Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI-24)
Demonstrations Track

3D-FuM: Benchmarking 3D Molecule Learning with Functional Groups

Tingwei Chen , **Jianpeng Chen** and **Dawei Zhou**
Department of Computer Science, Virginia Tech, VA, USA
cccchentingwei@gmail.com, {jianpengc, zhoud}@vt.edu

Abstract

Molecular graph representation learning plays a crucial role in various domains, such as drug discovery and chemical reaction prediction, where molecular graphs are typically depicted as 2D topological structures. However, recent insights highlight the critical role of 3D geometric information and functional groups in accurately predicting molecular properties, aspects often neglected in existing molecular graph benchmark datasets. To bridge the research gap, we introduce a comprehensive molecular learning benchmark named 3D-FUM, which incorporates both 3D geometric information and functional groups of a large number of molecules. 3D-FUM integrates 18 state-of-the-art algorithms and 19 evaluation metrics on three molecular learning tasks, including general molecule generation, conditional molecule generation, and property predictions. 3D-FUM, for the first time, take into consideration both 3D geometric information and molecular functional groups, which enables researchers and practitioners to effectively and impartially evaluate newly proposed methods in comparison to existing baselines across diverse datasets. Furthermore, we design a user interface for user-friendly interaction and development with the benchmark for evaluation metrics selection, parameter adjustment, and leaderboard comparison. To ensure accessibility and reproducibility, we open-source our benchmark 3D-FUM and experimental results at https://3dfunctiongroupmoleculedataset.github.io/3D-FuM/#/Home.

1 Background

The utilization of artificial intelligence (AI) model in chemistry, such as drug discovery [Doshi and Chepuri, 2022; Morselli Gysi *et al.*, 2021] and chemical reaction prediction [Nakao *et al.*, 2023; Unsleber *et al.*, 2022], has received much attention in recent years due to their enhanced efficiency in handling large-scale molecular data. These works concentrate on a wide range of topics and consistently contribute to various molecular-related tasks. Notably, the tasks of molecular conformation generation and molecular property prediction have gained significant attention and remarkable achievements. For example, Garcia Satorras *et al.* [2021] introduce E-NF for molecular generation, and Brandstetter *et al.* [2022] propose SEGNN for molecular energy prediction. More recently, Wang *et al.* [2022] and Peng *et al.* [2022] consider fragments to design their model and obtain improved performance, which demonstrates the significance of function groups for molecular-related tasks. On the other hand, to assess these AI models in real-world applications, choosing a comprehensive benchmark are imperative. In the realm of molecular research, QM9 [Ruddigkeit *et al.*, 2012] and ZINC [Sterling and Irwin, 2015] are commonly employed for evaluation.

Despite the remarkable achievements, previous studies overlook the importance of 3D and functional groups that play a pivotal role in molecular tasks. On the one hand, 3D information determines some crucial molecular properties, *e.g.*, lacking 3D info will lead to uniform molecular generation results [Patronov *et al.*, 2022]. On the other hand, with the help of fragments, researchers can explore a broader chemical space while screening a limited number of compounds. It leads to fewer inevitable interactions, which enhances the probability of achieving targeted interactions [Jinsong *et al.*, 2024; Xu *et al.*, 2024].

Although Some efforts have managed to introduce 3D molecular information [Thölke and Fabritiis, 2022; Liao and Smidt, 2023; Huang *et al.*, 2023] or molecular fragment information [Kong *et al.*, 2022], they fall short in incorporating them together. These works have demonstrated the effectiveness of incorporating 3D information and functional groups, highlighting the urgent need for a comprehensive benchmark and dataset that integrates them to prompt further advancements of AI models.

However, the integration of 3D information with denoised fragment details posed several challenges. **C1. Molecular representation:** It is challenging to determine representative functional groups that can enhance the universality and interpretability of molecular representation without introducing excessive noise. **C2. Benchmark and toolbox:** In molecular-related research, the absence of reliable benchmarks and tools for benchmarking models and unified evaluation metrics leads to a disarray of comparative results across various studies. **C3. Leaderboard:** The lack of an easily

