

Accurate Prediction of NMR Chemical Shifts: Integrating DFT Calculations with Three-Dimensional Graph Neural Networks

Chao Han,^{||} Dongdong Zhang,^{||} Song Xia, and Yingkai Zhang*



Cite This: *J. Chem. Theory Comput.* 2024, 20, 5250–5258



Read Online

ACCESS |



Metrics & More

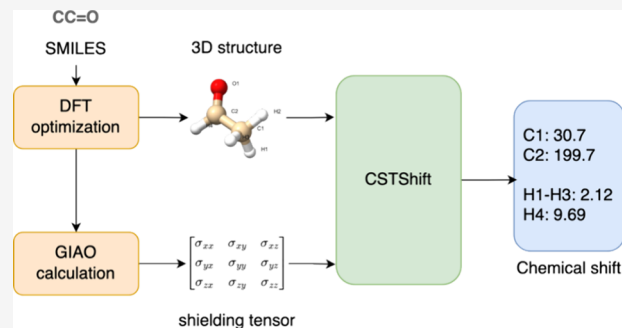


Article Recommendations



Supporting Information

ABSTRACT: Computer prediction of NMR chemical shifts plays an increasingly important role in molecular structure assignment and elucidation for organic molecule studies. Density functional theory (DFT) and gauge-including atomic orbital (GIAO) have established a framework to predict NMR chemical shifts but often at a significant computational expense with a limited prediction accuracy. Recent advancements in deep learning methods, especially graph neural networks (GNNs), have shown promise in improving the accuracy of predicting experimental chemical shifts, either by using 2D molecular topological features or 3D conformational representation. This study presents a new 3D GNN model to predict ^1H and ^{13}C chemical shifts, CSTShift, that combines atomic features with DFT-calculated shielding tensor descriptors, capturing both isotropic and anisotropic shielding effects. Utilizing the NMRShiftDB2 data set and conducting DFT optimization and GIAO calculations at the B3LYP/6-31G(d) level, we prepared the NMRShiftDB2-DFT data set of high-quality 3D structures and shielding tensors with corresponding experimentally measured ^1H and ^{13}C chemical shifts. The developed CSTShift models achieve the state-of-the-art prediction performance on both the NMRShiftDB2-DFT test data set and external CHESHIRE data set. Further case studies on identifying correct structures from two groups of constitutional isomers show its capability for structure assignment and elucidation. The source code and data are accessible at <https://yzhang.hpc.nyu.edu/IMA>.



INTRODUCTION

Nuclear magnetic resonance (NMR) spectroscopy stands as an indispensable tool in various domains of chemistry and biology.^{1–7} NMR chemical shifts, which represent the resonance frequency variation of spin-active nuclei due to distinct atomic local environments, are pivotal for determining molecular connectivity, stereochemistry, and conformation.^{8–11} The computational prediction of chemical shifts plays an increasingly important role in the structure elucidation of small organic molecules, especially when the chemical shift cannot be readily assigned from complex NMR spectra. Although density functional theory (DFT) calculations, enhanced by the gauge-including atomic orbital (GIAO) method, have paved the way for structural assignment and elucidation by predicting NMR chemical shift values,^{12–19} their accuracy is limited despite the significant computational resources required. Empirical scaling has been widely applied to improve NMR chemical shift predictions,¹⁷ in which a linear regression approach is used. The demonstrated accuracy enhancement underscores the potential of integrating more sophisticated, data-driven techniques in the NMR chemical shift prediction.

With the significant advancement of machine learning (ML) techniques and the accumulation of experimental data on NMR chemical shifts in recent years, ML has demonstrated success in the data-driven prediction of NMR chemical shifts

for small molecules, solids, and proteins.^{20–31} Particularly, graph neural networks (GNNs) have showcased their superior ability in representation learning across various chemical and biological tasks without the need for precomputed molecular descriptors or fingerprints, operating in an end-to-end manner.^{32,33} GNNs have found wide applications in chemistry and biology, encompassing molecular property prediction,^{34–39} molecule generation,^{40–42} and the prediction of NMR chemical shift values.^{43–55} Molecules can be modeled as either two-dimensional (2D) graphs, where edges represent chemical bonds connecting the atoms, or as three-dimensional (3D) graphs, where edges of interatomic distances provide more intricate conformational information. Jonas and Kuhn employed a 2D convolutional graph neural network to predict ^1H and ^{13}C chemical shifts along with uncertainties.⁴⁸ Subsequent works improved the design of 2D GNNs, resulting in enhanced performance.^{44,45,50} In comparison with models using 2D topological graphs, 3D GNNs take into account more

Received: March 30, 2024

Revised: May 25, 2024

Accepted: May 29, 2024

Published: June 6, 2024



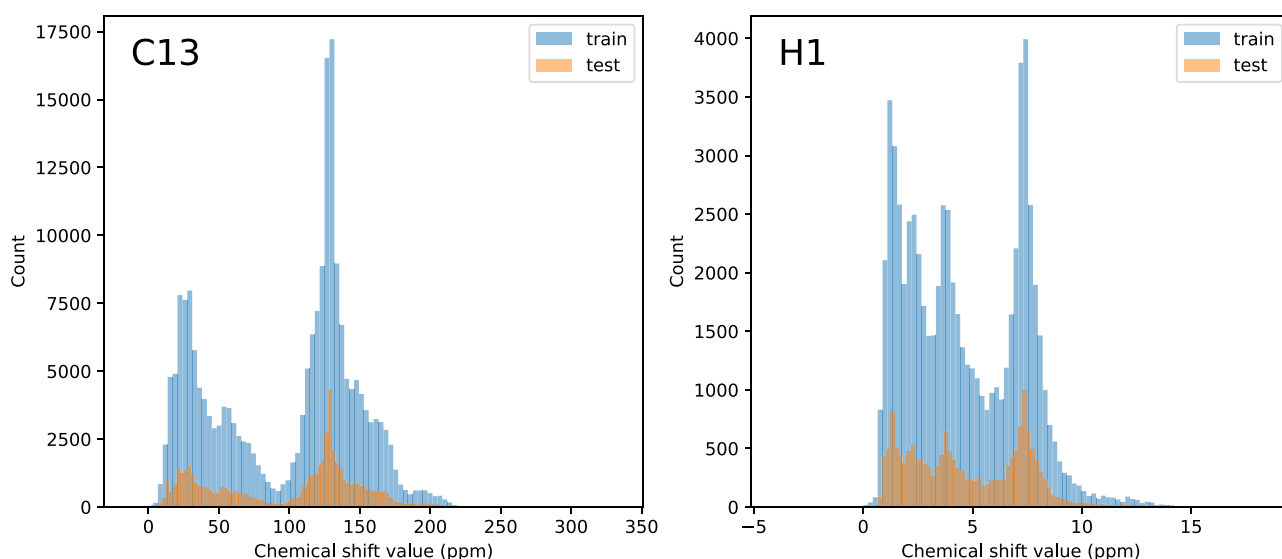


Figure 1. Data distribution of chemical shifts for ^{13}C (left) and ^1H (right) in NMRShiftDB2.

