Pretrained Graph Transformers for Chirality-related Tasks

Anonymous Authors[1]

Abstract

Graph Neural Networks (GNNs) play a fundamental role in cheminformatics. However, typical GNNs cannot capture the concept of chirality, which means they cannot distinguish chiral molecules from their enantiomers. The ability to distinguish between enantiomers is important as enantiomers may exhibit very distinct biochemical properties. Herein, we proposed a method that leverages the spatial structure of molecules to incorporate conformational information into GNNs. A **P**retrained **G**raph **T**ransformer was designed for **Chi**ral tasks (ChiPGT), which can iteratively optimize raw 3D enantiomer conformations generated by inexpensive methods such as RDKit. The results indicated that our ChiPGT outperforms current state-of-the-art GNNs in chiral-sensitive molecular property prediction tasks on the D4DCHP and REDS datasets.

1. Introduction

Chiral molecules are those whose mirror-image structures (enantiomers) cannot be overlapped by any combination of rotations, translations, and conformational changes (Kasprzyk-Hordern, 2010; Cahn et al., 1966). Most chiral molecules contain at least one chiral center, typically a carbon atom bonded to four different groups. Some chiral molecules do not contain a chiral center but have a chiral axis, known as axial chirality. Although enantiomers have many identical physical and chemical properties, they can exhibit vastly different behaviors when interacting with chiral environments. This makes chiral molecules critically important in various applications, particularly in pharmaceuticals, agrochemicals, and materials science (Kasprzyk-Hordern, 2010; Cahn et al., 1966; Brandt et al., 2017). For instance, the chirality of a drug molecule can significantly influence its therapeutic efficacy and safety. One enantiomer of a chiral drug might be therapeutically active, while the other could be inactive or even harmful (Nguyen et al., 2006).

Graph neural networks (GNNs) are a class of deep learning models that can process graph-structured data, including molecular graphs (Gilmer et al., 2017; Gasteiger et al., 2020; Reiser et al., 2022). Benefiting from graph representation of molecules, GNNs exhibit impressive learning capabilities of molecular structures and promising performance on many chemistry-related tasks, such as molecular property prediction (Yang et al., 2019; Chen & Schwaller, 2024), drug discovery (Brown et al., 2019; Li et al., 2022), and reaction prediction (Chen & Jung, 2022; Tu & Coley, 2022). However, typical GNNs fail to capture the concept of chirality, roughly meaning they cannot distinguish between a molecule and its enantiomer (Fig. 1). Traditional operations directly represent atoms as nodes and chemical bonds as edges, resulting in two-dimensional molecular graphs with identical topological structures for enantiomers.

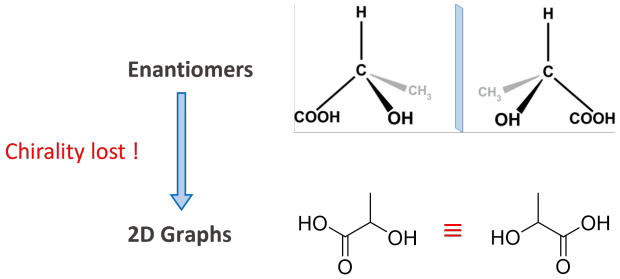

Figure 1. The 2D molecular graph loses chirality.

The recent developments of GNNs provide message passing with information about bond distances, bond angles, and torsion angles of conformational isomers (Schütt et al., 2017; Liu et al., 2022; Gasteiger et al., 2021). These methods theoretically enable the GNNs to differentiate enantiomers, while they exhibit poor performance on chirality-related tasks (Adams et al., 2021). Some researchers have attempted to design chirality-sensitive models explicitly to incorporate chiral information into GNNs, but the performance of these models still needs improvement (Schneider et al., 2018; Pattanaik et al., 2020; Mamede et al., 2021; Gaiński et al., 2023). Additionally, methods designed specifically for chiral centers not only fail to handle axial chiral molecules

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Under review at AI4Mat 2024 workshop. Do not distribute.

but may also lead to a decrease in the ability to handle non-chiral tasks (Gaiński et al., 2023).