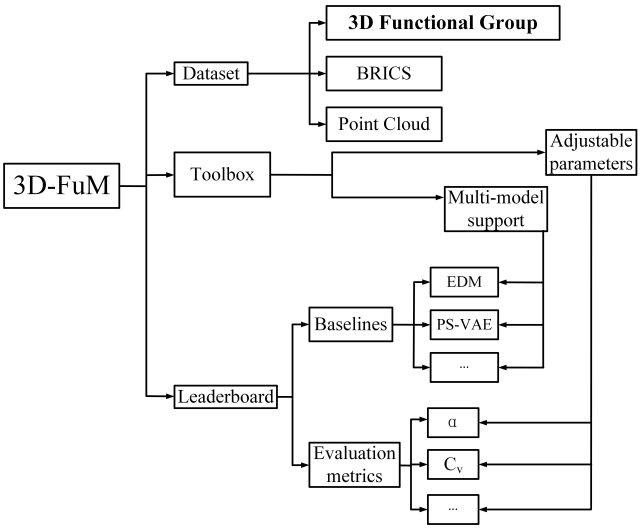


Figure 1: An overview of 3D-FuM. 3D-FUM comprises three integral components. In the Benchmark section, users gain access to downloads for three distinct datasets, each accompanied by visual representations showcasing its unique characteristics within the demo. In the Toolbox section, users are afforded versatile model support with adjustable parameters, enabling a customized approach to model training. Lastly, the Leaderboard section offers a diverse array of benchmark models and a range of evaluation metrics, providing a comprehensive platform for tracking and comparing the latest advancements in molecular research.

accessible leaderboard hinders researchers' utilization of AI models in molecular research. This impedes prompt assessments of model performances, leading to significant time investment in searching for applicable models.

In this paper, we introduce a comprehensive molecular learning benchmark based on functional groups for 3D molecule learning (namely 3D-FUM) based on the QM9 dataset and domain knowledge of functional groups in organic chemistry. By introducing specific functional groups, the 3D-FUM dataset can reduce noisy fragments while incorporating valuable chemical insights. Furthermore, to better enhance the field of AI-assisted molecular research, we summarize mainstream evaluation metrics and tasks. As illustrated in Figure 1, 3D-FUM is a benchmark with a user interface where users can choose diverse datasets and assess model performance across various metrics. Additionally, a leaderboard has been established to track the latest advancements in using AI models for molecular research, providing a comprehensive overview of the state-of-the-art results. This comprehensive strategy aims to bridge the gap between computational approaches and chemical domain knowledge, offering a more holistic and insightful benchmark for advancing the field.

2 System Design

3D-FUM is composed of a benchmark, a toolbox, and a leaderboard for easy model access, advancing the tasks associated with molecular generation and property prediction. Users can flexibly choose their preferred molecule representation (Section 2.1) and adjust the model parameters (Section 2.2), enabling the exploration of diverse models and parameter settings for different molecular data input while analyzing the interpretability with functional groups. Finally, users can report their model performance on the leaderboard for comparison (Section 2.3).

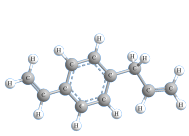
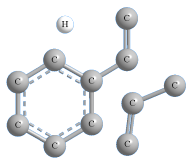
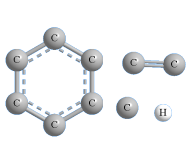
Original Molecule	Trivial	Non-Trivial
		

Table 1: An example comparison of trivial and non-trivial fragments generation. Trivial methods often result in two unrelated fragments for structures with carbon-carbon double bonds. In contrast, non-trivial approaches, *e.g.*, the one employed in 3D-FUM, create universal and interpretable fragments.

2.1 Benchmark

As shown in Table 1, trivial molecular structures, characterized by similar portions lacking a unified label, act as noise in datasets. Recent studies [Patrini *et al.*, 2017; Zhang *et al.*, 2021a; Dai *et al.*, 2021; Ling *et al.*, 2023; Chen *et al.*, 2024] reveal that neural networks generally suffer from overfitting when trained on datasets with noisy labels. In graphs, the noisy information (trivial structures) spreads through the topological structures, making it challenging to learn rational node representations under the impact of noise. Therefore, a model trained on data containing numerous trivial molecular structures is easy to exhibit suboptimal performance. To avoid the impact of trivial molecular structures, we devise 3D-FUM dataset, which considers concrete chemical insights (*i.e.*, functional groups). However, it is notable that the proposal of 3D-FUMdoes not mean obsolescence for the other two fragments generation methods (BRICS and point cloud). In reality, both the BRICS method [Polykovskiy *et al.*, 2020; Zhang *et al.*, 2021b; Zang *et al.*, 2023] and the point cloud method [Schütt *et al.*, 2017; Shi *et al.*, 2021; Vignac *et al.*, 2023] are easy to access, so they are continually being widely used by numerous researchers.