refined geometric factors, leading to superior performance than 2D GNNs when predicting molecular properties like molecular energies.^{34,37} Recent studies have developed 3D GNNs for predicting NMR chemical shifts of small organic molecules and proteins.^{43,46} However, direct comparisons between new 3D GNNs and 2D models are still required to illustrate the necessity of introducing molecular geometry information into the neural network. One novel study by Gao et al. constructed a 3D GNN by combining atomic embedding with calculated isotropic shielding constants to predict experimental chemical shifts in the few-shot setting, with a very small data set of 476 ^{13}C and 217 ^1H experimental chemical shifts including both training and test sets.⁵¹ Their DFT calculations employed M062X/6-31+G(d,p) for geometry optimization and mPW1PW91/6-311+G(2d,p) for the NMR GIAO calculation. Such a high level of DFT calculations is often recommended for the empirical scaling approach (DFT-LR) to predict NMR chemical shifts and is suitable for few-shot learning and small data sets but would be time-consuming for large molecules and for labeling a much bigger data set. Thus, it remains to be explored whether the combination of DFT-calculated NMR information and 3D GNN-learned atomic representation would be a promising approach to achieving accurate prediction of chemical shifts with a much larger data set. In addition, to make such an approach more applicable, the computational cost should also be controlled at an affordable level.

In this work, a 3D GNN model incorporating atomic descriptors from B3LYP/6-31G*-optimized geometries and calculated shielding tensors (CSTs) has been developed to predict ^1H and ^{13}C chemical shifts for small molecules to achieve state-of-the-art performance. CST descriptors are three normalized eigenvalues of the DFT-calculated NMR shielding tensors for all atoms within a given molecule, providing atomic environmental information derived from the electronic structure. Rooted in the sPhysNet architecture,^{56,57} our model CSTShift includes DFT-calculated NMR information by concatenating CST descriptors with atomic representations in the GNN. Based on the NMRShiftDB2^{58,59} data set containing experimental NMR chemical shifts for ^1H and ^{13}C of over 40,000 molecules, DFT optimization and GIAO

calculations at the level of B3LYP/6-31G(d) were conducted to prepare a large data set NMRShift2DB2-DFT comprising computational information and experimental NMR data for model development. With DFT-optimized 3D structures and the incorporation of calculated shielding tensors, our model reduced prediction errors to mean absolute error (MAE) values of 0.944 ppm on ^{13}C and 0.185 ppm on ^1H . CSTShift also outperformed the traditional DFT/GIAO-LR (linear regression) model and previous 3D GNN model on the CHESHIRE data set, achieving MAE values of 0.504 ppm on ^{13}C and 0.078 ppm on ^1H . Last, to demonstrate our model's ability in structural elucidation, we directly tested it on two groups of constitutional isomers, which exhibited superior performance compared to other reported models.

Data Sets. NMRShiftDB2-DFT. We conducted DFT optimization and NMR calculation to prepare the NMRShiftDB2-DFT data set for model training based on the NMRShiftDB2 data set revision 1624, accessible via <https://sourceforge.net/p/nmrshiftdb2/code/1624/>. This experimental data set was initially released by Kuhn et al.^{58,59} and further processed by their subsequent work.⁴⁸ The original NMRShiftDB2 contains over 40,000 compounds with ^1H and ^{13}C chemical shifts from NMR spectra measured under different solvent and temperature conditions. The processed version excluded molecules containing rarely occurred elements, those failing to pass the sanitization check in RDKit,⁶⁰ or those with more than 64 atoms. A total of 26,913 molecules with ^{13}C chemical shifts and 12,806 molecules with ^1H chemical shifts were selected, limited to elements of C, H, O, N, S, Cl, P, and F. It is worth noting that some molecules have multiple entries either in the train set or test set due to their chemical shift records of ^1H and ^{13}C being obtained from different measured spectra. We adopted this processed version and conducted geometry optimization and NMR calculations, as reported in a later section. The data distribution for ^{13}C and ^1H in the provided train/test set from NMRShiftDB2 is shown in Figure 1.

To equip molecules with 3D conformations, we applied three steps to generate and optimize structures of compounds in the NMRShiftDB2 data set. Given a molecule with its SMILES or RDKit mol object, a maximum of 300 conformers

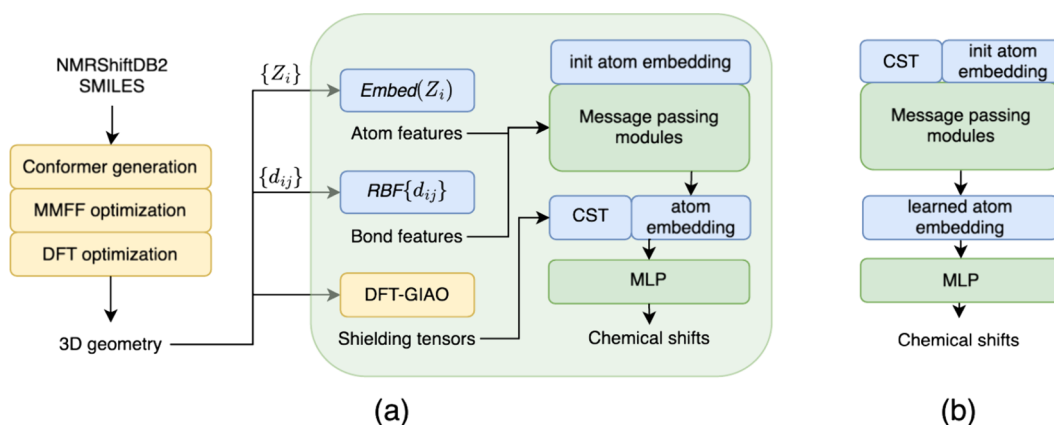


Figure 2. (a) Workflow of CSTShift to predict the NMR chemical shift for organic molecules. Molecular structures are generated by the three-step calculation. As a GNN model, CSTShift takes 3D geometry as input and then embeds atom type $\{Z_i\}$ and interatomic distances $\{d_{ij}\}$ using atomic embedding layers (Embed) and radial basis function (RBF). The atom embedding is updated in message-passing modules iteratively before being concatenated with CST descriptors derived from DFT-GIAO calculation based on previous optimized 3D structures. The concatenated embedding is fed into fully connected layers (MLP) to provide the final prediction of chemical shifts. (b) Alternative model architecture named CSTShift-emb in contrast to CSTShift-out in panel (a). For CSTShift-emb, CST descriptors are concatenated directly with initialized atom embeddings before message-passing modules. Detailed architecture is shown in Figure S1 in the Supporting Information.

were generated using ETKDG⁶¹ and optimized with an MMFF94⁶² force field. After removing similar conformers using Butina clustering⁶³ with a 0.2 Å RMSD cutoff, five conformers with the lowest MMFF energies for each molecule were optimized in the solvent phase (chloroform) using B3LYP/6-31G(d)⁶⁴ by Gaussian.⁶⁵ All conformers were optimized if the total amount was less than 5. In addition, charged molecules with a small amount were excluded. As only atom species are used for the initialization of the atomic embeddings, molecules with the same atom sets but different charges cannot be distinguished by the model. Among the DFT-optimized conformers, the one with the lowest energy was considered the final DFT-generated 3D structure for the molecule.

