DO-CoLM: Dynamic 3D Conformation Relationships Capture with Self-Adaptive Ordering Molecular Relational Modeling in Language Models

Zhuo Chen[1,2], **Jiahui Zhang**[1,2], **Sihan Wang**[1,2], **Hongxin Xiang**[3], **Jianmin Wang**[4], **Wenjie Du**[1,2,*] and **Yang Wang**[1,2,*]
[1]University of Science and Technology of China (USTC), Hefei, China
[2]Suzhou Institute for Advanced Research, USTC, Suzhou, China
[3]Hunan University, Hunan, China
[4]Yonsei University, Seoul, Korea
{czchenzhuo, kongping, sihanwang}@mail.ustc.edu.cn, xianghx@hnu.edu.cn, jmwang113@hotmail.com, duwenjie@mail.ustc.edu.cn, angyan@ustc.edu.cn

Abstract

Molecular Relational Learning (MRL) aims to understand interactions between molecular pairs, playing a critical role in advancing biochemical research. Recently, Large Language Models (LLMs), with their extensive knowledge bases and advanced reasoning capabilities, have emerged as powerful tools for MRL. However, existing LLMs, which primarily rely on SMILES strings and molecular graphs, face two major challenges. Firstly, they struggle to capture molecular stereochemistry and dynamics, as molecules possess multiple 3D conformations with varying reactivity and dynamic transformation relationships that are essential for accurately predicting molecular interactions but cannot be effectively represented by 1D SMILES or 2D molecular graphs. Secondly, these models do not consider the autoregressive nature of LLMs, overlooking the impact of input order on model performance. To address these issues, we propose a **D**ynamic relationship capture and self-adaptive **O**rdering 3D molecular **Co**nformation **LM** for MRL, termed as **DO-CoLM**. By introducing modules to dynamically model intra-molecular and inter-molecular conformational relationships and adaptively adjust the molecular modality input order, DO-CoLM achieves superior performance on 12 cross-domain datasets.

1 Introduction

Molecular Relationship Learning (MRL) aims to understand interactions between molecular pairs, a critical topic with applications in drug discovery and materials science. For example, drug-drug interactions (DDIs) are vital in pharmacology and drug development, while solute-solvent interactions (SSIs) play a key role in solution chemistry and chemical process design. However, validating these interactions experimentally is time-intensive and costly.

Large Language Models (LLMs), with their vast knowledge base and advanced reasoning capabilities, offer a promising solution [Taylor *et al.*, 2022a]. Models like ReactionT5 [Sagawa and Kojima, 2023] and MolTC [Fang *et al.*, 2024] extend LLMs to MRL by leveraging multimodal data, including molecular graphs (2D), chemical properties, and SMILES (1D), for accurate interaction predictions. Despite their potential, these models face unexplored challenges:

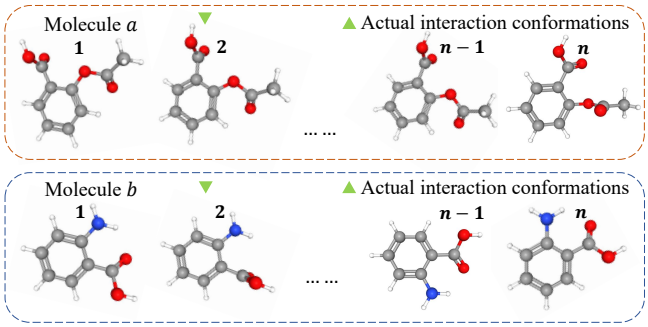

Figure 1: A molecule can have multiple conformations with mutual interconversion relationships [Skjærven *et al.*, 2011] and the actual molecular interactions are determined by the specific conformations [Mortier *et al.*, 2015].

Lack of the molecular stereochemistry and dynamics. Existing LLMs primarily rely on SMILES strings and molecular graphs for predicting molecular interactions [Fang *et al.*, 2024; Sagawa and Kojima, 2023], but these 1D and 2D representations fail to capture the 3D information of molecules [Liu *et al.*, 2023] and the dynamic nature of molecular conformations during chemical reactions [Morris *et al.*, 2019]. Essential properties such as stereochemistry (including chirality and stereoisomerism) and molecular dynamics [Andrade *et al.*, 2009] are crucial for understanding molecular interactions and predicting reaction mechanisms [Alonso *et al.*, 2006]. On the one hand, as shown in Figure 1, a molecule has multiple conformations that continuously interconvert [Skjærven *et al.*, 2011]; on the other hand, the interaction between two molecules occurs through the interaction of their specific conformations, determined by their 3D structures [Mortier *et al.*, 2015], such as the binding of small-molecule and target proteins [Wu *et al.*, 2022].