In this work, we adopt an approach that starts from the overall spatial structure of molecules and incorporates conformational information into GNNs. We designed a graph transformer (ChiPGT) capable of effectively handling chirality information. The ChiPGT's graph encoder is pre-trained on PCQM4MV2 dataset (Hu et al., 2021) with extensive DFT-optimized molecular geometries, and able to optimize raw 3D enantiomer conformations generated by inexpensive methods such as RDKit. The results prove that our ChiPGT outperforms current state-of-the-art methods in chiral-sensitive molecular property prediction tasks on the D4DCHP and REDS datasets.

2. Related Work

The most common method for incorporating chirality into GNNs is using local or global chirality labels (Schneider et al., 2018; Mamede et al., 2021). For each carbon atom with four nonequivalent groups (referred to as a chiral center), we can order the four groups according to specific rules. The position of the highest numbered group (4) should be oriented away from the observer, and the relative positions of the groups numbered (1), (2) and (3) determine the chiral center's clockwise (CW) or counterclockwise (CCW) direction, forming the chirality label. Both local and global chirality labels have very limited expressivity.

A theoretically feasible approach to incorporating chiral information into graph neural network models is to provide the set of torsion angles to 3D GNNs. However, even access to a complete set of torsion angles does not guarantee the expressivity for chirality-sensitive tasks, as torsion angles are sensitive to bond rotation and can be negated by reflections of non-chiral molecules(Adams et al., 2021). In(Adams et al., 2021), the authors proposed the ChIRo model, which embeds sets of torsion angles with common bonds rather than individual torsion angles. ChIRo explicitly models conformational flexibility by integrating a novel type of intramolecular bond rotation invariance into the architecture, thus reducing the need for extensive conformational data augmentation. Tests show that ChIRo effectively learns chiral information and achieves satisfactory performance on R/S classification tasks.

Another approach to incorporating chiral information into graph neural network models is to modify the message-passing scheme. In typical GNNs, incoming messages from neighboring nodes are treated as a set and aggregated using permutation-invariant functions such as summation or averaging. Since enantiomers have the same graph connectivity, symmetrical aggregators operating on different chiral centers will collapse their neighbors into identical representations regardless of chirality, making message passing unable to distinguish chiral graphs. In (Pattanaik et al., 2020), the authors proposed the Tetra-DMPNN model, which replaces the classic message-passing scheme with a chirality-sensitive scheme. The proposed aggregation scheme is guided by local chirality labels, and its effectiveness in tasks such as enantiomer ranking demonstrates its ability to distinguish enantiomers. In(Gaiński et al., 2023), the authors proposed a chirality-sensitive message-passing scheme that does not rely on distances, angles or torsion angles but instead uses the directions of neighbors around a node. The authors applied this method to construct the chiral edge neural network (ChiENN) layer in the context of molecular chirality which can be attached to any GNN model to achieve chiral awareness. ChiENN is the current state-of-the-art method for chirality-sensitive tasks.

3. Method

3.1. ChiPGT Architecture

As shown in Fig. 2, the ChiPGT model consists of a graph encoder and an output block. Based on Uni-Mol+ (Lu et al., 2023) architecture, our ChiPGT's graph encoder can iteratively update 3D conformations and atom embeddings using molecular graphs with coarse atomic coordinates as input. The initial atomic positions are generated from SMILES through inexpensive methods such as RDKit. ChiPGT's graph encoder is pre-trained on PCQM4MV2 dataset (Hu et al., 2021) with extensive DFT-optimized molecular geometries, and able to optimize molecular conformations towards an equilibrium state. Subsequently, the ChiPGT is fine-tuned on chirality-related datasets without DFT-optimized structures. The optimized 3D conformations and corresponding atomic embeddings output from the graph encoder are used to predict chirality-related properties through the output block.

3.2. Graph Encoder Backbone

In the ChiPGT's graph encoder, a raw conformation is iteratively updated to its target DFT equilibrium conformation, and the learned conformation will be encoded to atom embedding and then be used to predict the chirality-related properties. To effectively learn this update process towards the equilibrium conformation, the graph encoder maintains two tracks of representations: 1) Atom representation ($x \in \mathbb{R}^{n \times d_x}$, where d_x is the dimension of the atom representation) and 2) Pair representation ($p \in \mathbb{R}^{n \times n \times d_p}$, where d_p is the dimension of the pair representation).