To include additional information from NMR calculation, we construct the atomic CST descriptors using shielding tensors computed by the DFT-GIAO¹² method. The calculated shielding tensors of a nucleus describe the electronic shielding effect in the external magnetic field.⁶⁶ The average values of these 2-rank shielding tensor eigenvalues are often used to predict the experimental chemical shifts with empirical linear scaling.^{10,17} Strongly correlated to the chemical shifts, shielding tensors could provide great auxiliary information to update atomic embeddings in GNNs. The CST descriptors consist of three eigenvalues of the shielding tensors, including both isotropic and anisotropic effects. The descriptors are normalized for each element individually to provide element-specific information and achieve magnitude alignment with embeddings from previous network layers. Subsequently, these three-dimensional descriptors are concatenated with atomic embeddings, as shown in Figure 2.

To calculate the shielding tensors, the DFT-generated 3D structure undergoes NMR single-point calculation with B3LYP/6-31G(d) and the SMD continuum solvation model.⁶⁷ If solvent information from the experimental measurements for the compounds is explicitly provided, then the solvation modeling is based upon the given solvent. Otherwise, chloroform is used since it is the most used solvent in experimental measurements.

CHESHIRE. The CHESHIRE data set was first created by Rablen et al.⁶⁸ and expanded by Tantillo et al.¹⁷ It served as a benchmark for a series of DFT-based methods with different solvents and solvation models. In the CHESHIRE data set, the test set containing 80 small organic molecules was fitted by linear regression to provide scaling factors, while the probe set containing 25 small organic molecules was used to evaluate the fitting performance. Herein, the CHESHIRE test set serves as an external test set to evaluate the prediction performance.

METHODS

Model Architecture. We developed the graph neural network named CSTShift based on the PhysNet⁵⁷ and sPhysNet models⁵⁶ to predict chemical shifts, the workflow of which is shown in Figure 2. Taking 3D molecular structures optimized from DFT calculations as input, our model can effectively encode atomic environment information through Gaussian expansion of interatomic distances to update atom embeddings. Additionally, shielding tensors from DFT-GIAO calculations are concatenated with the atom embeddings to provide auxiliary information and improve the final prediction. Different solvent conditions are considered using implicit solvent modeling in the DFT optimization and NMR-GIAO calculations, while other experimental conditions, including temperatures, are not included in CSTShift models.

CSTShift constructs initial atom embeddings based solely on the atom species and initial bond features by expanding atomwise distances via radius basis functions (RBFs) from 3D molecular structures. The atom embeddings h_v^t are updated iteratively based on the framework of the message-passing neural network (MPNN):⁶⁹

$$m_v^{t+1} = M_t(h_v^t, d_{vw}) \\ = g_{\text{self}}(h_v^t) + \sum_{w \in N(v)} g_{\text{neighbor}}(h_w^t, \text{RBF}(d_{vw}))$$

$$h_v^{t+1} = U_t(h_v^t, m_v^{t+1}) = u^t \odot h_v^t + f(m_v^{t+1})$$

where h_v^0 is the initial atom features, g_{self} is an activation-first linear layer, g_{neighbor} is a neural network calculating the

interaction from h_w to h_v depending on the RBF expansion of the interatomic distance d_{vw} between atoms w and v , u^t is a learnable parameter vector, and f is a neural network consisting of one residual layer and one linear layer. The architecture is implemented using PyTorch⁷⁰ and Torch-geometric frameworks.⁷¹ Detailed hyperparameters can be found in Table S1 in the Supporting Information.

To further introduce auxiliary information for more accurate GNN predictions, we constructed CST (calculated shielding factor) descriptors from DFT NMR calculations. The components of the shielding tensor for a resonant nucleus N are calculated as the energy derivatives in the form of $\frac{\partial^2 E}{\partial B_i \partial \mu_N^j}$,

where E is the molecular energy, B is the external magnetic field, and μ_N is the nuclear magnetic moment for the atom N . The indices i and j run through 3 directions. Three eigenvalues of these 3×3 matrices for each nucleus are used as the CST descriptors of the corresponding atom after element-wise normalization.

By integration of messages from the atomic environment, the atom embeddings are concatenated with the element-wise normalized CST descriptors, in which DFT calculations directly provide additional information for each atom. Subsequently, the atomic readout layer r reduces the embedding dimension and generates the final prediction labels for each atom:

$$\hat{y}_v = r(\text{CONCAT}(h_v^t, \text{CST}_v))$$

The loss function takes the predicted $\hat{y}_v$ with the true chemical shifts on the NMR-active atoms to calculate the mean absolute error (MAE) and update the model's weights.

To explore the best architecture for leveraging atomic information from DFT calculations, we developed another variant model combining normalized CST descriptors and atom embedding. After initialization, atom embeddings are directly concatenated with CST before the following message-passing steps and output layers. This earlier concatenation enables atomic shielding tensors to revise embeddings for all atoms, while concatenation after message-passing updating keeps more direct influence from CST descriptors and maintains the capability for further transfer learning strategies. We explore the performance of these two implementations and present them in Results and Discussion.

Training and Evaluation Protocols. Our model is trained on the NMRShiftDB2-DFT data set to learn experimental chemical shifts. For comparison with other models, we divide our screened data set using the training and test split from the previous work⁴⁸ after removing molecules, which failed to obtain DFT-optimized structures. We reserved 5% of molecules from the training data set as the validation set. During training, the model parameters are updated using an AMSGrad⁷² optimizer with an initial learning rate of 0.001. The learning rate is scheduled by ReduceLROnPlateau with the decay factor being 0.5 and the patience epoch being 30. The MAE error on the validation data set is used to adjust the learning rate and stop training when the learning rate reaches 5×10^{-8} . Typically, fewer than 500 epochs are needed to complete the training.

The CHESHIRE data set serves as an external test data set to evaluate the model's ability to correctly assign chemical shifts for molecules. Two groups of constitutional isomers, TIC-10⁷³ and NHP,⁷⁴ are used to further illustrate our model's

ability for structure elucidation. It is important to note that when measuring our model's performance using MAE or root-mean-squared error (RMSE), the computation is done on the NMR-active atoms.

RESULTS AND DISCUSSION

Models Trained on NMRShiftDB2-DFT. During the generation of 3D structures by MMFF and DFT calculation, failed or charged molecules are excluded from the data set. Overlapping molecules in the CHESHIRE data set are also removed. In the reduced data set, the multiplicity still exists, where one molecule structure corresponds to several chemical shift records. While many are caused by measurements in different situations such as in different solvents, we notice that there is a portion of molecules sharing the same geometry and nonconflicting chemical shifts. Reasons might include repeating experiment records or partially examining target atom sets in different trials. We keep one unified record for each set to avoid biased training. The total process removed 2% ¹³C samples and 6% ¹H samples, as shown in Table 1.

Table 1. Molecule Counts in the Original NMRShiftDB2 and Processed NMRShiftDB2-DFT Data Set^a

data set	data split	¹³ C	¹ H
NMRShiftDB2	train	21,523	10,252
	test	5390	2554
NMRShiftDB2-DFT	train	21,064	9590
	test	5289	2393

^aStatistics of the original data set comes from the screened version, in which molecules have passed the sanitize check and are restricted by element types and number of atoms.⁴⁸ The processed NMRShiftDB2-DFT data set is used for the training and evaluation.

