this study, to ensure a maximally fair comparison with other descriptor-based techniques, we adopted KRR.

KRR combines the merits of ridge regression, which offers regularization to mitigate overfitting, with the kernel method, facilitating nonlinear regression. In kernel methods, the data – in the context of our study, the GNN-TL descriptors – are mapped into a high-dimensional feature space through a non-linear kernel function. The Laplacian and Gaussian kernels were applied:

$$k(\mathbf{G}_i, \mathbf{G}_j) = \exp(-\gamma \|\mathbf{G}_i - \mathbf{G}_j\|_p^p), \quad (1)$$

where  $\gamma$  is the hyperparameter of the kernel and  $p$  is the norm parameter that differentiates the type of kernel:  $p = 1$  for the Laplacian kernel and  $p = 2$  for the Gaussian kernel. In KRR, the predicted value  $\hat{\sigma}_t$  for the target chemical property of the target atom is derived from the GNN-TL descriptor  $\mathbf{G}_t$  as follows:

$$\hat{\sigma}_t(\mathbf{G}_t) = \sum_i^N \alpha_i k(\mathbf{G}_i, \mathbf{G}_t) \quad (2)$$

here,  $\alpha_i$  represents the  $i$ th element of the regression coefficient vector,  $\alpha$ , of size  $N$ . The regression coefficients are determined by solving a ridge-regularized least-squares problem, which can be reduced to:

$$\alpha = (\mathbf{K} + \lambda \mathbf{I})^{-1} \sigma \quad (3)$$

where  $\mathbf{I}$  denotes the identity matrix,  $\sigma$  denotes the chemical properties of each  $N$  training data samples, and  $\lambda$  denotes the regularization parameter. The matrix  $\mathbf{K}$ , is a kernel matrix, with elements given by  $k(\mathbf{G}_i, \mathbf{G}_j)$ .

All computations related to the KRR were executed using Scikit-learn v.1.2.2,<sup>87</sup> and the hyperparameters of each model were tuned using Optuna v.2.10.<sup>88</sup> For dataset sizes of up to 50K items, we conducted hyperparameter optimization for 100 iterations with ten-fold cross-validation, while for those at 100K, we limited the optimization to 10 iterations.

The quantum-kernel method leverages quantum computers to compute kernels,<sup>16,89,90</sup> which is achieved by embedding feature vectors generated by classical computers into quantum states. This method calculates the inner product of these

quantum states to derive the desired kernels. Embedding feature vectors into quantum states corresponds to mapping them onto a Hilbert space with dimensions raised to the power of two quantum bits (qubits). Using the kernel matrix constructed on a quantum computer, we performed a KRR, denoted as quantum KRR (QKRR).

In this study, we adopted the natural parameterized quantum circuits (NPQC) kernel, which has been demonstrated to possess performance characteristics similar to the Gaussian kernel, both theoretically and in actual hardware experiments.<sup>91–93</sup> All computations were conducted using Scikit-qulacs.<sup>87,94,95</sup> The quantum kernel was constructed in a 10-qubit space. Hyperparameters for the quantum kernel were determined through grid search. The determined parameters of NPQC kernel were  $c = 1.5$  and the repetition times of embedding 40. The regularization hyperparameter in QKRR was determined using 10 iterations of randomized search.

### 3 Results

In Section 3.1, because we deal with many elements, we compared the dimensional efficiency of our proposed GNN-TL descriptor to well-established physics-inspired descriptors. Note that the GNN-TL descriptor can better handle complex chemical systems by exploiting the GNN-based IAP architecture.

In Section 3.2, we focused on the accuracy of the GNN-TL descriptor in predicting NMR chemical shifts, which are key to understanding molecular details (*e.g.*, interatomic distances and bond angles). This scenario provides an ideal test for determining how well the GNN-TL descriptor works in our study.

Our analysis began by comparing quantum kernel learning, in which the kernels are tested using a quantum computer emulator with traditional kernel learning methods. We then checked the accuracy of the GNN-TL descriptors across the different pretrained GNN models.

Finally, we juxtaposed our GNN-TL descriptor using well-established physics-inspired descriptors. This comparison demonstrates the superiority of the proposed descriptor in terms of efficiency and accuracy. Furthermore, it highlights its potential for accurately predicting chemical properties, which is crucial for advancing research in the molecular and material sciences.

#### 3.1 Dimensional efficiency

At the atomic level, descriptors are tools designed to encode information about atoms within molecules or crystalline materials into vectors. Popular descriptors, such as SOAP and FCHL18, excel at intricately capturing the environment within an atom's cutoff radius. Although these descriptors have achieved significant success in various accuracy benchmarks, they also present challenges due to their large dimensions. Various strategies have been developed to address these challenges,<sup>34,96–98</sup> including refining the descriptor itself, using principal component analysis for dimensionality reduction, and exploring NNs to encode them. In particular, Christensen



**Table 1** Scaling of descriptor dimensions with respect to number of elemental species  $N_{\text{elem}}$ 

	SOAP <sup>a</sup>	FCHL19 <sup>a</sup>	SchNet GNN-TL	MEGNet GNN-TL	M3GNet GNN-TL	MACE-MP0-small GNN-TL	MACE-MP0-large GNN-TL	MACE-OFF23-small GNN-TL	MACE-OFF23-large GNN-TL
$N_{\text{elem}}$	$O(N_{\text{elem}}^2)$	$O(N_{\text{elem}}^2)$	$O(1)$	$O(1)$	$O(1)$	$O(1)$	$O(1)$	$O(1)$	$O(1)$
5	5740	740	128	32	64	128	256	96	224
10	22 680	2440	—	—	64	128	256	96	224
89	1 737 120	162 336	—	—	64	128	256	—	—

<sup>a</sup> SOAP and FCHL19 were generated by Dscribe 0.4.0 (ref. 28) and QML 0.4.0.12,<sup>100</sup> respectively. The default hyperparameters were selected as in QM9NMR paper.

*et al.* applied Behler's method of the atom-centered symmetry function<sup>99</sup> for NN potential to discretize FCHL18 (ref. 33) to derive a compact and accurate FCHL19.<sup>34</sup>

In Table 1, we present the scaling of the SOAP, FCHL19 and various GNN-TL descriptors in response to an increase in the number of elemental species considered. Additionally, for the QM9, QMugs,<sup>85</sup> and MPF.2021.8 or MPTrj datasets,<sup>10</sup> the descriptor dimensions corresponding to 5, 10, and 89 elemental species comprising each dataset are summarized, respectively. Remarkably, with an increase in the number of element types, both SOAP and FCHL19 exhibited quadratic scaling. As a snapshot, when representing five elements in the QM9 dataset, the SOAP and FCHL19 methods have dimensions of 5740 and 740, respectively. This dimensional disparity increases with the number of elemental types. Hence, to represent the 89 elements, the dimensions increased to 1 737 120 and 162 336, respectively. These dimensions are hundreds to tens of thousands of times larger than the compact GNN-TL descriptors, which ranges from 64 to 256 dimensions. Owing to its consistent dimensionality, irrespective of the increase in elements, the GNN-TL descriptors are overwhelmingly compact.

### 3.2 Prediction accuracy: NMR chemical shifts

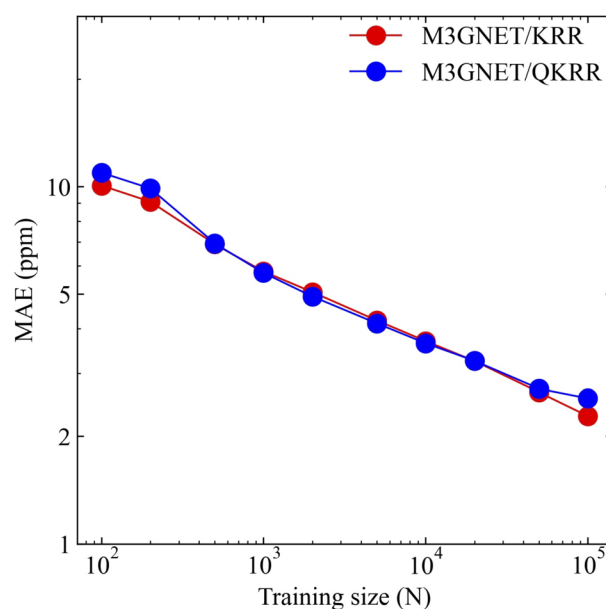
The NMR chemical shifts,  $\delta$ , were predicted using the chemical shielding constant of the reference substance,  $\sigma_{\text{ref}}$ , as the baseline. The NMR chemical shift was calculated using the following equation:

$$\delta = \sigma_{\text{ref}} - \sigma. \quad (4)$$

The reference substances selected for the various nuclei in this study are widely recognized and commonly adopted in the literature.<sup>29,101–104</sup> Specifically, tetramethylsilane was selected for both  $^1\text{H}$  and  $^{13}\text{C}$ , nitromethane ( $\text{MeNO}_2$ ) for  $^{15}\text{N}$ , water- $^{17}\text{O}$  ( $\text{H}_2^{17}\text{O}$ ) for  $^{17}\text{O}$ , and trichlorofluoromethane ( $\text{CFCl}_3$ ) for  $^{19}\text{F}$ . We determined the chemical shielding constants for these well-established reference substances as follows: 31.7608 ppm for  $^1\text{H}$ , 187.0521 ppm for  $^{13}\text{C}$ ,  $-147.8164$  ppm for  $^{15}\text{N}$ , 325.8642 ppm for  $^{17}\text{O}$ , and 171.2621 ppm for  $^{19}\text{F}$ . These constants were evaluated by calculations at the mPW1PW91 (ref. 105)/6-311+G(2d,p) level using density functional theory (DFT) and gauge-including atomic orbital (GIAO)<sup>106</sup> methods. Structure optimization was conducted at the B3LYP<sup>107</sup>/6-31G(2df,p) level in alignment with the methodologies employed

for the QM9 NMR dataset. All calculations were performed using the Gaussian 16 software suite.<sup>108</sup>