Figure 2 compared the fragments generated by the three approaches based on five typical molecular examples. 3D-FUM, as a novel approach, can effectively amalgamate the strengths of the point cloud method and the BRICS method. The amalgamation is strategically designed to preserve three fundamental structures commonly found in small organic molecules: aromatic rings, double bonds, and triple bonds. Each of these structures has been carefully selected for its possession of distinct and significant chemical properties.

Aromatic rings, known for the presence of aromaticity, introduce a layer of complexity to electron pairs within this structure, diverging from typical structural characteristics. Double and triple bonds, the significant reduction in bond length compared to conventional chemical bonds, coupled with a noteworthy increase in bond energy, makes these structures crucial for their unique chemical behavior. The integra-

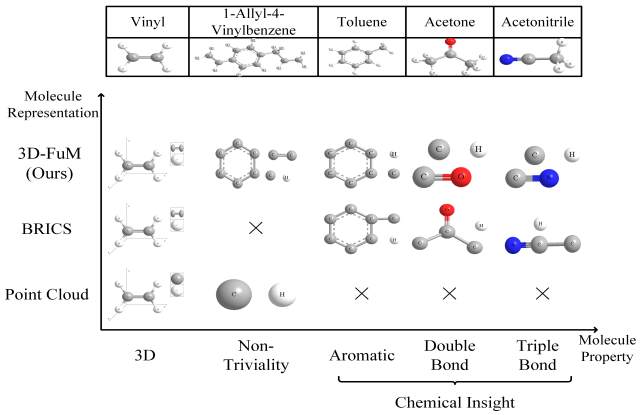


Figure 2: Comparison of different molecule representations. When comparing these three molecule representation methods, we analyze them from three main perspectives and five sub-perspectives. They are assessed based on their support for 3D information, potential data fragmentation issues (particularly notable in the BRICS method), and the introduction of chemical insights. Notably, the point cloud method lacks explicit chemical insights at the level of specific structures such as aromatic rings and double bonds.

tion of these classical structures with their specific chemical insights is aligned with the goal of endowing AI models with a sophisticated comprehension of molecular intricacies.

The remaining atoms, which typically do not exhibit specific chemical properties, are treated using the point cloud methodology. This thoughtful integration allows for a balanced representation, where the unique features of specific structures are meticulously preserved while the more generic characteristics of other atoms are efficiently captured. Finally, The proposed 3D-FUM dataset comprises a total of 133,885 molecules with 360 distinct types of functional groups.

2.2 Toolbox

To facilitate the accommodation of diverse models and evaluation metrics, 3D-FUM includes a generation panel, as illustrated in Figure 3. This panel serves as a central hub where users can craft model training commands tailored to their specific requirements. The design of the generation panel empowers users to flexibly customize various parameters according to their preferences.

Within the panel, users can navigate through an array of selectors and number inputs corresponding to essential elements of the model architecture and training configuration. This user-friendly interface allows for intricate adjustments, ensuring a nuanced control over the model's behavior. The generation panel offers a versatile platform for users to experiment and optimize their models, whether it entails modifying the model's base architecture, adjusting hyperparameters, or fine-tuning specific aspects of the training process.

The interactive panel allows users to intuitively refine their model specifications and observe the real-time impact of adjustments on generated commands. This process empowers users to create models that closely align with their objectives and enhances understanding of the interplay between different components and parameters.

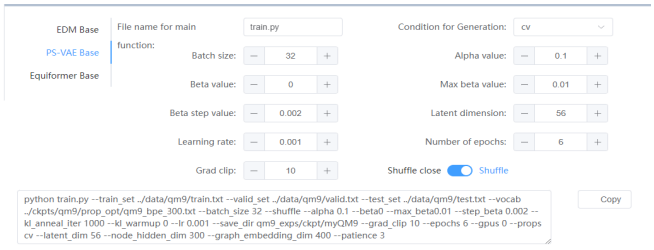


Figure 3: The user interface of 3D-FUM toolbox.

2.3 User Interface and Leaderboard

The versatility of our dataset is demonstrated through the provision of multiple evaluation metrics for assessing tasks related to molecular research assisted by AI models. As a result, we maintain a leaderboard to facilitate the tracking and measurement of performance across different datasets using various models.

3 Potential Impacts

3D-FUM is composed of a benchmark, a toolbox, a user interface, and a leaderboard. Consequently, it serves as a valuable resource for users seeking to augment the chemical insights of their models while steering clear of the confusion induced by trivial molecular fragments. This innovative approach accelerates the progress in utilizing AI models to assist chemistry research. We firmly believe that the proposed 3D-FUM dataset can extend its applicability to a broader spectrum of datasets, further fortifying chemical insights in AI-assisted chemical research.