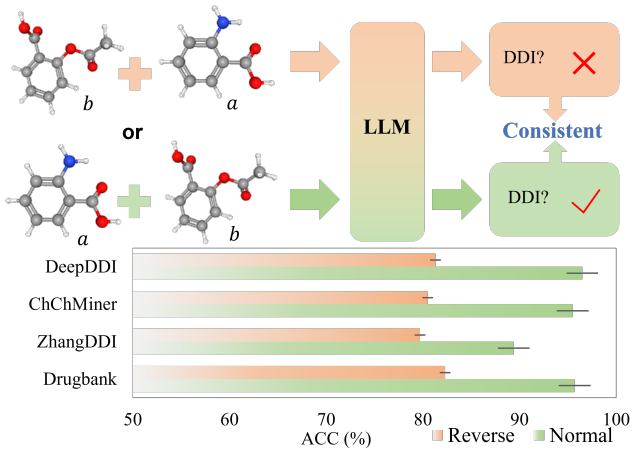

Figure 2: Swapping the molecular input order leads to different predictions and MolTC's results for DDI tasks (Normal) compared to when the SMILES order is swapped (Reverse).

Ignorance of the influence of input order. The performance of LLM is naturally influenced by the input sequence [Berglund *et al.*, 2023], which can have catastrophic consequences in certain MRL tasks. For instance, in DDI tasks, the SMILES order of input molecules should exhibit symmetry, meaning that swapping the SMILES input order should not affect the model's prediction results. However, as shown in Figure 2, swapping the input SMILES order causes the model to produce different predictions, leading to a decrease in MolTC's performance, with accuracy dropping by 11.58% on the Drugbank dataset, 9.24% on the ZhangDDI dataset, 12.81% on the ChChMiner dataset, and 9.08% on the DeepDDI dataset. Recent studies have highlighted this issue [Berglund *et al.*, 2023], which is attributed to the autoregressive training paradigm typically employed by LLMs, leading the model to treat input order as a feature. This can be helpful in some tasks, such as SSI tasks, but it should not affect model performance in symmetric tasks. Notably, in addition to the overall SMILES order, the internal order of molecular tokens also could affect the model's performance.

Considering the above issues, we propose **DO-CoLM**: a **D**ynamic relationship capture and self-adaptive **O**rdering 3D molecular **Co**nformation **LM** for MRL. Specifically, we design the **Dynamic Conformational Relationship Capture Module (DCRCM)**, to address the lack of molecular stereochemistry and dynamics. This module models both intramolecular conformation transformation and intermolecular conformation interaction relationships using heterogeneous graphs based on conformational information. Additionally, we introduce the **Adaptive Order Adjustment Module (AOAM)**, which ensures the model's robustness to perturbations in symmetric input orders and allows it to adaptively find the optimal sequence for asymmetric input orders from the model's perspective.

The main contributions of this paper could be summarized as follows:

- Given the current limitations of LLM in adequately considering molecular 3D information and dynamics, and the neglect of input order in MRL tasks, we propose DO-CoLM, a multimodal framework that adaptively integrates conformational information and accommodates molecular order.
- DO-CoLM enables dynamic learning of relationships between molecular conformations by constructing a heterogeneous graph and adaptively selecting the optimal conformation order, thereby achieving comprehensive molecular information capture and tolerance to molecular encoder tokens variations.
- The superiority of DO-CoLM is empirically validated through extensive experiments across 12 datasets from various domains, including DDI, CSI, and SSI tasks. These results demonstrate that our approach significantly outperforms existing methods.

2 Related Work

2.1 Molecular Relational Learning

Early research primarily relied on graph neural networks (GNNs) to construct predictive frameworks for MRL [Du *et al.*, ; Fu *et al.*, 2020; Sizhe Liu *et al.*, 2024]. For instance, Nyamabo et al. proposed a substructure-substructure interaction framework [Zhong *et al.*, 2024], which utilizes graph attention network (GAT) layers for substructure extraction and a co-attention layer to model interactions between different substructures. To further capture molecular interactions, Lee et al. [Lee *et al.*, 2023] introduced the Conditional Graph Information Bottleneck (CGIB) model, inspired by the information bottleneck theory. However, their models lack prior knowledge and fail to leverage the advantages of LLMs.