We first compared our results to other reported results predicting ¹³C and ¹H chemical shifts on NMRShiftDB2. Baseline models include the HOSE method and three graph neural networks. HOSE used atomic featurization and the nearest-neighbor approach to predict chemical shifts.⁷⁵ All the baseline graph neural networks embedded molecules using 2D features. To demonstrate individual contributions, we designed three variants of the CSTShift model: concatenating CST descriptors after embedding initialization (CSTShift-emb), concatenating before the output layers (CSTShift-out), and no concatenation (CSTShift-noCST). Finally, we proposed ensemble models of two variants with concatenation to further improve the performance. Five models are trained individually with random initialization and identical hyperparameters. The average prediction is used as the final ensemble prediction.

The performance comparison in Table 2 shows that the CSTShift model improves the prediction accuracy for ¹³C and achieves state-of-the-art performance on ¹H. The best performance provided by our model is at MAE values of 0.944 and 0.185 ppm for ¹³C and ¹H, respectively. The baseline model without CST descriptors achieves slightly better performance than previous 2D graph neural networks. Incorporating knowledge from CST descriptors further reduces the MAE by around 12% on ¹³C and 5% on ¹H. Both versions of the CST descriptor concatenation model show a similar level of prediction error. The two ensemble models provide the best prediction performance due to the elimination of potential bias from one single model.

Table 2. Performance of CSTShift Models on the NMRShiftDB2-DFT Data Set Compared with Other Reported Models^a

model	description	¹³ C test MAE (ppm)	¹ H test MAE (ppm)
HOSE ⁴⁸	nearest-neighbor search by atom environment encoding	2.850	0.330
GCN ⁴⁸	2D GNN	1.430	0.280
FCG ⁴⁵	2D GNN	1.355 ± 0.022	0.224 ± 0.002
scalable GNN ⁵⁰	2D GNN	1.271 ± 0.008	0.216 ± 0.001
CSTShift-noCST	3D GNN without concatenating CST descriptors	1.164 ± 0.005	0.206 ± 0.002
CSTShift-emb	3D GNN concatenating CST with input atomic embedding (Figure 2b)	1.019 ± 0.005	0.195 ± 0.001
CSTShift-out	3D GNN concatenating CST with message passing-updated embedding (Figure 2a)	1.043 ± 0.013	0.199 ± 0.001
CSTShift-emb (ensemble)	ensemble prediction of 5 CSTShift-emb models	0.944	0.185
CSTShift-out (ensemble)	ensemble prediction of 5 CSTShift-out models	<u>0.959</u>	<u>0.189</u>

^aThe best performance is in bold, and the second best one is underlined. Results of other baseline models are directly cited from corresponding works.

Table 3. Performance Comparison on the CHESHIRE Data Set^a

nuclei	metrics (ppm)	DFT-LR ¹⁷	ExpNN-ff ⁴³ (3D GNN)	CSTShift-noCST	CSTShift-emb (ensemble)	CSTShift-out (ensemble)
¹³ C	RMSE	2.769		2.679 ± 0.173	<u>0.826</u>	0.818
	MAE		1.500	1.276 ± 0.069	0.504	<u>0.505</u>
¹ H	RMSE	0.133		0.236 ± 0.052	0.100	<u>0.101</u>
	MAE			0.152 ± 0.018	<u>0.079</u>	0.078

^aExpNN-ff is a 3D GNN model, while the DFT-LR method uses the linear scaling for DFT calculation.

Direct Test on the CHESHIRE Data Set. To further evaluate the prediction accuracy under extrapolation scenarios, we applied our CSTShift models to the CHESHIRE data set for external validation. The CHESHIRE data set was originally utilized to determine scaling factors for ¹³C and ¹H by linearly fitting the DFT/GIAO-calculated isotropic shielding constants against the experimental chemical shifts. In this data set, 80 molecules in the test set were used for fitting, and 25 molecules in the probe set were used for evaluation.¹⁷ We compared three CSTShift variants with both DFT-LR (linear regression) and another 3D-based GNN model, ExpNN-ff⁴³ on the probe set, which contains 107 ¹³C and 156 ¹H nuclei. To build the data set with 3D structures and calculated shielding tensors, we optimized each molecule and then conducted single-point NMR calculations in chloroform at the level of B3LYP/6-31G(d). Following DFT-LR, we averaged the model output for nuclei with the same chemical environments as the final prediction, for example, three hydrogen atoms from a methyl group.

The performance on the CHESHIRE data set is summarized in Table 3, and the comparison between prediction values of the CSTShift-emb ensemble model and experimental chemical shifts is shown in Figure 3a. The two ensemble models with CST descriptors outperform both the linear model and ExpNN-ff on the ¹³C and ¹H prediction. The prediction error of the CSTShift-noCST model is similar or larger than that of the DFT-LR method, while the precision difference between this and the model with CST descriptors is also larger than the prediction on NMRShiftDB2-DFT. As molecules in CHESHIRE data set are not covered in the NMRShiftDB2 training data set, the worse performance of CSTShift-noCST could be attributed to the lack of similar molecules during training. For example, we compared the prediction performance of the ¹H chemical shifts of the quadricyclane in Figure 3b. This molecule has a bicyclic structure, which is not abundant in the ¹H chemical shift training data set. As a result, the CSTShift-noCST model has worse predictions than the DFT methods for the ¹H chemical shifts of the quadricyclane.

However, the CST descriptors provide the correction for novel molecules, resulting in a large reduction in prediction error for the concatenation models. The combination with DFT-calculated NMR information enables more accurate prediction in situations where there is a lack of similar molecules in the training data set.

Applications on the Structure Elucidation. Next, we applied our CSTShift model to elucidate the structures of two groups of organic compounds by predicting the ¹³C chemical shifts: TIC-10⁷³ and nevirapine hydrolysis product (NHP).⁷⁴ They represent important pharmaceutical compounds containing extended conjugated ring structures. Tumor necrosis factor (TNF)-related apoptosis-inducing ligand (TRAIL) is a cytokine that kills cancer cells but shows little toxicity against normal cells. The active pharmaceutical ingredient TIC-10 was reported to have the capability of inducing TRAIL expression. However, the structure was misassigned as (b) instead of the correct (a) shown in Figure 4.⁷⁶ The structure misassignment could significantly slow the drug development process, as other isomers might lose pharmacological efficacy. Among three isomers shown in Figure 4, only (a) is capable of inducing TRAIL expression. The second test case, NHP, is the hydrolysis product of nevirapine, which is a powerful inhibitor of HIV-1 reverse transcriptase and contributes to the treatment of HIV-1 infection. Four structures (d)–(g) shown in Figure 4 were proposed from the mass spectroscopy and NMR results, where the computational structure elucidation is also required to identify the correct isomer.