Update of Atom Representation The atom representation $\boldsymbol{x}^{(0)}$ is initialized by the embeddings of atom features. At

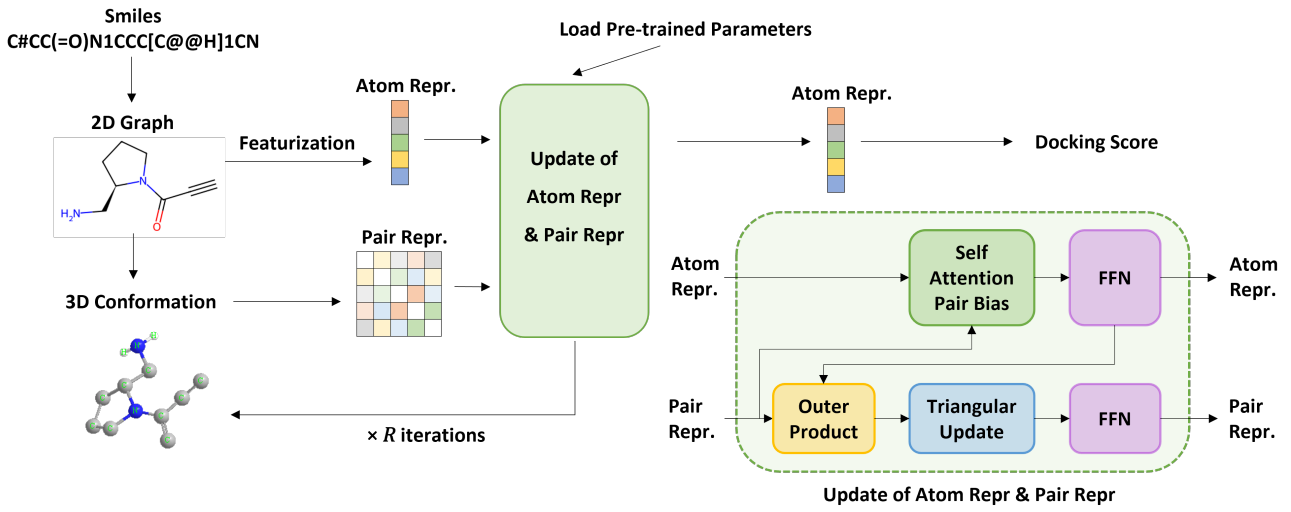

Figure 2. Model architecture of ChiPGT for molecular property prediction tasks.

l-th block, $\boldsymbol{x}^{(l)}$ is sequentially updated as follow:

$$\boldsymbol{x}^{(l)} = \boldsymbol{x}^{(l-1)} + \text{SelfAttention}\left(\boldsymbol{x}^{(l-1)}, \boldsymbol{p}^{(l-1)}\right), \quad (1)$$

$$x^{(l)} = x^{(l)} + \text{FFN}\left(x^{(l)}\right). \quad (2)$$

The SelfAttention function f_1 is denoted as:

$$\boldsymbol{Q}^{(l,h)} = \boldsymbol{x}^{(l-1)}\boldsymbol{W}_Q^{(l,h)}, \quad \boldsymbol{K}^{(l,h)} = \boldsymbol{x}^{(l-1)}\boldsymbol{W}_K^{(l,h)}, \quad (3)$$

$$\boldsymbol{B}^{(l,h)} = \boldsymbol{p}^{(l-1)}\boldsymbol{W}_B^{(l,h)}, \quad \boldsymbol{V}^{(l,h)} = \boldsymbol{x}^{(l-1)}\boldsymbol{W}_V^{(l,h)}, \quad (4)$$

$$f_1 = \text{softmax}\left(\frac{\boldsymbol{Q}^{(l,h)}\left(\boldsymbol{K}^{(l,h)}\right)^T}{\sqrt{d_h}} + \boldsymbol{B}^{(l,h)}\right)\boldsymbol{V}^{(l,h)}, \quad (5)$$

where d_h is the head dimension, $\boldsymbol{W}_Q^{(l,h)}, \boldsymbol{W}_K^{(l,h)}, \boldsymbol{W}_V^{(l,h)} \in \mathbb{R}^{d_x \times d_h}$, $\boldsymbol{W}_B^{(l,h)} \in \mathbb{R}^{d_p \times 1}$. FFN is a feedforward network with one hidden layer. For simplicity, layer normalizations are omitted. Compared to the standard Transformer layer, the only difference here is the usage of attention bias term $B^{(l,h)}$ to incorporate $\boldsymbol{p}^{(l-1)}$ from the pair representation track.