In our study, we utilized the QM9NMR dataset, which contains approximately 134K small organic molecules containing C, N, O, and F (excluding H), with each molecule having no more than nine atoms.<sup>29,82</sup> This dataset provides the detailed NMR chemical shielding constants for these molecules. To analyze how the model accuracy changes with training data size, we adopted an approach similar to that used in the original publication of the QM9NMR dataset.<sup>29</sup> Specifically, for  $^{13}\text{C}$ , of a total of 831K data points, we randomly withheld 50K data points to build our test set. Subsequently, from the remaining  $^{13}\text{C}$  NMR chemical shifts, we randomly selected subsets containing 100, 200, 500, 1K, 2K, 5K, 10K, 50K, 100K, and 200K data points to create various training sets. For the other isotopes (*i.e.*,  $^1\text{H}$ ,  $^{15}\text{N}$ ,  $^{17}\text{O}$ , and  $^{19}\text{F}$ ), the test sets were similarly established by withholding 50K, 30K, 50K, and 1K data points, respectively. The training size for  $^{19}\text{F}$  was set to 2K, whereas the other



**Fig. 2** Log-log plot of the training size ( $N$ ) and MAE for the  $^{13}\text{C}$  NMR chemical shielding constant prediction model. The red and blue colors represent the results of the KRR with the Laplacian kernel and QKRR with the NPQC kernel using GNN-TL descriptors from the pre-trained M3GNet model, respectively.





isotopes were trained on datasets of 100K data points. In addition to the QM9 NMR dataset, we sought to validate the performance of our model on external datasets. Hence, we employed the two sets of molecules provided in another study;<sup>29</sup> one consisting of 40 drug molecules from the GDB17 universe and another containing 12 drugs with 17 or more heavy atoms.

Fig. 2 shows the relationship between the mean absolute error (MAE) for the  $^{13}\text{C}$  NMR shielding constant predictions and the training data size. Both QKRR and KRR demonstrated consistent improvements in predictive accuracy with an increase in training size. Notably, the quantum kernel exhibited a performance comparable to that of the Laplacian kernel. For a training size of 100K, the MAE for the  $^{13}\text{C}$  predictions was 2.28 ppm. In a comparative study by Gupta *et al.*, the KRR models using the Coulomb matrix (CM),<sup>109</sup> SOAP, and FCHL descriptors reported MAEs of approximately 4, 2.1, and 1.88 ppm, respectively, for the same training size.<sup>29</sup> Compared with the CM descriptor, our GNN-TL descriptor showed significantly better predictive capabilities, achieving an MAE that was nearly half that of the CM descriptor. Although our method did not exceed the accuracy levels of SOAP and FCHL, the performance of the GNN-TL descriptor was competitive, highlighting its potential as a robust descriptor.

Next, we compared the performance of the GNN-TL descriptors derived from different IAP architectures. Recently, independent of our work, a predictive model for  $^{13}\text{C}$  NMR chemical shielding was proposed using a pretrained IAP known as SchNet, which is a pioneering GNN used as a descriptor.<sup>110</sup> This model was trained on 400 data points of  $^{13}\text{C}$  NMR chemical shielding constants of the molecules in QM9 dataset,<sup>82</sup> with the SchNet GNN-TL descriptor as an input to a feed-forward NN for regression. The predictive accuracy of the SchNet/NN was a root mean-squared error (RMSE) of 12.8 ppm. In pursuit of a fair comparison with their model, we applied KRR using pre-trained MEGNet, M3GNet and MACE GNN-TL descriptors, setting our training data size to 400 data points of  $^{13}\text{C}$  NMR chemical shielding constants. To account for the influence of random sampling, we created 10 different training sets, each comprising 400 data points. The effect of potential data bias was then quantified by calculating the mean RMSE and standard deviation (STD) for each model. Detailed verification including kernel function dependencies can be found in the Appendix. The results of this comparative study are summarized in Table 2. In Table 2, the results for KRR using the Gaussian kernel,

which showed superior accuracy compared to the Laplacian kernel, are presented.

In contrast to the SchNet/NN model's RMSE of 12.8 ppm, the MEGNet/KRR model shows significantly lower predictive accuracy with an RMSE of  $20.08 \pm 0.55$  ppm, suggesting that the MEGNet descriptor is less effective for  $^{13}\text{C}$  NMR chemical shielding data. The M3GNet/KRR model demonstrates a substantial improvement with an RMSE of  $10.02 \pm 0.37$  ppm. Models using MACE descriptors show even greater accuracy: the MACE-MP-0-small/KRR and MACE-MP-0-large/KRR models achieve RMSEs of  $9.77 \pm 0.34$  ppm and  $9.74 \pm 0.27$  ppm, respectively. The best performance is observed with the MACE-OFF23-small/KRR model, which has an RMSE of  $8.05 \pm 0.19$  ppm, with the MACE-OFF23-large/KRR model close behind at  $8.15 \pm 0.42$  ppm. These results highlight the superior performance of the MACE descriptors, particularly MACE-OFF23-small, in enhancing the accuracy of KRR models for predicting  $^{13}\text{C}$  NMR chemical shielding. A more detailed discussion of the nuances of these architectural differences is presented in Section 4.1.

The accuracy of KRR models incorporating the M3GNet GNN-TL descriptor with a Laplacian kernel for NMR chemical shifts was evaluated for each test set of the five different nuclei. Table 3 lists the statistical performance metrics for predicting NMR chemical shifts. Across all elements, the MAE for the test set remained below 5 ppm. The MAE for  $^1\text{H}$  and  $^{19}\text{F}$  were notably low at 0.18 ppm and 2.65 ppm, respectively, indicating a high degree of prediction accuracy for these nuclei in the unseen molecular environments. The MAE for  $^{17}\text{O}$ , although higher at 4.95 ppm, still reflects a reasonable predictive capability, given the complexity of the oxygen chemical shifts. The STD and interquartile range (IQR) values in the Table 3 represent the distribution of chemical shifts within the training data, rather than the accuracy of the model itself. Thus, the higher STD and IQR values for  $^{17}\text{O}$  do not indicate a lack of model precision but rather the natural variability inherent in the  $^{17}\text{O}$  chemical shifts within the training data. The MAE/STD ratio can still offer insights into model performance relative to data variability. For example, the relatively low ratio of  $^{17}\text{O}$  (2.21%) suggests that the model predictions are consistent with the diversity of the training data. On the other hand, the higher ratios for  $^1\text{H}$  (9.09%) and  $^{19}\text{F}$  (7.78%) indicate that the accuracy of the models are not as high as desired, particularly when considering the range of chemical shifts represented in the training dataset. The maximum absolute error (MaxAE) for all nuclei is comparable to the STD of the training data. This is attributed to random sampling and is expected to improve with

**Table 2** The architecture dependence of the predictive performance. For KRR, the Gaussian kernel was applied

GNN-TL descriptor/regressor	RMSE (ppm)
SchNet/NN <sup>a</sup>	12.8
MEGNet/KRR	$20.08 \pm 0.55$
M3GNet/KRR	$10.02 \pm 0.37$
MACE-MP-0-small/KRR	$9.77 \pm 0.34$
MACE-MP-0-large/KRR	$9.74 \pm 0.27$
MACE-OFF23-small/KRR	$8.05 \pm 0.19$
MACE-OFF23-large/KRR	$8.15 \pm 0.42$

<sup>a</sup> The value is taken from ref. 110.

**Table 3** Predictive performance and data variability of NMR shielding constants for 5 elements

	$^1\text{H}$	$^{13}\text{C}$	$^{15}\text{N}$	$^{17}\text{O}$	$^{19}\text{F}$
MAE (ppm)	0.18	2.28	3.42	4.95	2.65
MaxAE (ppm)	7.50	68.58	71.62	279.84	39.31
STD (ppm)	1.98	51.96	119.58	224.40	34.07
IQR (ppm)	2.34	59.93	211.19	354.25	36.77
MAE/STD (%)	9.09	4.38	2.86	2.21	7.78



the application of more sophisticated data point sampling techniques, such as active learning.

Subsequently, these models were employed to predict the NMR chemical shifts of a single molecule  $C_5H_5N_2OF$  containing five elements that was not included in the training data. The results are shown in Fig. 3. The MAE for each nucleus were found to be 0.08 ppm for  $^1H$ , 1.03 ppm for  $^{13}C$ , 6.45 ppm for  $^{15}N$ , 2.86 ppm for  $^{17}O$ , and 6.73 ppm for  $^{19}F$ . The remarkably low MAE

for  $^1H$  and  $^{13}C$  underscores the high accuracy of our model for these nuclei, with predictions that closely mirror the calculated values. The model performed well for the more challenging  $^{15}N$  and  $^{17}O$  nuclei, where the chemical shifts can be significantly affected by subtle changes in the molecular structure and environment, as indicated by the MAE values. The  $^{19}F$  nucleus, while having a higher MAE, showed excellent agreement with the DFT/GIAO calculations, suggesting that the model predictions were robust, even for nuclei with typically higher chemical shift ranges. These results demonstrate the strong predictive power and potential of the model as a reliable tool for accurately predicting NMR chemical shifts across a variety of nuclei, even in molecules beyond the scope of the training data.