The proposed benchmark aims to expedite the process for researchers to construct model frameworks swiftly, thereby transforming into an educational toolkit for molecular research, contributing to model advancement, and enhancing accessibility and insights in AI-assisted chemical research for both researchers and practitioners.

4 Conclusion

In this paper, we introduce a benchmark named 3D-FUM, which firstly incorporated 3D geometric information and functional groups for AI-assisted molecular tasks. The proposed 3D-FUM is composed of three components, *i.e.*, benchmark, toolbox, and leaderboard. By introducing 18 state-of-the-art algorithms and 19 evaluation metrics across three molecular tasks, we conduct an in-depth investigation into popular models and evaluation metrics in AI-assisted molecular research. With the easily accessible user interfaces, 3D-FUM can effectively provide insights for researchers and practitioners, thereby fostering innovation and improvement of AI models for molecular research. As an open-source benchmark, 3D-FUM and its experimental results can be accessed through https://3dfunctiongroupmoleculedataset.github.io/3D-FuM/#/Home.

Contribution Statement

The first two authors equally contributed to this paper during their internships at Virginia Tech.

References

[Brandstetter *et al.*, 2022] Johannes Brandstetter, Rob Hesselink, Elise van der Pol, Erik J. Bekkers, and Max Welling. Geometric and physical quantities improve E(3) equivariant message passing. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net, 2022.

[Chen *et al.*, 2024] Jianpeng Chen, Yujing Wang, Ming Zeng, Zongyi Xiang, Bitan Hou, Yunhai Tong, Ole J. Mengshoel, and Yazhou Ren. Customizing graph neural networks using path reweighting. *Information Sciences*, 674:120681, 2024.

[Dai *et al.*, 2021] Enyan Dai, Charu Aggarwal, and Suhang Wang. Nrgnn: Learning a label noise resistant graph neural network on sparsely and noisily labeled graphs. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 227–236, 2021.

[Doshi and Chepuri, 2022] Siddhant Doshi and Sundeep Prabhakar Chepuri. A computational approach to drug repurposing using graph neural networks. *Computers in Biology and Medicine*, 150:105992, 2022.

[Garcia Satorras *et al.*, 2021] Victor Garcia Satorras, Emiel Hoogeboom, Fabian Fuchs, Ingmar Posner, and Max Welling. E (n) equivariant normalizing flows. *Advances in Neural Information Processing Systems*, 34:4181–4192, 2021.

[Huang *et al.*, 2023] Lei Huang, Hengtong Zhang, Tingyang Xu, and Ka-Chun Wong. Mdm: Molecular diffusion model for 3d molecule generation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 5105–5112, 2023.

[Jinsong *et al.*, 2024] Shao Jinsong, Jia Qifeng, Chen Xing, Yajie Hao, and Li Wang. Molecular fragmentation as a crucial step in the ai-based drug development pathway. *Communications Chemistry*, 7(1):20, 2024.

[Kong *et al.*, 2022] Xiangzhe Kong, Wenbing Huang, Zhixing Tan, and Yang Liu. Molecule generation by principal subgraph mining and assembling. *Advances in Neural Information Processing Systems*, 35:2550–2563, 2022.

[Liao and Smidt, 2023] Yi-Lun Liao and Tess E. Smidt. Equiformer: Equivariant graph attention transformer for 3d atomistic graphs. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023.

[Ling *et al.*, 2023] Yawen Ling, Jianpeng Chen, Yazhou Ren, Xiaorong Pu, Jie Xu, Xiaofeng Zhu, and Lifang He. Dual label-guided graph refinement for multi-view graph clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 8791–8798, 2023.

[Morselli Gysi *et al.*, 2021] Deisy Morselli Gysi, Ítalo Do Valle, Marinka Zitnik, Asher Ameli, Xiao Gan, Onur Varol, Susan Dina Ghiassian, JJ Patten, Robert A Davey, Joseph Loscalzo, et al. Network medicine framework for identifying drug-repurposing opportunities for covid-19. *Proceedings of the National Academy of Sciences*, 118(19):e2025581118, 2021.

[Nakao *et al.*, 2023] Atsuyuki Nakao, Yu Harabuchi, Satoshi Maeda, and Koji Tsuda. Exploring the quantum chemical energy landscape with gnn-guided artificial force. *Journal of Chemical Theory and Computation*, 19(3):713–717, 2023.