Update of Pair Representation The pair representation $\boldsymbol{p}^{(0)}$ is initialized by the positional encoding ψ. The update process of pair representation begins with an outer product of $\boldsymbol{x}^{(l)}$, followed by a $\mathcal{O}\left(n^3\right)$ triangular multiplication, and is then concluded with an FFN layer. Formally, at l-th block, $\boldsymbol{p}^{(l)}$ is sequentially updated as follow:

$$\boldsymbol{p}^{(l)} = \boldsymbol{p}^{(l-1)} + \text{OuterProduct}\left(\boldsymbol{x}^{(l)}\right), \quad (6)$$

$$\boldsymbol{p}^{(l)} = \boldsymbol{p}^{(l)} + \text{TriangularUpdate}\left(\boldsymbol{p}^{(l)}\right), \quad (7)$$

$$\boldsymbol{p}^{(l)} = \boldsymbol{p}^{(l)} + \text{FFN}\left(\boldsymbol{p}^{(l)}\right). \quad (8)$$

The OuterProduct function (f_2) is used for atom-to-pair communication, denoted as:

$$\boldsymbol{a} = \boldsymbol{x}^{(l)}\boldsymbol{W}_{O1}^{(l)}, \quad \boldsymbol{b} = \boldsymbol{x}^{(l)}\boldsymbol{W}_{O2}^{(l)}, \quad (9)$$

$$\boldsymbol{o}_{i,j} = \text{flatten}\left(\boldsymbol{a}_i \otimes \boldsymbol{b}_j\right), \quad (10)$$

$$f_2 = \boldsymbol{o}\boldsymbol{W}_{O3}^{(l)}, \quad (11)$$

where $\boldsymbol{W}_{O1}^{(l)}, \boldsymbol{W}_{O2}^{(l)} \in \mathbb{R}^{d_z \times d_o}$, d_o is the hidden dimension of OuterProduct, and $\boldsymbol{W}_{O3}^{(l)} \in \mathbb{R}^{d_o^2 \times d_p}$, $\boldsymbol{o} = [\boldsymbol{o}_{i,j}]$. Please note that $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{o}$ are temporary variables in the OuterProduct function. TriangularUpdate function (f_3) is used to enhance pair representation further, denoted as:

$$\boldsymbol{a} = \text{sigmoid}\left(\boldsymbol{p}^{(l)}\boldsymbol{W}_{T1}^{(l)}\right) \odot \left(\boldsymbol{p}^{(l)}\boldsymbol{W}_{T2}^{(l)}\right), \quad (12)$$

$$\boldsymbol{b} = \text{sigmoid}\left(\boldsymbol{p}^{(l)}\boldsymbol{W}_{T3}^{(l)}\right) \odot \left(\boldsymbol{p}^{(l)}\boldsymbol{W}_{T1}^{(l)}\right), \quad (13)$$

$$\boldsymbol{o}_{i,j} = \sum_k \boldsymbol{a}_{i,k} \odot \boldsymbol{b}_{j,k} + \sum_k \boldsymbol{a}_{k,i} \odot \boldsymbol{b}_{k,j}, \quad (14)$$

$$f_3 = \text{sigmoid}\left(\boldsymbol{p}^{(l)}\boldsymbol{W}_{T5}^{(l)}\right) \odot \left(\boldsymbol{o}\boldsymbol{W}_{T6}^{(l)}\right), \quad (15)$$

where $\boldsymbol{W}_{T1}^{(l)}, \boldsymbol{W}_{T2}^{(l)}, \boldsymbol{W}_{T3}^{(l)}, \boldsymbol{W}_{T4}^{(l)} \in \mathbb{R}^{d_p \times d_t}$, $\boldsymbol{W}_{T5}^{(l)} \in \mathbb{R}^{d_p \times d_p}$, $\boldsymbol{W}_{T6}^{(l)} \in \mathbb{R}^{d_t \times d_p}$, $o = [\boldsymbol{o}_{i,j}]$, and d_t is the hidden dimension of TriangularUpdate. $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{o}$ are temporary variables.