Since both ensemble models show similar prediction accuracy across two experimental data sets, we only show the ¹³C chemical shift prediction RMSE of the CSTShift-emb ensemble model for both TIC-10 and NHP isomers in Figure 4. We prepared the 3D structure optimization and GIAO calculation for each isomer with the same process in DMSO at the level of B3LYP/6-31G(d). Compound (a) is selected as the predicted structure with the lowest RMSE among three TIC-10 isomers, which was reported as the pharmacologically active isomer.⁷³ For NHP isomers, our model also achieves

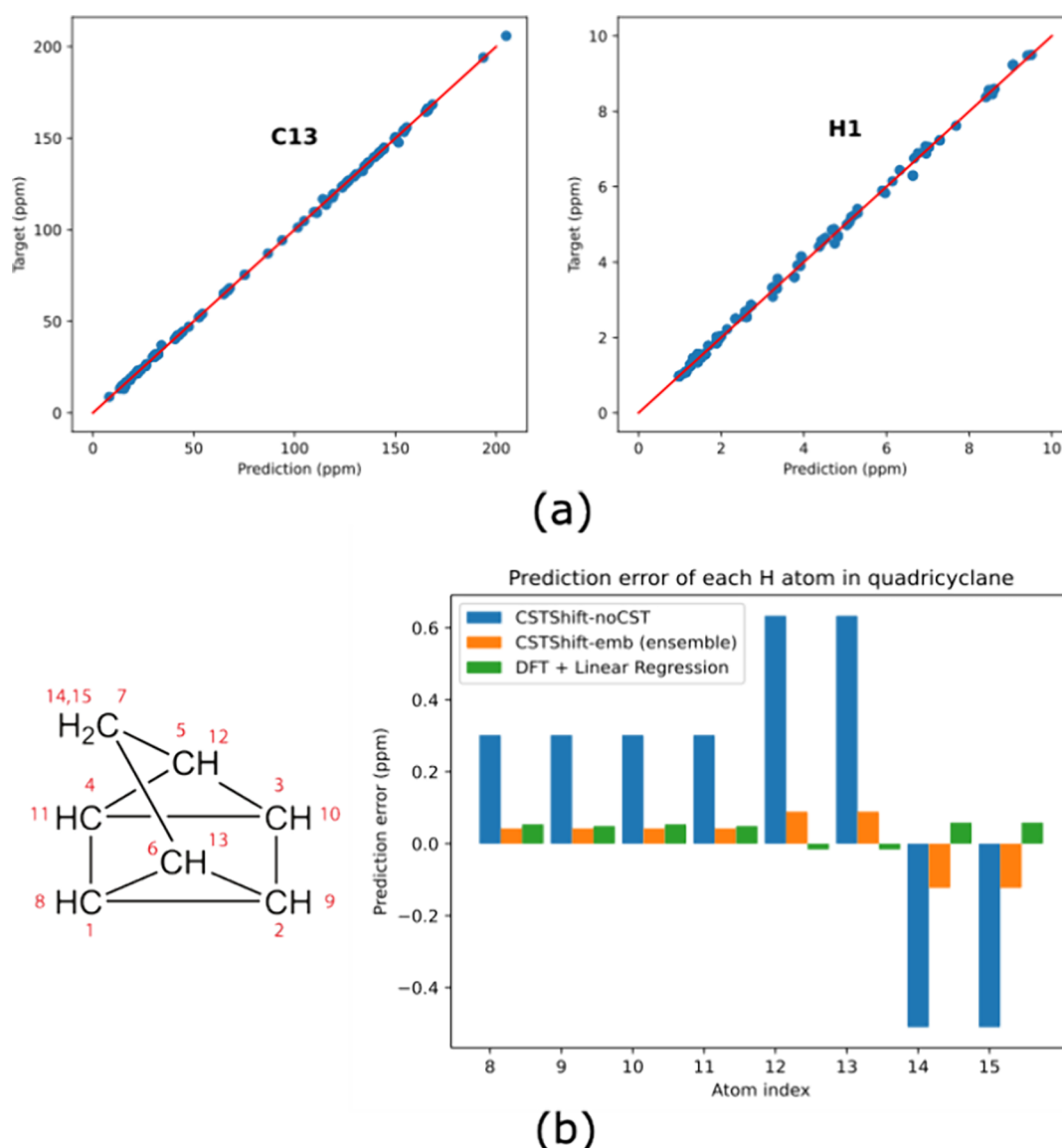


Figure 3. (a) Scatter plots between the predicted and experimental chemical shifts for ^{13}C and ^1H in the CHESHIRE data set by the CSTShift-emb ensemble model. (b) Structure and atom numbering of quadracyclane, with the prediction error comparison on its hydrogen atoms.

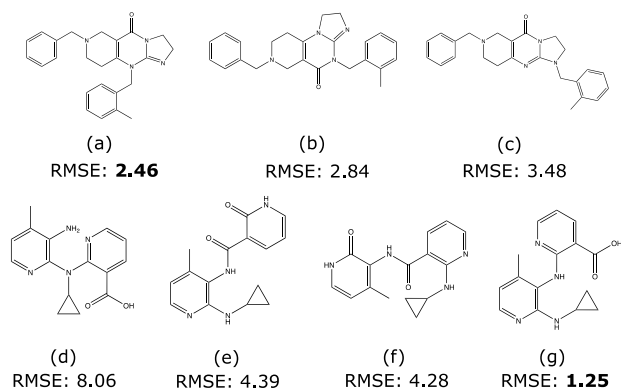


Figure 4. Isomer structures and CSTShift-emb ensemble prediction RMSE (unit: ppm) of TIC-10 (a–c) and NHP (d–g). (a) and (g) are the active isomers for TIC-10 and NHP, respectively.

success elucidation by selecting (g) with the lowest RMSE. Compared with the other two works conducting the same prediction task, our model has better prediction accuracy for

the correct structures (a) and (g). Detailed comparisons are shown in Table S4 in the Supporting Information.

CONCLUSIONS

In this work, we developed CSTShift, a 3D geometry-based GNN model that integrates atomic descriptors derived from DFT-calculated shielding tensors (CSTs) for accurate prediction of ^{13}C and ^1H chemical shifts in small molecules. Leveraging the extensive NMRShiftDB2 data set, which includes experimental chemical shifts for over 40,000 molecules, we conducted DFT optimization and GIAO calculations at the B3LYP/6-31G(d) level to assemble a comprehensive data set NMRShiftDB2-DFT that includes experimentally determined chemical shifts as well as DFT-optimized geometries and calculated shielding tensor (CST) descriptors. Our developed CSTShift model has yielded the best prediction accuracy across two experimental benchmarks, including the external CHESHIRE data set, achieving an MAE of 0.504 ppm for ^{13}C and 0.078 ppm for ^1H . Our model also successfully predicted the bioactive isomer in terms of lowest

RMSE values among two groups of constitutional isomers, showcasing its capability in structure elucidation. As only one 3D structure of each molecule is used as input for our current model, providing multiple conformers to implement the Boltzmann-weighted average prediction has the potential to further improve the model's robustness and applicability. Further improvement of the CSTShift model could also focus on bypassing or reducing the computational cost of DFT calculations for CST descriptors, which will be especially beneficial when scaling up to larger data sets or complex molecular systems.