2.2 LLMs in the Molecular Domain

LLMs have been widely applied in the molecular domain [Zheng *et al.*, 2025; Jablonka *et al.*, 2024]. In 1D, model like MolT5 [Edwards *et al.*, 2022] tokenizes SMILES strings; in 2D, methods such as Text2Mol [Edwards *et al.*, 2021], MolCA [Liu *et al.*, 2023], and DrugChat [Liang *et al.*, 2023] combine molecular graphs with LLMs; and in 3D, MolLM [Tang *et al.*, 2024] and 3D-MoLM [Li *et al.*, 2024] capture spatial features using attention and 3D encoders. Multimodal models like ReactionT5 [Sagawa and Kojima, 2023] and MolTC [Fang *et al.*, 2024] integrate 1D, 2D, and chemical property data for prediction. However, most existing work focuses on single molecules or 1D and 2D data, overlooking the role of 3D conformations in MRL.

2.3 Input Order

Previous studies show that input sequence order significantly affects model performance. For instance, the improved Seq2Seq model [Vinyals *et al.*, 2015] reveals that even unordered data can benefit from an optimal input order. PointNet [Qi *et al.*, 2017] mitigates order sensitivity using symmetry functions. Work on the reversal curse [Berglund *et al.*, 2023] and inter-modal order [Tan *et al.*, 2024] further highlights order's impact in LMs. However, most focus on single-modal inputs or coarse inter-modal order, neglecting fine-grained token-level order across modalities.

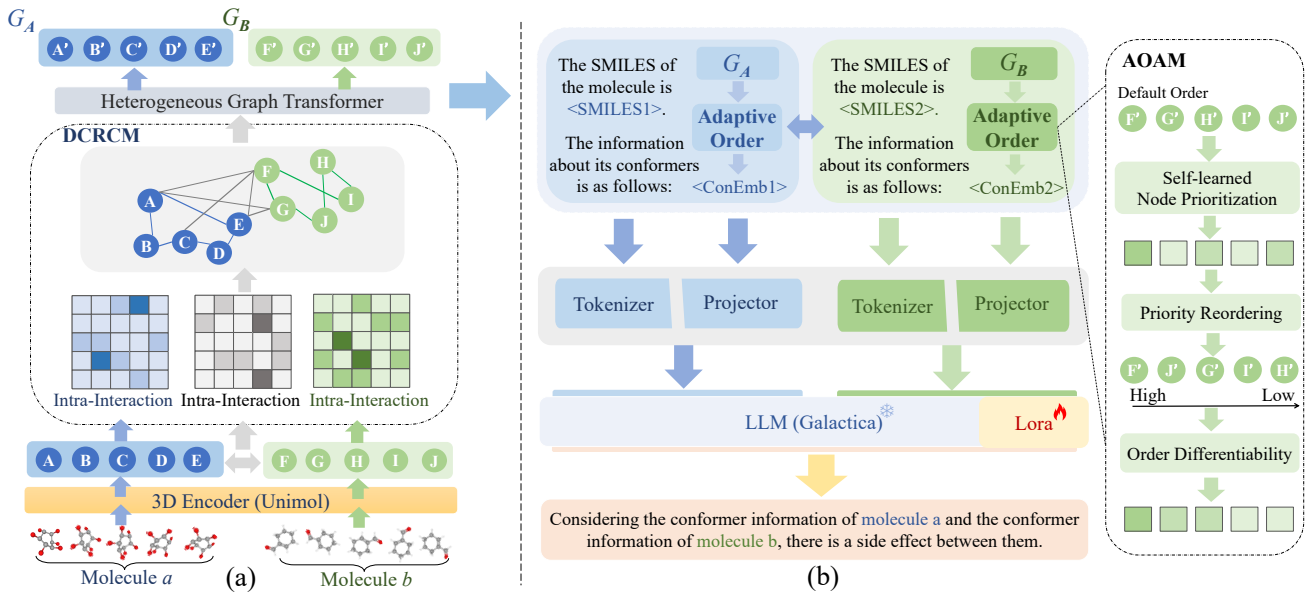

Figure 3: The architecture of DO-CoLM. (a) is to encode 3D molecular conformations and dynamically model their relationships. (b) is designed to adaptively adjust the input order and perform modality alignment to generate the output.