4. Experiments

4.1. Datasets

For chirality-related property prediction tasks, we conduct our experiments on two datasets designed specifically to evaluate the capability of a model to express chirality.

D4DCHP(Pattanaik et al., 2020) is a dataset assessing chiral interactions of molecules in the D4 dopamine receptor protein. There are a large number of chiral centers in complex proteins, making protein pockets a chiral environment, so enantiomers can show different interaction energies when complexed with target proteins.

Ranking Enantiomers by Docking Scores dataset(later referred to as REDS dataset) (Adams et al., 2021) is also a dataset consists of docking scores for enantiomeric pairs of molecules. The difference is, REDS defined two different tasks: enantiomer ranking and docking score prediction. The former involves predicting which molecule in an enantiomeric pair has relatively lower binding affinity in a chiral protein pocket, while the latter, sharing the same molecular data as the former, focuses on regression for docking score.

4.2. Baselines

We benchmark our method, ChiPGT, in comparison to several GNN architectures, including general GNNs: GPS (Rampášek et al., 2022), SAN (Kreuzer et al., 2021), DMPNN (Yang et al., 2019), and chirality-sensitive GNNs: Tetra-DMPNN (Pattanaik et al., 2020), ChIRo (Adams et al., 2021), ChiENN (Gaiński et al., 2023), with ChiENN being the current state-of-the-art method for chiral-sensitive tasks. Models not specifically designed for chirality are considered to add chiral atom labels to their node features or stack them with the chirality-sensitive layer of ChiENN.

4.3. Evaluation Metrics

For the molecular property prediction task on D4DCHP, the evaluation metric employed is the Root Mean Square Error (RMSE), which evaluates the regression accuracy for docking scores prediction. For the two different tasks in REDS, the evaluation metrics are: Enantiomer ranking accuracy, which evaluates the accuracy of predicting the molecule with lower binding energy in enantiomeric pairs, and Mean Absolute Error (MAE), which evaluates the regression accuracy for docking score prediction.

4.4. Training details

Before model training, we first use the RDKit method to generate 8 initial conformations for each molecule at a cost of approximately 0.005s per molecule.

For model training, we choose Mean Squared Error (MSE) as the loss function, used the AdamW optimizer with a learning rate of 5×10^{-5} for D4DCHP dataset and 2×10^{-4} for REDS dataset, a batch size of 128 (molecules). We trained for 30 epochs on D4DCHP dataset (287,468 molecules) and for 250 epochs on REDS dataset (69,120 molecules), 10% of which are warm-up steps.

The training took approximately 5 hours for each dataset, utilizing one NVIDIA A6000 GPU. The inference on the D4DCHP test set (28,746 molecules) took approximately 9 minutes, and on the REDS test set (10,386 molecules) took approximately 2 minutes, utilizing one NVIDIA A6000 GPU. The reported statistics are averages of five runs, using standord training/validation/testing split and different initial random seeds.

4.5. Results

Table 1. The comparison between the ChiPGT and baseline models on D4DCHP dataset. The results of baselines are from Tetra-DMPNN work (Pattanaik et al., 2020).

METHOD	BINDING AFFINITY RMSE ↓
GCN + PERM	6.67 ± 0.06
GIN + PERM	6.39 ± 0.06
DMPNN + PERM	6.39 ± 0.06
GCN + PERM-CAT	6.62 ± 0.06
GIN + PERM-CAT	6.37 ± 0.07
DMPNN + PERM-CAT	6.38 ± 0.06
CHIPGT (OUR METHOD)	$\mathbf{5.99 \pm 0.01}$

As depicted in Table 1 and 2, it can be observed that our method achieved state-of-the-art performance on regression tasks across two chiral-related datasets. On D4DCHP dataset, our ChiPGT model significantly outperformed the chiral message passing baselines in the prediction of docking score, achieving a root mean square error (RMSE) of 5.99 kcal/mol. This represents a reduction of approximately 6% compared to the previous best method combination (GIN + PERM-CAT), which uses chirality-sensitive aggregation functions for message passing and achieves an RMSE of 6.37 kcal/mol. On the REDS dataset, we compared the ChiPGT method with the current state-of-the-art chiral-sensitive GNNs for the docking score prediction. ChiPGT achieved the lowest mean absolute error (MAE) of 0.249 kcal/mol, surpassing the current best ChiENN method combined with the SAN, which had an MAE of 0.257 kcal/mol.