■ ASSOCIATED CONTENT

Data Availability Statement

The 3D geometries for each compound in NMRShiftDB2-DFT, CHESHIRE, TIC-10, and NHP can be accessible at <https://yzhang.hpc.nyu.edu/IMA>, where the data processing, model building, and training procedure are also provided. RDKit 2022.09 version, PyTorch 1.12.1 version, and PyTorch Geometrics are used for model building/training, respectively.

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.4c00422>.

Model hyperparameters and training setting (Table S1); detailed architecture of the CSTShift model (Figure S1); heavy atom number distributions of molecules in NMRShiftDB2-DFT (Figure S2); predicted ^{13}C chemical shifts of the CHESHIRE data set (Table S2); predicted ^1H chemical shifts of the CHESHIRE data set (Table S3); RMSE comparison for each isomer of TIC-10 and NHP groups (Table S4); atom numbering of TIC-10 isomers (Figure S3); difference between CSTShift-emb ensemble model prediction and experimental ^{13}C chemical shifts of each isomer in TIC-10 (Figure S4); atom numbering of NHP isomers (Figure S5); difference between CSTShift-emb ensemble model prediction and experimental ^{13}C chemical shifts of each isomer in NHP (Figure S6) (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Yingkai Zhang – Department of Chemistry, New York University, New York, New York 10003, United States; Simons Center for Computational Physical Chemistry at New York University, New York, New York 10003, United States; NYU-ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai 200062, China; orcid.org/0000-0002-4984-3354; Email: yingkai.zhang@nyu.edu

Authors

Chao Han – Department of Chemistry, New York University, New York, New York 10003, United States; orcid.org/0009-0008-9979-4475

Dongdong Zhang – Department of Chemistry, New York University, New York, New York 10003, United States

Song Xia – Department of Chemistry, New York University, New York, New York 10003, United States

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acs.jctc.4c00422>

Author Contributions

^{||}C.H. and D.Z. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the U.S. National Institutes of Health (R35-GM127040). S.X. acknowledges support from a Margaret and Herman Sokol fellowship and a graduate fellowship from the Simons Center for Computational Physical Chemistry (SCCPC) at NYU. We thank NYU-ITS for providing computational resources.

■ REFERENCES

- (1) Sanders, J. K. M.; Hunter, B. K. *Modern NMR Spectroscopy*; Oxford University Press: Oxford, 1987.
- (2) Marion, D. An Introduction to Biological NMR Spectroscopy. *Molecular & Cellular Proteomics* **2013**, *12* (11), 3006–3025.
- (3) Bifulco, G.; Dambruoso, P.; Gomez-Paloma, L.; Riccio, R. Determination of Relative Configuration in Organic Compounds by NMR Spectroscopy and Computational Methods. *Chem. Rev.* **2007**, *107* (9), 3744–3779.
- (4) Becker, W.; Bhattiprolu, K. C.; Gubensäk, N.; Zangger, K. Investigating Protein–Ligand Interactions by Solution Nuclear Magnetic Resonance Spectroscopy. *ChemPhysChem* **2018**, *19* (8), 895–906.
- (5) Gossert, A. D.; Jahnke, W. NMR in Drug Discovery: A Practical Guide to Identification and Validation of Ligands Interacting with Biological Macromolecules. *Prog. Nucl. Magn. Reson. Spectrosc.* **2016**, *97*, 82–125.
- (6) Stockman, B. J.; Dalvit, C. NMR Screening Techniques in Drug Discovery and Drug Design. *Prog. Nucl. Magn. Reson. Spectrosc.* **2002**, *41* (3–4), 187–231.
- (7) Wishart, D. S.; Nip, A. M. Protein Chemical Shift Analysis: A Practical Guide. *Biochem. Cell Biol.* **1998**, *76* (2–3), 153–163.
- (8) Willoughby, P. H.; Jansma, M. J.; Hoyer, T. R. A Guide to Small-Molecule Structure Assignment through Computation of (^1H and ^{13}C) NMR Chemical Shifts. *Nat. Protoc.* **2014**, *9* (3), 643–660.
- (9) Wishart, D. S.; Sykes, B. D.; Richards, F. M. Relationship between Nuclear Magnetic Resonance Chemical Shift and Protein Secondary Structure. *J. Mol. Biol.* **1991**, *222* (2), 311–333.
- (10) Jonas, E.; Kuhn, S.; Schlörer, N. Prediction of Chemical Shift in NMR: A Review. *Magn. Reson. Chem.* **2022**, *60* (11), 1021–1031.
- (11) Cavalli, A.; Salvatella, X.; Dobson, C. M.; Vendruscolo, M. Protein Structure Determination from NMR Chemical Shifts. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (23), 9615–9620.
- (12) Wolinski, K.; Hinton, J. F.; Pulay, P. Efficient Implementation of the Gauge-Independent Atomic Orbital Method for NMR Chemical Shift Calculations. *J. Am. Chem. Soc.* **1990**, *112* (23), 8251–8260.
- (13) Muñoz, M. A.; Joseph-Nathan, P. DFT-GIAO ^1H and ^{13}C NMR Prediction of Chemical Shifts for the Configurational Assignment of 6 β -Hydroxyhyoscyamine Diastereoisomers. *Magn. Reson. Chem.* **2009**, *47* (7), 578–584.
- (14) 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.; Gonnella, N. C.; Horspool, K.; Hansen, G.; Senanayake, C. H. 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* (10), 5135–5145.
- (15) Bühl, M.; van Mourik, T. NMR Spectroscopy: Quantum-Chemical Calculations. *WIREs Computational Molecular Science* **2011**, *1* (4), 634–647.
- (16) Zanardi, M. M.; Sortino, M. A.; Sarotti, A. M. On the Effect of Intramolecular H-Bonding in the Configurational Assessment of Polyhydroxylated Compounds with Computational Methods. The Hyacinthacines Case. *Carbohydr. Res.* **2019**, *474*, 72–79.