We then expanded our assessment to evaluate the predictive ability of our model for molecules larger than those in the QM9 NMR dataset. As such, we incorporate the test sets provided in ref. 29, which comprised 40 drug molecules from the GDB17 universe and another set containing 12 drugs with 17 or more heavy atoms. See ref. 29 for the structures of these molecules.

Table 4 presents the benchmark results for each test set using our M3GNet GNN-TL descriptor and MACE-OFF23-small GNN-TL descriptor. For comparison, we used the FCHL descriptor from Gupta's study.<sup>29</sup> To ensure a fair comparison, we employed our GNN-TL descriptor models trained on a size of 100K  $^{13}C$  chemical shielding constants. For both models, an increased molecular size in the dataset correlated with deterioration of the MAE value. Notably, although our M3GNet GNN-TL descriptor did not match the 1.88 ppm value achieved by the FCHL descriptor for the QM9 50K test set, our model exhibited an MAE value that was approximately 0.3 ppm lower for the 40 GDB17 dataset test. The MACE-OFF23-small GNN-TL descriptor showed even better performance, with an MAE of 1.87 ppm for the QM9 50K test set, closely matching the FCHL descriptor, and significantly outperforming it for the 40 GDB17 dataset with an MAE of 2.83 ppm. For the set of 12 drugs with 17 or more heavy atoms, the M3GNet descriptor showed an MAE of 4.21 ppm, while the MACE-OFF23-small descriptor showed an MAE of 3.85 ppm. Notably, the M3GNet descriptor's accuracy is comparable to the FCHL descriptor. The results were nearly identical for the set of 12 drugs with 17 or more heavy atoms, highlighting that the M3GNet GNN-TL descriptor was less affected by increasing molecular size. On the other hand, the MACE-OFF23-small descriptor significantly outperforms FCHL with an MAE of 3.85 ppm, highlighting its superior predictive performance.

For a detailed comparison, Fig. 6 illustrates the molecule-specific MAE values for both drug test sets. The molecular

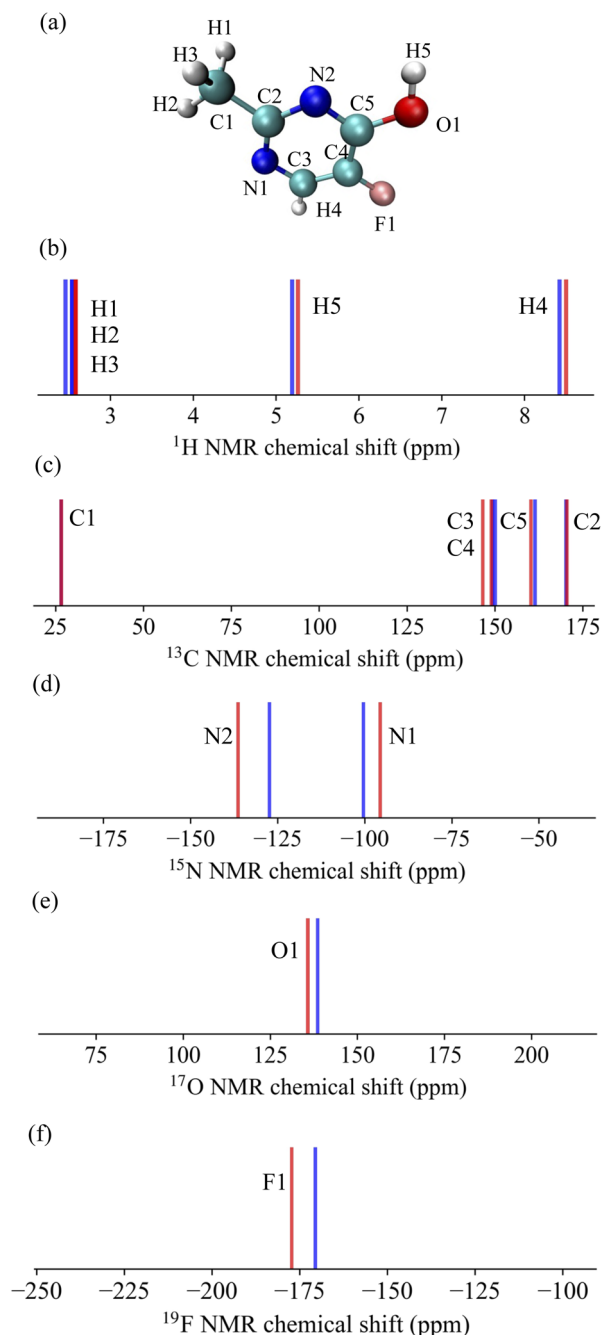


Fig. 3 Predicted NMR chemical shifts for (a) a single molecule, randomly selected from the QM9NMR dataset and not included in the training data, for (b)  $^1H$ , (c)  $^{13}C$ , (d)  $^{15}N$ , (e)  $^{17}O$ , and (f)  $^{19}F$ . These predictions (represented by red lines) are compared with the calculated values at the DFT/GIAO level, which are considered as the correct values (depicted by blue lines).

Table 4 The MAE values for the prediction of the 50K QM9NMR hold out set, 40 drug molecules from GDB17 universe and the other containing 12 drugs with 17 or more heavy atoms. The values in parentheses indicate MaxAE. All units are in ppm

	FCHL <sup>a</sup>	M3GNet GNN-TL	MACE-OFF23-small GNN-TL
50K QM9	1.88	2.28 (68.58)	1.87 (59.76)
40 drugs	3.7	3.46 (29.86)	2.83 (16.08)
12 drugs	4.2	4.21 (20.48)	3.85 (24.70)

<sup>a</sup> FCHL results are taken from ref. 29.



structures are provided in ref. 29. Our M3GNet and MACE-OFF-small GNN-TL descriptor-based prediction models ensured that the highest MAE values for individual molecules across both test sets remained below 10 ppm. Intriguingly, the desflurane molecule, which posed the greatest challenge, showed MAE values of 53.3 ppm, 9.35 ppm and 8.31 ppm for the FCHL, M3GNet and MACE-OFF23-small GNN-TL descriptor models, respectively. This suggests an approximately 80% reduction in the MAE with our descriptor, which is likely attributable to differences in the encompassed descriptor domain.

The cutoff radius for the FCHL descriptor was determined through a grid search,<sup>29</sup> which settled at 4.0 Å. In this scenario, the two fluorine atoms in the terminal trifluoromethyl group (CF<sub>3</sub>) of the desflurane molecule, which lie beyond 4 Å from the CF<sub>2</sub>H carbon, were neglected. In contrast, our M3GNet descriptor had a 6 Å cutoff radius during the initial graph configuration and a 5 Å cutoff for three-body interactions during graph convolution, capturing the entire CF<sub>3</sub> group. This suggests that the descriptor adequately accounts for the influence of the terminal trifluoromethyl group. Additionally, the intrinsic ability of GNN-TL descriptors to account for environments beyond their cutoff radius, owing to graph convolution, may have contributed to the substantial improvement in MAE. Notably, the MACE-OFF23-small model, with a cutoff value of 4.5 Å, achieves the highest accuracy, even though it does not capture the fluorine element at a distance of 4.65 Å in the CF<sub>3</sub> group. In summary, the proposed M3GNet and MACE GNN-TL descriptors demonstrate the capability of predicting <sup>13</sup>C NMR chemical shifts for molecules outside the training dataset with an accuracy comparable to that of the state-of-the-art FCHL descriptor.

Lastly, to explore further practical applications of the constructed models, we validated the NMR chemical shielding constants obtained using semi-empirical PM7-level geometries as inputs against the NMR chemical shift values obtained using DFT/GIAO-level structures from the training data. This validation was performed on the QM9 50K holdout set and two drug molecule test sets, as provided by ref. 29. The <sup>13</sup>C prediction model employed was the M3GNet/KRR model. The MAE values for each molecule in the drug datasets can be found in Fig. 6b and d. For the QM9 50K holdout set, the result was 3.61 ppm, showing a significant deterioration of 1.33 ppm compared to when DFT-level geometries were used as inputs. Conversely, predictions for the 40 drugs and 12 drugs test sets showed only minor deteriorations of 0.23 ppm and 0.04 ppm, respectively. These results suggest that even when using more readily available PM7-level geometries as inputs, the transferability of the model remains robust for extrapolative predictions on larger molecules compared to the training data.