4.6. Ablation Study

We conduct an ablation study to study the impact of pretraining (Table 3). The ChiPGT models trained from scratch exhibited an increase in RMSE from 5.95 to 6.50 on the D4DCHP dataset, while MAE increased from 0.248 to

0.283 and the enantiomer ranking accuracy decreased from 0.738 to 0.665 on the REDS dataset. The results prove that ChiPGT's prediction performance on chirality-related tasks can be obviously improved through pre-training on the PCQM4MV2 dataset (Hu et al., 2021) with extensive DFT-optimized molecular geometries.

Table 2. The comparison between the ChiPGT and baseline models on REDS dataset. The results of baselines are from ChiENN work (Gaiński et al., 2023).

METHOD	ENANTIOMER RANKING	DOCKING SCORE
	R.ACCURACY ↑	MAE ↓
DMPNN	0.000±0.000	0.310±0.001
GPS	0.000±0.000	0.330±0.003
SAN	0.000±0.000	0.317±0.004
DMPNN + TAGS	0.701±0.003	0.285±0.001
GPS + TAGS	0.669±0.037	0.318±0.004
SAN + TAGS	0.722±0.004	0.278±0.003
CHIRO	0.691±0.006	0.359±0.009
CHIENN	0.760±0.002	0.275±0.003
GPS + CHIENN	0.753±0.004	0.258±0.001
SAN + CHIENN	**0.764±0.005**	0.257± 0.002
CHIPGT	0.732± 0.004	**0.249± 0.001**

Table 3. Ablation study on D4DCHP and REDS.

METHOD	D4DCHP	REDS	
	RMSE ↓	R.ACC ↑	MAE ↓
FROM SCRATCH	6.50±0.01	0.665±0.011	0.283±0.004
FINETUNING	**5.99±0.01**	**0.732±0.004**	**0.249±0.001**

5. Conclusion

In this work, we propose a method that leverages the spatial structure of molecules to incorporate conformational information into GNNs, aiming to address the challenges in handling chiral molecules. We employ the proposed method to construct a pre-trained graph Transformer specifically for chiral tasks (ChiPGT). The ChiPGT's graph encoder is pre-trained on PCQM4MV2 dataset with extensive DFT-optimized molecular geometries, and able to optimize raw 3D enantiomer conformations generated by inexpensive methods such as RDKit. The experimental results demonstrate that the ChiPGT significantly outperforms current state-of-the-art GNNs in chirality-sensitive molecular property prediction tasks.

References

Adams, K., Pattanaik, L., and Coley, C. W. Learning 3D Representations of Molecular Chirality with Invariance to Bond Rotations, 2021. arXiv:2110.04383 [cs].

Brandt, J. R., Salerno, F., and Fuchter, M. J. The added value of small-molecule chirality in technological applications. *Nature Reviews Chemistry*, 1(6):0045, 2017.

Brown, N., Fiscato, M., Segler, M. H., and Vaucher, A. C. GuacaMol: Benchmarking Models for de Novo Molecular Design. *Journal of Chemical Information and Modeling*, 59(3):1096–1108, 2019.

Cahn, R. S., Ingold, C., and Prelog, V. Specification of molecular chirality. *Angewandte Chemie International Edition in English*, 5(4):385–415, 1966.

Chen, J. and Schwaller, P. Molecular hypergraph neural networks. *The Journal of Chemical Physics*, 160(14), 2024.

Chen, S. and Jung, Y. A generalized-template-based graph neural network for accurate organic reactivity prediction. *Nature Machine Intelligence*, 4(9):772–780, 2022.

Gaiński, P., Koziarski, M., Tabor, J., and Śmieja, M. ChiENN: Embracing Molecular Chirality with Graph Neural Networks. In *Machine Learning and Knowledge Discovery in Databases: Research Track: European Conference, ECML PKDD 2023, Turin, Italy, September 18–22, 2023, Proceedings, Part III*, pp. 36–52, Berlin, Heidelberg, 2023. Springer-Verlag.