- (17) Lodewyk, M. W.; Siebert, M. R.; Tantillo, D. J. Computational Prediction of ^1H and ^{13}C Chemical Shifts: A Useful Tool for Natural Product, Mechanistic, and Synthetic Organic Chemistry. *Chem. Rev.* **2012**, *112* (3), 1839–1862.
- (18) Dittmer, A.; Stoychev, G. L.; Maganas, D.; Auer, A. A.; Neese, F. Computation of NMR Shielding Constants for Solids Using an Embedded Cluster Approach with DFT, Double-Hybrid DFT, and MP2. *J. Chem. Theory Comput.* **2020**, *16* (11), 6950–6967.
- (19) Barone, G.; Duca, D.; Silvestri, A.; Gomez-Paloma, L.; Riccio, R.; Bifulco, G. Determination of the Relative Stereochemistry of Flexible Organic Compounds by Ab Initio Methods: Conformational Analysis and Boltzmann-Averaged GIAO ^{13}C NMR Chemical Shifts. *Chem. - Eur. J.* **2002**, *8* (14), 3240–3245.
- (20) Paruzzo, F. M.; Hofstetter, A.; Musil, F.; De, S.; Ceriotti, M.; Emsley, L. Chemical Shifts in Molecular Solids by Machine Learning. *Nat. Commun.* **2018**, *9* (1), 4501.
- (21) Chen, D.; Wang, Z.; Guo, D.; Orekhov, V.; Qu, X. Review and Prospect: Deep Learning in Nuclear Magnetic Resonance Spectroscopy. *Chemistry – A European Journal* **2020**, *26* (46), 10391–10401.
- (22) Kuhn, S.; Egert, B.; Neumann, S.; Steinbeck, C. Building Blocks for Automated Elucidation of Metabolites: Machine Learning Methods for NMR Prediction. *BMC Bioinformatics* **2008**, *9* (1), 400.
- (23) Han, B.; Liu, Y.; Ginzinger, S. W.; Wishart, D. S. SHIFTX2: Significantly Improved Protein Chemical Shift Prediction. *J. Biomol. NMR* **2011**, *50* (1), 43–57.
- (24) Meiler, J.; Maier, W.; Will, M.; Meusinger, R. Using Neural Networks for ^{13}C NMR Chemical Shift Prediction—Comparison with Traditional Methods. *J. Magn. Reson.* **2002**, *157* (2), 242–252.
- (25) Xia, S.; Chen, E.; Zhang, Y. Integrated Molecular Modeling and Machine Learning for Drug Design. *J. Chem. Theory Comput.* **2023**, *19* (21), 7478–7495.
- (26) Aires-de-Sousa, J.; Hemmer, M. C.; Gasteiger, J. Prediction of ^1H NMR Chemical Shifts Using Neural Networks. *Anal. Chem.* **2002**, *74* (1), 80–90.
- (27) Gao, P.; Zhang, J.; Peng, Q.; Zhang, J.; Glezakou, V.-A. General Protocol for the Accurate Prediction of Molecular $^{13}\text{C}/^1\text{H}$ NMR Chemical Shifts via Machine Learning Augmented DFT. *J. Chem. Inf. Model.* **2020**, *60* (8), 3746–3754.
- (28) Li, J.; Bennett, K. C.; Liu, Y.; Martin, M. V.; Head-Gordon, T. Accurate Prediction of Chemical Shifts for Aqueous Protein Structure on “Real World” Data. *Chem. Sci.* **2020**, *11* (12), 3180–3191.
- (29) 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* (2), 508–515.
- (30) Unzueta, P. A.; Greenwell, C. S.; Beran, G. J. O. Predicting Density Functional Theory-Quality Nuclear Magnetic Resonance Chemical Shifts via Δ -Machine Learning. *J. Chem. Theory Comput.* **2021**, *17* (2), 826–840.
- (31) Li, J.; Liang, J.; Wang, Z.; Ptaszek, A. L.; Liu, X.; Ganoe, B.; Head-Gordon, M.; Head-Gordon, T. Highly Accurate Prediction of NMR Chemical Shifts from Low-Level Quantum Mechanics Calculations Using Machine Learning. *J. Chem. Theory Comput.* **2024**, *20* (5), 2152–2166.
- (32) Zhou, J.; Cui, G.; Hu, S.; Zhang, Z.; Yang, C.; Liu, Z.; Wang, L.; Li, C.; Sun, M. Graph Neural Networks: A Review of Methods and Applications. *AI Open* **2020**, *1*, 57–81.
- (33) Reiser, P.; Neubert, M.; Eberhard, A.; Torresi, L.; Zhou, C.; Shao, C.; Metni, H.; van Hoesel, C.; Schopmans, H.; Sommer, T.; Friederich, P. Graph Neural Networks for Materials Science and Chemistry. *Commun. Mater.* **2022**, *3* (1), 1–18.
- (34) 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* (24), No. 241722.
- (35) Yang, K.; Swanson, K.; Jin, W.; Coley, C.; Eiden, P.; Gao, H.; Guzman-Perez, A.; Hopper, T.; Kelley, B.; Mathea, M.; Palmer, A.; Settels, V.; Jaakkola, T.; Jensen, K.; Barzilay, R. Analyzing Learned Molecular Representations for Property Prediction. *J. Chem. Inf. Model.* **2019**, *59* (8), 3370–3388.
- (36) Xiong, Z.; Wang, D.; Liu, X.; Zhong, F.; Wan, X.; Li, X.; Li, Z.; Luo, X.; Chen, K.; Jiang, H.; Zheng, M. Pushing the Boundaries of Molecular Representation for Drug Discovery with the Graph Attention Mechanism. *J. Med. Chem.* **2020**, *63* (16), 8749–8760.
- (37) Gasteiger, J.; Groß, J.; Günnemann, S. *Directional Message Passing for Molecular Graphs*. 2022, arXiv:2003.03123. arXiv.org ePrint archive. DOI: 10.48550/arXiv.2003.03123.
- (38) Zhang, D.; Xia, S.; Zhang, Y. Accurate Prediction of Aqueous Free Solvation Energies Using 3D Atomic Feature-Based Graph Neural Network with Transfer Learning. *J. Chem. Inf. Model.* **2022**, *62* (8), 1840–1848.
- (39) Xia, S.; Zhang, D.; Zhang, Y. Multitask Deep Ensemble Prediction of Molecular Energetics in Solution: From Quantum Mechanics to Experimental Properties. *J. Chem. Theory Comput.* **2023**, *19* (2), 659–668.
- (40) Elton, D. C.; Boukouvalas, Z.; Fuge, M. D.; Chung, P. W. Deep Learning for Molecular Design—a Review of the State of the Art. *Mol. Syst. Des. Eng.* **2019**, *4* (4), 828–849.
- (41) Xiong, J.; Xiong, Z.; Chen, K.; Jiang, H.; Zheng, M. Graph Neural Networks for Automated *de Novo* Drug Design. *Drug Discovery Today* **2021**, *26* (6), 1382–1393.
- (42) Hoogetboom, E.; Satorras, V. G.; Vignac, C.; Welling, M. Equivariant Diffusion for Molecule Generation in 3D. In *Proceedings of the 39th International Conference on Machine Learning*; PMLR, 2022; pp 8867–8887.
- (43) Guan, Y.; Shree Sowndarya, S. V.; Gallegos, L. C.; St. John, P. C.; Paton, R. S. Real-Time Prediction of ^1H and ^{13}C Chemical Shifts with DFT Accuracy Using a 3D Graph Neural Network. *Chem. Sci.* **2021**, *12* (36), 12012–12026.
- (44) Kang, S.; Kwon, Y.; Lee, D.; Choi, Y.-S. Predictive Modeling of NMR Chemical Shifts without Using Atomic-Level Annotations. *J. Chem. Inf. Model.* **2020**, *60* (8), 3765–3769.
- (45) 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* (4), 2024–2030.
- (46) Yang, Z.; Chakraborty, M.; White, A. D. Predicting Chemical Shifts with Graph Neural Networks. *Chem. Sci.* **2021**, *12* (32), 10802–10809.
- (47) Han, H.; Choi, S. Transfer Learning from Simulation to Experimental Data: NMR Chemical Shift Predictions. *J. Phys. Chem. Lett.* **2021**, *12* (14), 3662–3668.
- (48) Jonas, E.; Kuhn, S. Rapid Prediction of NMR Spectral Properties with Quantified Uncertainty. *Journal of Cheminformatics* **2019**, *11* (1), 50.
- (49) Venetos, M. C.; Wen, M.; Persson, K. A. Machine Learning Full NMR Chemical Shift Tensors of Silicon Oxides with Equivariant Graph Neural Networks. *J. Phys. Chem. A* **2023**, *127* (10), 2388–2398.
- (50) Han, J.; Kang, H.; Kang, S.; Kwon, Y.; Lee, D.; Choi, Y.-S. Scalable Graph Neural Network for NMR Chemical Shift Prediction. *Phys. Chem. Chem. Phys.* **2022**, *24* (43), 26870–26878.
- (51) 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**, *11* (22), 9812–9818.
- (52) Bänkestad, M.; Dorst, K. M.; Widmalm, G.; Rönnols, J. *Carbohydrate NMR Chemical Shift Predictions Using E(3) Equivariant Graph Neural Networks*. 2023, arXiv:2311.12657. arXiv.org ePrint archive. DOI: 10.48550/arXiv.2311.12657.
- (53) Chen, Z.; Badman, R. P.; Foley, L.; Woods, R.; Hong, P. *GlycoNMR: Dataset and Benchmarks for NMR Chemical Shift Prediction of Carbohydrates with Graph Neural Networks*. 2023, arXiv:2311.17134. arXiv.org ePrint archive. DOI: 10.48550/arXiv.2311.17134.
- (54) Rull, H.; Fischer, M.; Kuhn, S. NMR Shift Prediction from Small Data Quantities. *J. Cheminform* **2023**, *15* (1), 114.