## 4 Discussion

### 4.1 Influence of architectural choices on GNN-TL descriptor performance

In our exploration of different architectures for generating GNN-TL descriptors, we observed several patterns. First, as shown in Tables 1 and 2, it is important to note that the accuracy of GNN-TL descriptors does not necessarily improve with

an increase in the dimensionality of the descriptors. With this in mind, we discuss the architecture of each GNN-based IAP. SchNet, which operates on GNN-based local descriptors to evaluate systems as summations of atomic energies, accounts only for pairwise interactions. This limited inclusion could potentially constrain expressions, leading to inadequate representational power. The subpar performance of MEGNet during transfer learning may be attributed to its architectural design as it integrates atomic (local) descriptors into molecular (global) descriptors through concatenation. This means that the final piece of information passed to the MLP is not extracted directly from the end of the GNN layer, which might not be the optimal representation for targeted atomic-wise property prediction; however, it is expected to be suitable for molecule-wise property predictions. Moreover, the M3GNet architecture, which considers three-body interactions, has the potential to capture the three-dimensional structure of molecules with high resolution. Additionally, the MACE model, an E(3)-equivariant GNN, has demonstrated high performance as an IAP, suggesting that the outputs of its GNN layers are highly accurate in representing molecular structures. The GNN-TL descriptors from the pre-trained E(3) equivariant GNN-based IAP, MACE, show the best predictions. Although the equivariant operations in the GNN layer are not essential for predictions of scalar target properties such as energy or isotropic NMR chemical shifts, the invariant vectors obtained from the equivariant operations may contribute to the sophistication of the GNN-TL descriptors.

### 4.2 Significance of dataset size and diversity

The M3GNet training regimen incorporates data from 187 687 ionic steps spanning 62 783 compounds, including 187 687 energies, 16 875 138 force components, and 1 689 183 stress components. This diverse dataset covers 89 elements from the periodic table. The model is not limited to learning only the energies associated with these elements but extends to atomic-level forces. Moreover, M3GNet training includes not only stable structures but also the processes of structure optimization. The ingestion of vast amounts of data from crystalline systems may have endowed the M3GNet with enhanced expression, potentially making it adept at interpolating molecular systems. The pre-trained MACE-MP0 model was trained using ten times more energy data of crystalline systems, potentially contributing to the improved accuracy of the <sup>13</sup>C NMR chemical shift predictions shown in Table 2. On the other hand, the MACE-OFF23 model, which is specialized for molecules containing 10 elemental species, was trained on a dataset comprising about 1 M energy data points, with structures containing up to 150 atoms. This extensive training dataset might make it more suitable for predicting molecular NMR chemical shifts. Thus, the training data for IAPs, much like their architectures, could be a crucial factor in determining the performance of the descriptors.

### 4.3 Potential for transfer learning on quantum computer

There is a potential for leveraging quantum computation approaches.<sup>111</sup> Specifically, our 10 qubit QKRR, facilitated by a simulator, demonstrated a performance comparable to that of



state-of-the-art KRR. This is underpinned by the theoretical equivalence of the NPQC with the Gaussian kernel. The quantum kernel method stands out because of its capability to compute with fewer measurement iterations than other quantum computation methodologies, such as quantum neural networks.<sup>112</sup> In particular, our proposed M3GNet GNN-TL descriptor can be feasibly realized with a minimum of six qubits, enabling evaluations with a quantum bit count that is more efficient than traditional descriptors, such as SOAP.

However, embedding for higher-dimensional SOAP appears to be a challenge, possibly due to noise. From a futuristic perspective, there is excitement about the possibility of developing kernels that traditional computers cannot express, as well as accelerating the inversion calculations of kernel matrices using quantum algorithms. The constant scaling property of our proposed method concerning element number dimensions may significantly contribute to real-time material exploration powered by quantum-classical hybrid algorithms in the near future.

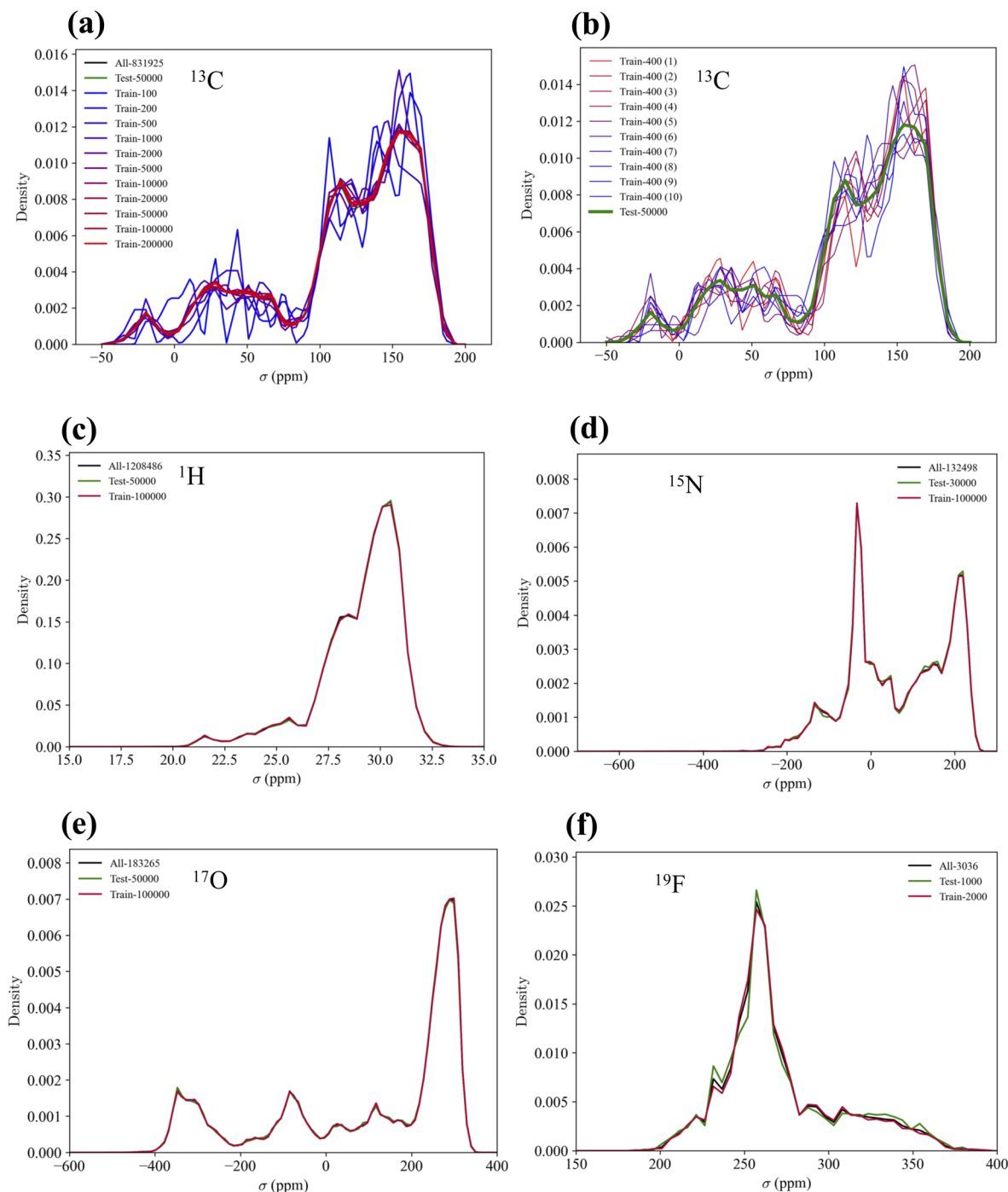


Fig. 4 Distributions of the NMR shielding constants of the training subsets and test set sampled from the of the QM9NMR dataset for the five elemental species: (a)  $^{13}\text{C}$  (for dataset size dependency), (b)  $^{13}\text{C}$  (for potential data bias), (c)  $^1\text{H}$ , (d)  $^{15}\text{N}$ , (e)  $^{17}\text{O}$ , and (f)  $^{19}\text{F}$ , respectively.





## 5 Conclusion

The dynamics of machine learning and its extensive applications across various domains are driving cutting-edge research. Our endeavor to integrate transfer learning with pretrained IAP GNNs for NMR chemical shift prediction offers a paradigm shift in efficiency and scalability. The GNN-TL descriptor presents an unparalleled advantage in terms of scalability due to its consistent dimensionality, irrespective of the number of elements.

Comparative evaluations with other renowned descriptors, such as SOAP, suggest that the GNN-TL descriptor can match, if not surpass, the performance of its contemporaries while maintaining a more compact representation. This is especially important when factoring large datasets, where dimensionality can exponentially burgeon.

Architectural choice plays a pivotal role in the performance of GNN-TL descriptors. Moreover, the diversity and vastness of the training dataset, which encompasses myriad elemental types and structural configurations, augment the robustness and versatility of the GNN.

Our proposed model has immense potential for creating a unified framework capable of predicting various atomic and molecular properties simultaneously, presenting profound implications for accelerated material and molecular research. This potential union of multiple predictions can usher in an era of comprehensive understanding and quicker innovations, possibly revolutionizing fields, such as catalysis, drug discovery, and material design.

The union of transfer learning with pretrained GNNs not only augments prediction accuracy but also drastically reduces learning costs, presenting a cost-effective and efficient alternative to more computationally intensive methods. As we move toward an era in which data-driven insights and models govern the pace of innovation, our research offers a promising pathway for future endeavors in the domain of chemical property predictions with both classical and quantum computers.