[Patrini *et al.*, 2017] Giorgio Patrini, Alessandro Rozza, Aditya Krishna Menon, Richard Nock, and Lizhen Qu. Making deep neural networks robust to label noise: A loss correction approach. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1944–1952, 2017.

[Patronov *et al.*, 2022] Atanas Patronov, Kostas Papadopoulos, and Ola Engkvist. Has artificial intelligence impacted drug discovery? *Artificial Intelligence in Drug Design*, pages 153–176, 2022.

[Peng *et al.*, 2022] Xingang Peng, Shitong Luo, Jiaqi Guan, Qi Xie, Jian Peng, and Jianzhu Ma. Pocket2mol: Efficient molecular sampling based on 3d protein pockets. In *International Conference on Machine Learning*, pages 17644–17655. PMLR, 2022.

[Polykovskiy *et al.*, 2020] Daniil Polykovskiy, Alexander Zhebrak, Benjamin Sanchez-Lengeling, Sergey Golovanov, Oktai Tatanov, Stanislav Belyaev, Rauf Kurbanov, Aleksey Artamonov, Vladimir Aladinskiy, Mark Veselov, et al. Molecular sets (moses): a benchmarking platform for molecular generation models. *Frontiers in pharmacology*, 11:565644, 2020.

[Ruddigkeit *et al.*, 2012] Lars Ruddigkeit, Ruud Van Deursen, Lorenz C Blum, and Jean-Louis Reymond. Enumeration of 166 billion organic small molecules in the chemical universe database gdb-17. *Journal of chemical information and modeling*, 52(11):2864–2875, 2012.

[Schütt *et al.*, 2017] Kristof Schütt, Pieter-Jan Kindermans, Huziel Enoc Sauceda Felix, Stefan Chmiela, Alexandre Tkatchenko, and Klaus-Robert Müller. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. *Advances in neural information processing systems*, 30, 2017.

[Shi *et al.*, 2021] Chence Shi, Shitong Luo, Minkai Xu, and Jian Tang. Learning gradient fields for molecular conformation generation. In *International conference on machine learning*, pages 9558–9568. PMLR, 2021.

[Sterling and Irwin, 2015] Teague Sterling and John J Irwin. Zinc 15–ligand discovery for everyone. *Journal of chemical information and modeling*, 55(11):2324–2337, 2015.

[Thölke and Fabritiis, 2022] Philipp Thölke and Gianni De Fabritiis. Equivariant transformers for neural network based molecular potentials. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net, 2022.

[Unsleber *et al.*, 2022] Jan P Unsleber, Stephanie A Grimmel, and Markus Reiher. Chemoton 2.0: Autonomous exploration of chemical reaction networks. *Journal of Chemical Theory and Computation*, 18(9):5393–5409, 2022.

[Vignac *et al.*, 2023] Clément Vignac, Igor Krawczuk, Antoine Siraudin, Bohan Wang, Volkan Cevher, and Pascal Frossard. Digress: Discrete denoising diffusion for graph generation. In *Proceedings of the 11th International Conference on Learning Representations*, 2023.

[Wang *et al.*, 2022] Yuyang Wang, Rishikesh Magar, Chen Liang, and Amir Barati Farimani. Improving molecular contrastive learning via faulty negative mitigation and decomposed fragment contrast. *Journal of Chemical Information and Modeling*, 62(11):2713–2725, 2022.

[Xu *et al.*, 2024] Dapeng Xu, Wei Tang, Feng Li, Ke Zhou, Junlin Wu, Alexander Fuqua, Yu-ting Chen, Dong Ju Lee, Jianting Qin, Andrea R Tao, et al. Insights on na3ps4 solid-state electrolyte dry films: Interfacial stability and dry room compatibility. *ACS Energy Letters*, 9:2293–2302, 2024.

[Zang *et al.*, 2023] Xuan Zang, Xianbing Zhao, and Buzhou Tang. Hierarchical molecular graph self-supervised learning for property prediction. *Communications Chemistry*, 6(1):34, 2023.

[Zhang *et al.*, 2021a] Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals. Understanding deep learning (still) requires rethinking generalization. *Communications of the ACM*, 64(3):107–115, 2021.

[Zhang *et al.*, 2021b] Zaixi Zhang, Qi Liu, Hao Wang, Chengqiang Lu, and Chee-Kong Lee. Motif-based graph self-supervised learning for molecular property prediction. *Advances in Neural Information Processing Systems*, 34:15870–15882, 2021.