- (55) Hack, J.; Jordan, M.; Schmitt, A.; Raru, M.; Zorn, H. S.; Seyfarth, A.; Eulenberger, I.; Geitner, R. IIm-NMR-P31: An Open-Access ^{31}P Nuclear Magnetic Resonance Database and Data-Driven Prediction of ^{31}P NMR Shifts. *J. Cheminform* **2023**, *15* (1), 122.
- (56) Lu, J.; Xia, S.; Lu, J.; Zhang, Y. Dataset Construction to Explore Chemical Space with 3D Geometry and Deep Learning. *J. Chem. Inf. Model.* **2021**, *61* (3), 1095–1104.
- (57) Unke, O. T.; Muwly, M. PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments, and Partial Charges. *J. Chem. Theory Comput.* **2019**, *15* (6), 3678–3693.
- (58) Steinbeck, C.; Krause, S.; Kuhn, S. NMRShiftDB - Constructing a Free Chemical Information System with Open-Source Components. *J. Chem. Inf. Comput. Sci.* **2003**, *43* (6), 1733–1739.
- (59) Kuhn, S.; Schlörer, N. E. Facilitating Quality Control for Spectra Assignments of Small Organic Molecules: Nmrshiftdb2 – a Free in-House NMR Database with Integrated LIMS for Academic Service Laboratories. *Magn. Reson. Chem.* **2015**, *53* (8), 582–589.
- (60) Landrum, G. RDKit: Open-Source Cheminformatics. <http://www.RDKit.org>.
- (61) Riniker, S.; Landrum, G. A. Better Informed Distance Geometry: Using What We Know To Improve Conformation Generation. *J. Chem. Inf. Model.* **2015**, *55* (12), 2562–2574.
- (62) Halgren, T. A. Merck Molecular Force Field. II. MMFF94 van Der Waals and Electrostatic Parameters for Intermolecular Interactions. *J. Comput. Chem.* **1996**, *17* (5–6), 520–552.
- (63) Butina, D. Unsupervised Data Base Clustering Based on Daylight's Fingerprint and Tanimoto Similarity: A Fast and Automated Way To Cluster Small and Large Data Sets. *J. Chem. Inf. Comput. Sci.* **1999**, *39* (4), 747–750.
- (64) Becke, A. D. Density-functional thermochemistry. I. The effect of the exchange-only gradient correction. *J. Chem. Phys.* **1993**, *98* (7), 5648–5652.
- (65) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams, D. J.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian* 16, 2016.
- (66) Saitō, H.; Ando, I.; Ramamoorthy, A. Chemical Shift Tensor – the Heart of NMR: Insights into Biological Aspects of Proteins. *Prog. Nucl. Magn. Reson. Spectrosc.* **2010**, *57* (2), 181–228.
- (67) 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* (18), 6378–6396.
- (68) Rablen, P. R.; Pearlman, S. A.; Finkbiner, J. A Comparison of Density Functional Methods for the Estimation of Proton Chemical Shifts with Chemical Accuracy. *J. Phys. Chem. A* **1999**, *103* (36), 7357–7363.
- (69) Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; Dahl, G. E. Neural Message Passing for Quantum Chemistry. In *Proceedings of the 34th International Conference on Machine Learning*; PMLR, 2017; pp 1263–1272.
- (70) Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; Desmaison, A.; Kopf, A.; Yang, E.; DeVito, Z.; Raison, M.; Tejani, A.; Chilamkurthy, S.; Steiner, B.; Fang, L.; Bai, J.; Chintala, S. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems*; 2019; Vol. 32.
- (71) Fey, M.; Lenssen, J. E. Fast Graph Representation Learning with PyTorch Geometric. 2019, arXiv: 1903.02428. arXiv.org ePrint archive. DOI: 10.48550/arXiv.1903.02428.
- (72) Reddi, S. J.; Kale, S.; Kumar, S. On the Convergence of Adam and Beyond. 2019, arXiv: 1904.09237. arXiv.org ePrint archive. DOI: 10.48550/arXiv.1904.09237.
- (73) Jacob, N. T.; Lockner, J. W.; Kravchenko, V. V.; Janda, K. D. Pharmacophore Reassignment for Induction of the Immunosurveillance Cytokine TRAIL. *Angew. Chem.* **2014**, *126* (26), 6746–6749.
- (74) Hargrave, K. D.; Proudfoot, J. R.; Grozinger, K. G.; Cullen, E.; Kapadia, S. R.; Patel, U. R.; Fuchs, V. U.; Mauldin, S. C.; Vitous, J.; Behnke, M. L.; Klunder, J. M.; Pal, K.; Skiles, J. W.; McNeil, D. W.; Rose, J. M.; Chow, G. C.; Skoog, M. T.; Wu, J. C.; Schmidt, G.; Engel, W. W.; Eberlein, W. G.; Saboe, T. D.; Campbell, S. J.; Rosenthal, A. S.; Adams, J.; et al. *J. Med. Chem.* **1991**, *34* (7), 2231–2241.
- (75) Kuhn, S.; Johnson, S. R. Stereo-Aware Extension of HOSE Codes. *ACS Omega* **2019**, *4* (4), 7323–7329.
- (76) Allen, J. E.; Krigsfeld, G.; Mayes, P. A.; Patel, L.; Dicker, D. T.; Patel, A. S.; Dolloff, N. G.; Messaris, E.; Scata, K. A.; Wang, W.; Zhou, J.-Y.; Wu, G. S.; El-Deiry, W. S. Dual Inactivation of Akt and ERK by TIC10 Signals Foxo3a Nuclear Translocation, TRAIL Gene Induction, and Potent Antitumor Effects. *Sci. Transl. Med.* **2013**, *5* (171), 171ra17–171ra17.