Note added – as we were finalizing this manuscript, we became aware of recent articles<sup>86,110,113</sup> that also utilize intermediate information from graph neural network potentials. In Section 3.2, we added a direct comparison between our results and theirs. Elijošius *et al.* applied the pre-trained MACE descriptor to generative modeling of molecules.<sup>86</sup>

## Data availability

Data and code required to reproduce the figures and tables related to the GNN-TL descriptors and the NMR shielding constants prediction models presented in the manuscript is publicly accessible on GitHub at <https://www.github.com/TShiotaSS/gnn-tl>. The dataset utilized for the prediction of NMR chemical shifts, specifically the QM9NMR dataset, is available at DOI: <https://www.doi.org/10.17172/NOMAD/2021.10.16-1> and GitHub at <https://www.moldisgroup.github.io/qm9nmr/>. Results of the DFT/GIAO calculations for isolated atoms, used for NMR chemical shift

computations, are included within the manuscript. The GNN-TL descriptor vectors for the QM9NMR datasets are available at DOI: <https://www.doi.org/10.6084/m9.figshare.25484068.v2>. We have modified the code to extract GNN-TL descriptors from the pretrained M3GNet model on Github at <https://www.github.com/materialsvirtuallab/m3gnet>, and this adapted version can be found at <https://www.github.com/TShiotaSS/gnn-tl/tree/main/scripts/m3gnet>. The code used to extract GNN-TL descriptors from the pretrained MEGNet model can be found on GitHub at <https://www.github.com/materialsvirtuallab/megnet/blob/master/megnet/utis/descriptor.py>. The code used to generate descriptors from the pretrained MACE models can be found on GitHub at <https://www.github.com/ACESuit/mace/blob/main/mace/calculators/mace.py>. The implementation for quantum kernel ridge regression used in this study is available at <https://www.github.com/Qulacs-Osaka/scikit-qulacs/tree/main/skqulacs/qkrr>.

## Conflicts of interest

There are no conflicts to declare.

## Appendix

### Distribution of datasets for each NMR chemical shift prediction model

The distributions of the training and test sets sampled from the QM9NMR dataset are shown in Fig. 4. Fig. 4(a) shows that above 5K, the distribution is in good agreement with the overall distribution of the <sup>13</sup>C NMR shielding constants. For the other elemental species, the distributions of the training and test sets were in good agreement with the overall distribution.

**Table 5** Accuracy (measured by RMSE) of GNN-TL/KRR models trained on 400 <sup>13</sup>C NMR chemical shift values for different kernel functions. All units are in ppm

GNN-TL descriptor	Gaussian kernel	Laplacian kernel
MEGNet	20.08 ± 0.55	21.12 ± 0.56
M3GNet	10.02 ± 0.37	10.31 ± 0.38
MACE-MP-0-small	9.77 ± 0.34	10.78 ± 0.31
MACE-MP-0-large	9.74 ± 0.27	10.17 ± 0.30
MACE-OFF23-small	8.05 ± 0.19	8.64 ± 0.13
MACE-OFF23-large	8.15 ± 0.42	8.77 ± 0.21

**Table 6** The kernel function dependency of accuracy (MAE) for the prediction of the 50K QM9NMR hold out set, 40 drug molecules from GDB17 universe and the other containing 12 drugs with 17 or more heavy atoms. All units are in ppm

	M3GNet		MACE-OFF23-small	
	Gaussian	Laplacian	Gaussian	Laplacian
50K QM9	2.35	2.28	1.87	2.10
40 drugs	3.98	3.46	2.83	3.21
12 drugs	5.14	4.21	3.85	3.93



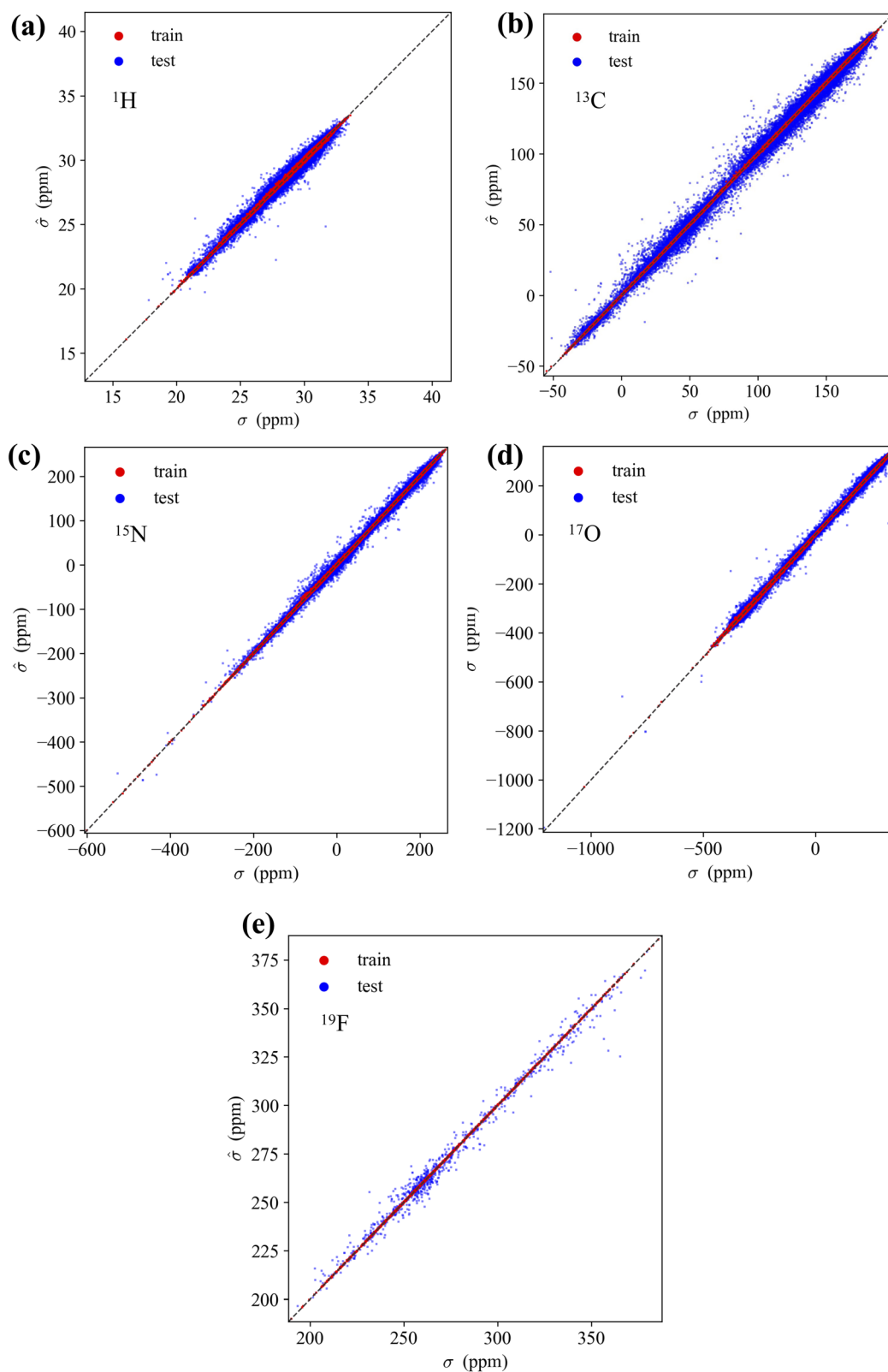
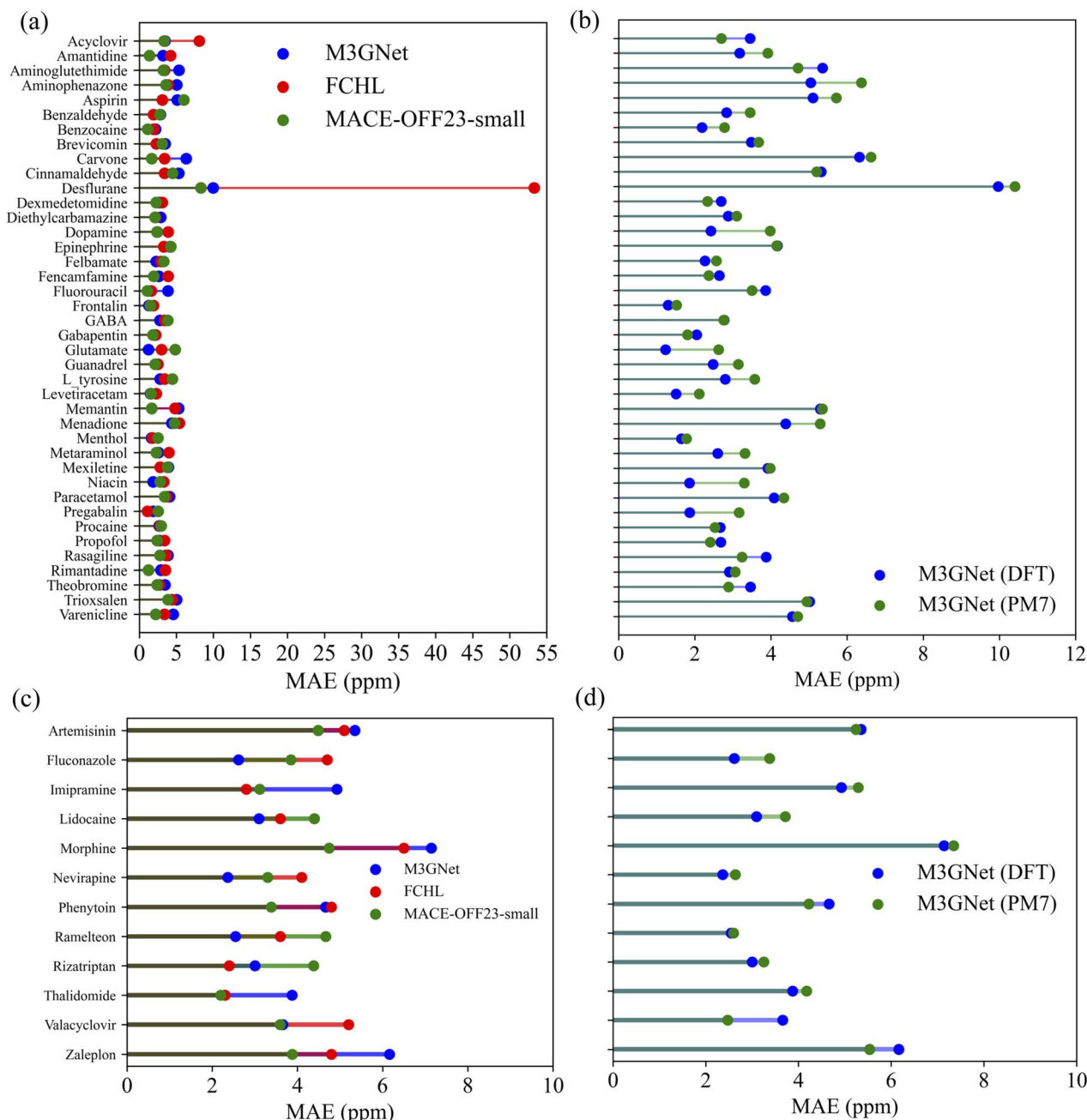