3 Our Proposed Model: DO-CoLM

In this section, we present the architecture of DO-CoLM. As shown in Figure 3, DO-CoLM consists of five key steps. First, the 3D Encoder is used to obtain molecular conformer feature embeddings. Next, the DCRCM dynamically models the intramolecular and intermolecular conformational relationships. Then, the AOAM adaptively adjusts the input order based on the asymmetry of the input data to optimize the model's performance. Afterward, the Alignment module ensures seamless integration by aligning molecular modality data with textual data. Finally, the LLM generates the model's output based on the processed information.

3.1 Conformer Encoder

Let $\mathcal{C}^a = [\mathcal{C}_1^a, \ldots, \mathcal{C}_{n_a}^a]$ and $\mathcal{C}^b = [\mathcal{C}_1^b, \ldots, \mathcal{C}_{n_b}^b]$ represent the conformer pairs for molecule a and molecule b, where $\mathcal{C}_i^a$ and $\mathcal{C}_i^b$ are the i-th conformations, n_a and n_b are the number of conformations for each molecule. We leverage Uni-Mol [Zhou *et al.*, 2023], a powerful 3D molecular feature encoder f, along with attention pooling p, to capture the embeddings of conformer pairs.

$$\begin{aligned} \mathbf{V}^a &= [v_1^a, v_2^a, \ldots, v_{n_a}^a], \text{where } v_i^a = p(f(\mathcal{C}_i^a)), \\ \mathbf{V}^b &= [v_1^b, v_2^b, \ldots, v_{n_b}^b], \text{where } v_i^b = p(f(\mathcal{C}_i^b)), \end{aligned} \tag{1}$$

where $\mathbf{V}^a$ and $\mathbf{V}^b$ represent the sets of conformation feature embeddings for the two molecules.

3.2 Dynamic Conformational Relationship Capture Module

Due to the dynamic nature of molecular conformations, accurately modeling the transitions and interactions of molecular conformations requires complex simulations and experiments. Therefore, a conformation transition relationship graph based on predefined structures is not feasible. To address this, we propose the DCRCM, which allows the model to dynamically model intramolecular transitions and intermolecular interactions.

We define this relationship as the Heterogeneous Molecular Conformation Interaction Graph (*HMCG*). The *HMCG* is represented by $\mathcal{G}_{\text{HMCG}} = (\mathbf{V}, \mathcal{E}, \mathbf{A}, \mathcal{T}, \mathcal{R})$, where $\mathbf{V} = \mathbf{V}^a \cup \mathbf{V}^b$ is the set of conformer nodes, $\mathcal{E}$ is the edge set with three types: intra-molecular edges within a, intra-molecular edges within b, and inter-molecular edges between a and b. $\mathbf{A}$ is the adjacency matrix representing node connectivity, $\mathcal{T}$ denotes node types, and $\mathcal{R}$ denotes edge types.

To model the intra-molecular conformation transformation edges $\mathcal{E}_{\text{intra}}$, we first obtain $\mathbf{A}_{\text{intra}}$ through the following algorithm, taking molecule a as an example.

$$\begin{aligned} \mathbf{A}_{\text{intra}}^{\text{a}} &= \sigma_1(\sigma_2(\mathbf{U}_1^a(\mathbf{U}_2^a)^T) - \sigma_2(\mathbf{U}_2^a(\mathbf{U}_1^a)^T)) \\ &= \sigma_1(\sigma_2(\theta_1\mathbf{V}^a(\theta_2\mathbf{V}^a)^T) - \sigma_2(\theta_2\mathbf{V}^a(\theta_1\mathbf{V}^a)^T)), \end{aligned} \tag{2}$$

where $\mathbf{U}_1^a$ represents the representation learning for the intra-molecular conformation as the source node, and $\mathbf{U}_2^a$ as the target node, θ_1 and θ_2 represent the learnable vectors for learning the source and target nodes, respectively, and σ_1 and σ_2 are the activation functions. Such an operation ensures that the generated graph is a directed and asymmetric graph.