Gasteiger, J., Groß, J., and Günnemann, S. Directional message passing for molecular graphs. *arXiv preprint arXiv:2003.03123*, 2020.

Gasteiger, J., Becker, F., and Günnemann, S. GemNet: Universal Directional Graph Neural Networks for Molecules. In *Advances in Neural Information Processing Systems*, volume 34, pp. 6790–6802. Curran Associates, Inc., 2021.

Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., and Dahl, G. E. Neural message passing for quantum chemistry. In *International conference on machine learning*, pp. 1263–1272. PMLR, 2017.

Hu, W., Fey, M., Ren, H., Nakata, M., Dong, Y., and Leskovec, J. OGB-LSC: A Large-Scale Challenge for Machine Learning on Graphs. *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, 1, 2021.

Kasprzyk-Hordern, B. Pharmacologically active compounds in the environment and their chirality. *Chemical Society Reviews*, 39(11):4466–4503, 2010.

Kreuzer, D., Beaini, D., Hamilton, W., Létourneau, V., and Tossou, P. Rethinking Graph Transformers with Spectral Attention. In *Advances in Neural Information Processing Systems*, volume 34, pp. 21618–21629. Curran Associates, Inc., 2021.

Li, M. M., Huang, K., and Zitnik, M. Graph representation learning in biomedicine and healthcare. *Nature Biomedical Engineering*, 6(12):1353–1369, 2022.

Liu, Y., Wang, L., Liu, M., Zhang, X., Oztekin, B., and Ji, S. Spherical Message Passing for 3D Graph Networks, 2022. URL http://arxiv.org/abs/2102.05013. arXiv:2102.05013 [cs].

Lu, S., Gao, Z., He, D., Zhang, L., and Ke, G. Highly accurate quantum chemical property prediction with uni-mol+. *arXiv preprint arXiv:2303.16982*, 2023.

Mamede, R., de Almeida, B. S., Chen, M., Zhang, Q., and Aires-de Sousa, J. Machine Learning Classification of One-Chiral-Center Organic Molecules According to Optical Rotation. *Journal of Chemical Information and Modeling*, 61(1):67–75, 2021.

Nguyen, L. A., He, H., and Pham-Huy, C. Chiral drugs: an overview. *International journal of biomedical science: IJBS*, 2(2):85–100, 2006.

Pattanaik, L., Ganea, O.-E., Coley, I., Jensen, K. F., Green, W. H., and Coley, C. W. Message Passing Networks for Molecules with Tetrahedral Chirality, 2020. arXiv:2012.00094 [cs, q-bio].

Rampášek, L., Galkin, M., Dwivedi, V. P., Luu, A. T., Wolf, G., and Beaini, D. Recipe for a General, Powerful, Scalable Graph Transformer. *Advances in Neural Information Processing Systems*, 35:14501–14515, 2022.

Reiser, P., Neubert, M., Eberhard, A., Torresi, L., Zhou, C., Shao, C., Metni, H., van Hoesel, C., Schopmans, H., Sommer, T., et al. Graph neural networks for materials science and chemistry. *Communications Materials*, 3(1): 93, 2022.

Schneider, N., Lewis, R. A., Fechner, N., and Ertl, P. Chiral Cliffs: Investigating the Influence of Chirality on Binding Affinity. *ChemMedChem*, 13(13):1315–1324, 2018.

Schütt, K., Kindermans, P.-J., Sauceda Felix, H. E., Chmiela, S., Tkatchenko, A., and Müller, K.-R. SchNet: A continuous-filter convolutional neural network for modeling quantum interactions. In *Advances in Neural Information Processing Systems*, 2017.

Tu, Z. and Coley, C. W. Permutation Invariant Graph-to-Sequence Model for Template-Free Retrosynthesis and Reaction Prediction. *Journal of Chemical Information and Modeling*, 62(15):3503–3513, 2022.

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., and Barzilay, R. Analyzing Learned Molecular Representations for Property Prediction. *Journal of Chemical Information and Modeling*, 59(8):3370–3388, 2019.