Fig. 5 Scatterplots for the training set (red) and test set (blue) showing NMR chemical shifts from the QM9NMR dataset, using the M3GNet GNN-TL/KRR model constructed with QM9NMR data for the five elemental species: (a)  $^1\text{H}$ , (b)  $^{13}\text{C}$ , (c)  $^{15}\text{N}$ , (d)  $^{17}\text{O}$ , and (e)  $^{19}\text{F}$ , respectively.



**Fig. 6** Comparison of  $^{13}\text{C}$  NMR shielding constant predictions using different descriptors for (a) 40 drug molecules from the GDB17 universe and (d) 12 drugs with 17 or more heavy atoms. The predictions were made using the KRR model with the FCHL descriptor (red), the M3GNet GNN-TL descriptor (blue), and the MACE-OFF23-small GNN-TL descriptor (green). The FCHL results were taken from ref. 29. The results for the M3GNet/KRR model using DFT-level geometries and PM7-level geometries are shown in (b) and (d), respectively.

### Kernel function dependency for various GNN-TL descriptors

The accuracy of KRR models using Gaussian and Laplacian kernels was evaluated. Table 5 presents the mean RMSE and its standard deviation for predictions on the 50K holdout set by models trained on 400 data points of  $^{13}\text{C}$  using Gaussian and Laplacian kernels. For all models using GNN-TL descriptors, the mean RMSE of models with Gaussian kernel was found to be more accurate than those with Laplacian kernel. However, the variation in accuracy due to dataset sampling (standard

deviation) was found to have a greater impact than kernel choice in models with MEGNet and M3GNet GNN-TL descriptors. On the other hand, in models with MACE GNN-TL descriptors, the impact of kernel choice was more significant than the variation due to dataset sampling.

Next, Table 6 shows the accuracy of KRR models using M3GNet and MACE-OFF23-small GNN-TL descriptors trained on a 100K  $^{13}\text{C}$  training set. Unlike models trained on the 400  $^{13}\text{C}$  training set, the KRR models with M3GNet GNN-TL descriptors consistently showed higher accuracy with the Laplacian kernel



compared to the Gaussian kernel. Conversely, the results for MACE-OFF23-small GNN-TL descriptors were similar to those for models trained on the 400  $^{13}\text{C}$  training set, with the Gaussian kernel models demonstrating higher accuracy. This suggests that the appropriate kernel function may vary depending on the size of the training data.

Finally, these results indicate the choice of kernel functions for KRR models as presented in the Results section of this paper. For models trained on 400  $^{13}\text{C}$  data points, all KRR models using GNN-TL descriptors employed the Gaussian kernel. In contrast, for models trained on 100K  $^{13}\text{C}$  data points, the Laplacian kernel was used for KRR models with M3GNet GNN-TL descriptors, whereas the Gaussian kernel was employed for models with MACE-OFF23-small GNN-TL descriptors.

### Validation of learning accuracy of NMR chemical shift prediction

Fig. 5 illustrates the accuracy of the KRR models trained using the M3GNet GNN-TL descriptor for five elemental species. The MAE values for the NMR shielding constant, for train/test, are as follows: for  $^1\text{H}$ , 0.0344/0.1767;  $^{13}\text{C}$ , 0.1420/2.2798 ppm;  $^{15}\text{N}$ , 0.3910/3.4157 ppm;  $^{17}\text{O}$  0.8881/4.9509 ppm; and  $^{19}\text{F}$  0.0864/2.6518 ppm.

The accuracy of the GNN-TL descriptors was also validated using the molecular structures of two drug molecule data sets reported in ref. 29. The predicted  $^{13}\text{C}$  NMR shielding constants for each drug molecule using the M3GNet and MACE-OFF23 GNN-TL/KRR models are shown in Fig. 6a and c. These predictions are accompanied by the values predicted by the FCHL/KRR model.<sup>29</sup> The prediction results of the M3GNet/KRR model using PM7-level optimized geometries, along with the prediction results using DFT-level geometries, are shown in Fig. 6b and d.

## Acknowledgements

We thank Nobuki Inoue and Tuan Minh Do for fruitful discussions. This project was supported by funding from the MEXT Quantum Leap Flagship Program (MEXTQLEAP) through Grant No. JPMXS0120319794, and the JST COI-NEXT Program through Grant No. JPMJPF2014. The completion of this research was partially facilitated by the JSPS Grants-in-Aid for Scientific Research (KAKENHI), specifically Grant No. JP23H03819 and JP21K18933. We thank the Supercomputer Center, the Institute for Solid State Physics, the University of Tokyo for the use of the facilities. This work was also achieved through the use of SQUID at the Cybermedia Center, Osaka University.

## References

- J.-L. Reymond, R. van Deursen, L. C. Blum and L. Ruddigkeit, *MedChemComm*, 2010, **1**, 30.
- L. Ruddigkeit, R. van Deursen, L. C. Blum and J.-L. Reymond, *J. Chem. Inf. Model.*, 2012, **52**, 2864–2875.
- R. Gómez-Bombarelli, J. Aguilera-Iparraguirre, T. D. Hirzel, D. Duvenaud, D. Maclaurin, M. A. Blood-Forsythe, H. S. Chae, M. Einzinger, D.-G. Ha, T. Wu, G. Markopoulos, S. Jeon, H. Kang, H. Miyazaki, M. Numata, S. Kim, W. Huang, S. I. Hong, M. Baldo, R. P. Adams and A. Aspuru-Guzik, *Nat. Mater.*, 2016, **15**, 1120–1127.
- S. Curtarolo, G. L. W. Hart, M. B. Nardelli, N. Mingo, S. Sanvito and O. Levy, *Nat. Mater.*, 2013, **12**, 191–201.
- W. Nie, Q. Wan, J. Sun, M. Chen, M. Gao and S. Chen, *Nat. Commun.*, 2023, **14**, 6671.
- A. R. Oganov, C. J. Pickard, Q. Zhu and R. J. Needs, *Nat. Rev. Mater.*, 2019, **4**, 331–348.
- G. K. Pierens, *J. Comput. Chem.*, 2014, **35**, 1388–1394.
- J. D. Hartman, R. A. Kudla, G. M. Day, L. J. Mueller and G. J. O. Beran, *Phys. Chem. Chem. Phys.*, 2016, **18**, 21686–21709.
- J. B. K. Büning and S. Grimme, *J. Chem. Theory Comput.*, 2023, **19**, 3601–3615.
- C. Chen and S. P. Ong, *Nat. Comput. Sci.*, 2022, **2**, 718–728.
- M. W. Lodewyk, M. R. Siebert and D. J. Tantillo, *Chem. Rev.*, 2011, **112**, 1839–1862.
- G. Lauro, P. Das, R. Riccio, D. S. Reddy and G. Bifulco, *J. Org. Chem.*, 2020, **85**, 3297–3306.
- K. Hansen, F. Biegler, R. Ramakrishnan, W. Pronobis, O. A. Von Lilienfeld, K.-R. Müller and A. Tkatchenko, *J. Phys. Chem. Lett.*, 2015, **6**, 2326–2331.
- V. L. Deringer, A. P. Bartók, N. Bernstein, D. M. Wilkins, M. Ceriotti and G. Csányi, *Chem. Rev.*, 2021, **121**, 10073–10141.
- F. A. Faber, L. Hutchison, B. Huang, J. Gilmer, S. S. Schoenholz, G. E. Dahl, O. Vinyals, S. Kearnes, P. F. Riley and O. A. von Lilienfeld, *J. Chem. Theory Comput.*, 2017, **13**, 5255–5264.
- M. Sajjan, J. Li, R. Selvarajan, S. H. Sureshbabu, S. S. Kale, R. Gupta, V. Singh and S. Kais, *Chem. Soc. Rev.*, 2022, **51**, 6475–6573.
- J. A. Keith, V. Vassilev-Galindo, B. Cheng, S. Chmiela, M. Gastegger, K.-R. Müller and A. Tkatchenko, *Chem. Rev.*, 2021, **121**, 9816–9872.
- K. Wan, J. He and X. Shi, *Adv. Mater.*, 2023, 2305758.
- Z. Wu, B. Ramsundar, E. N. Feinberg, J. Gomes, C. Geniesse, A. S. Pappu, K. Leswing and V. Pande, *Chem. Sci.*, 2018, **9**, 513–530.
- P. Reiser, M. Neubert, A. Eberhard, L. Torresi, C. Zhou, C. Shao, H. Metni, C. van Hoesel, H. Schopmans, T. Sommer, et al., *Commun. Mater.*, 2022, **3**, 93.
- A. P. Bartók, S. De, C. Poelking, N. Bernstein, J. R. Kermode, G. Csányi and M. Ceriotti, *Sci. Adv.*, 2017, **3**, e1701816.
- E. Kocer, T. W. Ko and J. Behler, *Annu. Rev. Phys. Chem.*, 2022, **73**, 163–186.
- M. F. Langer, A. Goeaamann and M. Rupp, *npj Comput. Mater.*, 2022, **8**, 41.
- Z. Liu, L. Lin, Q. Jia, Z. Cheng, Y. Jiang, Y. Guo and J. Ma, *J. Chem. Inf. Model.*, 2021, **61**, 1066–1082.
- A. Merchant, S. Batzner, S. S. Schoenholz, M. Aykol, G. Cheon and E. D. Cubuk, *Nature*, 2023, **624**, 80–85.
- P. Gao, J. Zhang, Q. Peng, J. Zhang and V.-A. Glezakou, *J. Chem. Inf. Model.*, 2020, **60**, 3746–3754.





- 27 S. J. Y. Macalino, V. Gosu, S. Hong and S. Choi, *Arch. Pharmacol. Res.*, 2015, **38**, 1686–1701.
- 28 L. Himanen, M. O. Jäger, E. V. Morooka, F. Federici Canova, Y. S. Ranawat, D. Z. Gao, P. Rinke and A. S. Foster, *Comput. Phys. Commun.*, 2020, **247**, 106949.
- 29 A. Gupta, S. Chakraborty and R. Ramakrishnan, *Machine Learning: Science and Technology*, 2021, **2**, 035010.
- 30 W. Gerrard, L. A. Bratholm, M. J. Packer, A. J. Mulholland, D. R. Glowacki and C. P. Butts, *Chem. Sci.*, 2020, **11**, 508–515.
- 31 M. J. Willatt, F. Musil and M. Ceriotti, *J. Chem. Phys.*, 2019, **150**, 154110.
- 32 A. P. Bartók, R. Kondor and G. Csányi, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2013, **87**, 184115.
- 33 F. A. Faber, A. S. Christensen, B. Huang and O. A. von Lilienfeld, *J. Chem. Phys.*, 2018, **148**, 241717.
- 34 A. S. Christensen, L. A. Bratholm, F. A. Faber and O. Anatole von Lilienfeld, *J. Chem. Phys.*, 2020, **152**, 044107.
- 35 A. Kabylda, V. Vassilev-Galindo, S. Chmiela, I. Poltavsky and A. Tkatchenko, *Nat. Commun.*, 2023, **14**, 3562.
- 36 F. Musil, A. Grisafi, A. P. Bartók, C. Ortner, G. Csányi and M. Ceriotti, *Chem. Rev.*, 2021, **121**, 9759–9815.
- 37 A. P. Bartók, M. C. Payne, R. Kondor and G. Csányi, *Phys. Rev. Lett.*, 2010, **104**, 136403.
- 38 B. Parsaeifard, D. Sankar De, A. S. Christensen, F. A. Faber, E. Kocer, S. De, J. Behler, O. Anatole von Lilienfeld and S. Goedecker, *Machine Learning: Science and Technology*, 2021, **2**, 015018.
- 39 D. Khan, S. Heinen and O. A. von Lilienfeld, *J. Chem. Phys.*, 2023, **159**, 034106.
- 40 M. Rupp, R. Ramakrishnan and O. A. von Lilienfeld, *J. Phys. Chem. Lett.*, 2015, **6**, 3309–3313.
- 41 M. Cordova, E. A. Engel, A. Stefaniuk, F. Paruzzo, A. Hofstetter, M. Ceriotti and L. Emsley, *J. Phys. Chem. C*, 2022, **126**, 16710–16720.
- 42 K. J. Kohlhoff, P. Robustelli, A. Cavalli, X. Salvatella and M. Vendruscolo, *J. Am. Chem. Soc.*, 2009, **131**, 13894–13895.
- 43 M. Tsitsvero, J. Pirillo, Y. Hijikata and T. Komatsuzaki, *J. Chem. Phys.*, 2023, **158**, 194108.
- 44 F. M. Paruzzo, A. Hofstetter, F. Musil, S. De, M. Ceriotti and L. Emsley, *Nat. Commun.*, 2018, **9**, 4501.
- 45 V. Fung, J. Zhang, E. Juarez and B. G. Sumpter, *npj Comput. Mater.*, 2021, **7**, 84.
- 46 Y. Guan, S. V. S. Sowndarya, L. C. Gallegos, P. C. S. John and R. S. Paton, *Chem. Sci.*, 2021, **12**, 12012–12026.
- 47 S. Liu, J. Li, K. C. Bennett, B. Ganoe, T. Stauch, M. Head-Gordon, A. Hexemer, D. Ushizima and T. Head-Gordon, *J. Phys. Chem. Lett.*, 2019, **10**, 4558–4565.
- 48 H. Han and S. Choi, *J. Phys. Chem. Lett.*, 2021, **12**, 3662–3668.
- 49 Y. Kwon, D. Lee, Y.-S. Choi, M. Kang and S. Kang, *J. Chem. Inf. Model.*, 2020, **60**, 2024–2030.
- 50 E. Jonas and S. Kuhn, *J. Cheminf.*, 2019, **11**, 1–7.
- 51 K. T. Schütt, P.-J. Kindermans, H. E. Sauceda, S. Chmiela, A. Tkatchenko and K.-R. Müller, *SchNet: A continuous-filter convolutional neural network for modeling quantum interactions*, 2017.
- 52 C. Chen, W. Ye, Y. Zuo, C. Zheng and S. P. Ong, *Chem. Mater.*, 2019, **31**, 3564–3572.
- 53 J. Gastegger, F. Becker and S. Günnemann, *Advances in Neural Information Processing Systems*, 2021, vol. 34, pp. 6790–6802.
- 54 Y.-L. Liao and T. Smidt, *Equiformer: Equivariant Graph Attention Transformer for 3D Atomistic Graphs*, 2023.
- 55 Y.-L. Liao, B. Wood, A. Das and T. Smidt, *EquiformerV2: Improved Equivariant Transformer for Scaling to Higher-Degree Representations*, 2024.
- 56 K. Xu, W. Hu, J. Leskovec and S. Jegelka, *How Powerful are Graph Neural Networks?*, 2019.
- 57 H. Stärk, D. Beaini, G. Corso, P. Tossou, C. Dallago, S. Günnemann and P. Lió, *Proceedings of the 39th International Conference on Machine Learning*, 2022, pp. 20479–20502.
- 58 D. Jiang, Z. Wu, C.-Y. Hsieh, G. Chen, B. Liao, Z. Wang, C. Shen, D. Cao, J. Wu and T. Hou, *J. Cheminf.*, 2021, **13**, 1–23.
- 59 S. Kang, Y. Kwon, D. Lee and Y.-S. Choi, *J. Chem. Inf. Model.*, 2020, **60**, 3765–3769.
- 60 S. Takamoto, C. Shinagawa, D. Motoki, K. Nakago, W. Li, I. Kurata, T. Watanabe, Y. Yayama, H. Iriguchi, Y. Asano, et al., *Nat. Commun.*, 2022, **13**, 2991.
- 61 A. Musaelian, S. Batzner, A. Johansson, L. Sun, C. J. Owen, M. Kornbluth and B. Kozinsky, *Nat. Commun.*, 2023, **14**, 579.
- 62 I. Batatia, D. P. Kovacs, G. N. C. Simm, C. Ortner and G. Csányi, *Advances in Neural Information Processing Systems*, 2022.
- 63 D. P. Kovács, J. H. Moore, N. J. Browning, I. Batatia, J. T. Horton, V. Kapil, I.-B. Magdău, D. J. Cole and G. Csányi, *arXiv*, 2023, preprint, arXiv:2312.15211, DOI: [10.48550/arXiv.2312.15211](https://doi.org/10.48550/arXiv.2312.15211).
- 64 I. Batatia, P. Benner, Y. Chiang, A. M. Elena, D. P. Kovács, J. Riebesell, X. R. Advincula, M. Asta, W. J. Baldwin, N. Bernstein, A. Bhowmik, S. M. Blau, V. Cărare, J. P. Darby, S. De, F. D. Pia, V. L. Deringer, R. Elijošius, Z. El-Machachi, E. Fako, A. C. Ferrari, A. Genreith-Schriever, J. George, R. E. A. Goodall, C. P. Grey, S. Han, W. Handley, H. H. Heenen, K. Hermansson, C. Holm, J. Jaafar, S. Hofmann, K. S. Jakob, H. Jung, V. Kapil, A. D. Kaplan, N. Karimitari, N. Kroupa, J. Kullgren, M. C. Kuner, D. Kuryla, G. Liepuoniute, J. T. Margraf, I.-B. Magdău, A. Michaelides, J. H. Moore, A. A. Naik, S. P. Niblett, S. W. Norwood, N. O'Neill, C. Ortner, K. A. Persson, K. Reuter, A. S. Rosen, L. L. Schaaf, C. Schran, E. Sivonxay, T. K. Stenczel, V. Svahn, C. Sutton, C. van der Oord, E. Varga-Umbrich, T. Vegge, M. Vondrák, Y. Wang, W. C. Witt, F. Zills and G. Csányi, *A foundation model for atomistic materials chemistry*, 2023.
- 65 A. Merchant, S. Batzner, S. S. Schoenholz, M. Aykol, G. Cheon and E. D. Cubuk, *Nature*, 2023, **624**, 80–85.
- 66 B. Simon, *Nat. Commun.*, 2002, **13**(1), 2453.
- 67 J. Riebesell, R. E. A. Goodall, P. Benner, Y. Chiang, B. Deng, A. A. Lee, A. Jain and K. A. Persson, *Matbench Discovery – A*



- framework to evaluate machine learning crystal stability predictions, 2024.
- 68 J. Han, H. Kang, S. Kang, Y. Kwon, D. Lee and Y.-S. Choi, *Phys. Chem. Chem. Phys.*, 2022, **24**, 26870–26878.
  - 69 N. Grimblat, M. M. Zanardi and A. M. Sarotti, *J. Org. Chem.*, 2015, **80**, 12526–12534.
  - 70 V. A. Semenov and L. B. Krivdin, *Magn. Reson. Chem.*, 2020, **58**, 56–64.
  - 71 T. Schaefer, J. Peeling and G. H. Penner, *Can. J. Chem.*, 1986, **64**, 2162–2167.
  - 72 H. Fukaya and T. Ono, *J. Comput. Chem.*, 2004, **25**, 51–60.
  - 73 H. Chen, S. Viel, F. Ziarelli and L. Peng, *Chem. Soc. Rev.*, 2013, **42**, 7971–7982.
  - 74 J.-X. Yu, R. R. Hallac, S. Chiguru and R. P. Mason, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2013, **70**, 25–49.
  - 75 W. Gerrard, C. Yiu and C. P. Butts, *Magn. Reson. Chem.*, 2022, **60**, 1087–1092.
  - 76 L. B. Krivdin, *Magn. Reson. Chem.*, 2023, **61**, 507–529.
  - 77 K. Matsuzaki, S. Hayashi and W. Nakanishi, *RSC Adv.*, 2024, **14**, 14340–14356.
  - 78 T. Xie, X. Fu, O.-E. Ganea, R. Barzilay and T. Jaakkola, *arXiv*, 2021, preprint, arXiv:2110.06197, DOI: [10.48550/arXiv.2110.06197](https://doi.org/10.48550/arXiv.2110.06197).
  - 79 L. Wu, C. Gong, X. Liu, M. Ye and Q. Liu, *Advances in Neural Information Processing Systems*, 2022, vol. 35, pp. 36533–36545.
  - 80 S. Zaidi, M. Schaarschmidt, J. Martens, H. Kim, Y. W. Teh, A. Sanchez-Gonzalez, P. Battaglia, R. Pascanu and J. Godwin, *arXiv*, 2022, preprint, arXiv:2206.00133, DOI: [10.48550/arXiv.2206.00133](https://doi.org/10.48550/arXiv.2206.00133).
  - 81 S. Jia, A. R. Parthasarathy, R. Feng, G. Cong, C. Zhang and V. Fung, *Digital Discovery*, 2024, **3**, 586–593.
  - 82 R. Ramakrishnan, P. O. Dral, M. Rupp and O. A. Von Lilienfeld, *Sci. Data*, 2014, **1**, 1–7.
  - 83 B. Deng, P. Zhong, K. Jun, J. Riebesell, K. Han, C. J. Bartel and G. Ceder, *Nat. Mach. Intell.*, 2023, **5**, 1031–1041.
  - 84 P. Eastman, P. K. Behara, D. L. Dotson, R. Galvelis, J. E. Herr, J. T. Horton, Y. Mao, J. D. Chodera, B. P. Pritchard, Y. Wang, et al., *Sci. Data*, 2023, **10**, 11.
  - 85 C. Isert, K. Atz, J. Jiménez-Luna and G. Schneider, *Sci. Data*, 2022, **9**, 273.
  - 86 R. Eljosi, F. Zills, I. Batatia, S. W. Norwood, D. P. Kovács, C. Holm and G. Csányi, *arXiv*, 2024, preprint, arXiv:2402.08708, DOI: [10.48550/arXiv.2402.08708](https://doi.org/10.48550/arXiv.2402.08708).
  - 87 F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, *J. Mach. Learn. Res.*, 2011, **12**, 2825–2830.
  - 88 T. Akiba, S. Sano, T. Yanase, T. Ohta and M. Koyama, *Optuna: A Next-generation Hyperparameter Optimization Framework*, 2019.
  - 89 M. Schuld and N. Killoran, *Phys. Rev. Lett.*, 2019, **122**, 040504.
  - 90 T. Kusumoto, K. Mitarai, K. Fujii, M. Kitagawa and M. Negoro, *npj Quantum Inf.*, 2021, **7**, 94.
  - 91 T. Haug, C. N. Self and M. S. Kim, *Machine Learning: Science and Technology*, 2023, **4**, 015005.
  - 92 T. Haug and M. Kim, *Phys. Rev. A*, 2022, **106**, 052611.
  - 93 M. Benedetti, E. Lloyd, S. Sack and M. Fiorentini, *Quantum Sci. Technol.*, 2019, **4**, 043001.
  - 94 <https://github.com/Qulacs-Osaka/scikit-qulacs>.
  - 95 Y. Suzuki, Y. Kawase, Y. Masumura, Y. Hiraga, M. Nakadai, J. Chen, K. M. Nakanishi, K. Mitarai, R. Imai, S. Tamiya, T. Yamamoto, T. Yan, T. Kawakubo, Y. O. Nakagawa, Y. Ibe, Y. Zhang, H. Yamashita, H. Yoshimura, A. Hayashi and K. Fujii, *Quantum*, 2021, **5**, 559.
  - 96 N. Lapanitsyna, G. Fraux, M. A. Springer, S. De and M. Ceriotti, *Phys. Rev. Mater.*, 2023, **7**, 045802.
  - 97 M. J. Willatt, F. Musil and M. Ceriotti, *Phys. Chem. Chem. Phys.*, 2018, **20**, 29661–29668.
  - 98 S. Li, Y. Liu, D. Chen, Y. Jiang, Z. Nie and F. Pan, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2022, **12**, e1558.
  - 99 J. Behler, *J. Chem. Phys.*, 2011, **134**, 074106.
  - 100 A. S. Christensen, L. A. Bratholm, F. A. Faber, B. Huang, A. Tkatchenko, K. R. Müller and O. A. von Lilienfeld, *QML: A Python Toolkit for Quantum Machine Learning*, 2017, <https://github.com/qmlcode/qml>.
  - 101 D. Xin, C. A. Sader, U. Fischer, K. Wagner, P.-J. Jones, M. Xing, K. R. Fandrick and N. C. Gonnella, *Org. Biomol. Chem.*, 2017, **15**, 928–936.
  - 102 C. Puzzarini, G. Cazzoli, M. E. Harding, J. Vázquez and J. Gauss, *J. Chem. Phys.*, 2009, **131**, 234304.
  - 103 R. E. Wasylshen and D. L. Bryce, *J. Chem. Phys.*, 2002, **117**, 10061–10066.
  - 104 C. P. Rosenau, B. J. Jelner, A. D. Gossert and A. Togni, *Angew. Chem., Int. Ed.*, 2018, **57**, 9528–9533.
  - 105 C. Adamo and V. Barone, *J. Chem. Phys.*, 1998, **108**, 664–675.
  - 106 R. Ditchfield, *J. Chem. Phys.*, 1972, **56**, 5688–5691.
  - 107 P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *J. Phys. Chem.*, 1994, **98**, 11623–11627.
  - 108 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr, J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian 16, Revision C.01*, Gaussian Inc., Wallingford CT, 2016.
  - 109 M. Rupp, A. Tkatchenko, K.-R. Müller and O. A. Von Lilienfeld, *Phys. Rev. Lett.*, 2012, **108**, 058301.



- 110 A. M. El-Samman, S. De Castro, B. Morton and S. De Baerdemacker, *Can. J. Chem.*, 2023, **102**(4), 275–288.
- 111 M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio and P. J. Coles, *Nat. Rev. Phys.*, 2021, **3**, 625–644.
- 112 K. Mitarai, M. Negoro, M. Kitagawa and K. Fujii, *Phys. Rev. A*, 2018, **98**, 032309.
- 113 A. M. El-Samman, I. A. Husain, M. Huynh, S. De Castro, B. Morton and S. De Baerdemacker, *Digital Discovery*, 2024, **3**, 544–557.