To reduce the complexity of the graph, we perform a sparsification operation on it.

$$\begin{aligned} \text{idx}, \text{idy} = \text{argtopk}(A_{\text{intra}}^a[:,:]) \quad \text{idx} \neq \text{idy}, \\ \mathbf{A}_{\text{intra}}^{\text{a}}[-\text{idx}, -\text{idy}] = 0, \end{aligned} \tag{3}$$

where *argtopk*$(\cdot)$ returns the indices of the most likely conformation nodes for transformation. The weights of other edges are set to 0. For the conformation transformation edges in molecule b, we adopt the same approach to obtain $\mathbf{A}_{\text{intra}}^{\text{b}}$.

Since the inter-molecular conformation interaction is symmetric, we use a bidirectional graph to construct the inter-molecular conformation interaction edges $\mathcal{E}_{\text{inter}}$. The algorithm to obtain the adjacency matrix $\mathbf{A}_{\text{inter}}$ representing the edges is as follows:

$$\begin{aligned}\mathbf{A}_{\text{inter}} &= \sigma_1(\sigma_2(\mathbf{U}^a(\mathbf{U}^b)^T)) \\ &= \sigma_1(\sigma_2(\Theta_a\mathbf{V}^a(\Theta_b\mathbf{V}^b)^T)),\end{aligned} \tag{4}$$

where $\mathbf{U}^a$ and $\mathbf{U}^b$ represent the relationship representations of the source node and target node for the inter-molecular conformation interaction, respectively. Here, we use molecule a as the source node and molecule b as the target node, and Θ_a and Θ_b are learnable vectors for learning the relationship between the source node and the target node. After applying the same sparsification operation as in Equation (3), we obtain $\mathbf{A}_{\text{inter}}$.

After obtaining $\mathcal{G}_{\text{HMCG}}$ through the above operation, we aggregate the node information on the graph using $\mathbf{Attention}(\cdot)$ and $\mathbf{Message}(\cdot)$. The specific computation formula is as follows:

$$\begin{aligned}&\mathbf{Attention}(u,e,v) \\ &= \underset{\forall u\in\mathcal{N}(v)}{\text{Softmax}}(\underset{i\in[1,n_h]}{\Vert} F_1^{\tau(u)}(h_u^{(l-1)})\mathbf{W}_1^{\rho(e)}F_2^{\tau(v)}(h_v^{(l-1)})), \\ &\mathbf{Message}(u,e,v) = \underset{i\in[1,n_h]}{\Vert} F_3^{\tau(u)}(h_u^{(l-1)})\mathbf{W}_2^{\rho(e)},\end{aligned} \tag{5}$$

where $\mathcal{N}(v)$ represents the set of nodes connected to node v, $F^{\tau(u)}$ denotes a fully connected layer with parameters $\{\mathbf{W}^{\tau(v)}, \mathbf{b}^{\tau(v)}\}$, $\rho(e)$ represents the type of edge, $\Vert$ represents the concatenation operation, and n_h denotes the number of attention heads. Then update the node information.

$$\begin{aligned}\widetilde{h}_v^{(l)} &= \underset{\forall u\in\mathcal{N}(v)}{\oplus}(\mathbf{Attention}(u,e,v)\cdot\mathbf{Message}(u,e,v)), \\ h_v^{(l)} &= \mathbf{W}_1^{\tau(v)}\cdot(\sigma(\widetilde{h}_v^{(l)})) + \mathbf{b}_1^{\tau(v)} + h_v^{(l-1)},\end{aligned} \tag{6}$$

where $\oplus$ represents the aggregation operation, $h_v^{(l)}$ is an aggregated feature representation of the conformation.

3.3 Adaptive Order Adjustment Module

Due to the sensitivity of large models to input order, different input orders can significantly impact model performance. In our model, in addition to **the overall SMILES order**, there is also **the internal order of nodes** which also impacts the performance of the model. For these two types of order, we have designed different strategies to enable the model to achieve optimal performance.

The overall order of SMILES. In the DDI task, there are only two possible input molecular sequences: molecular a followed by molecular b, and molecular b followed by molecular a. These two sequences are symmetric and should not affect the model's experimental results. Therefore, we can use a simple operation to enable the model to understand these two different sequences, ensuring that the overall order of the input molecules does not influence the model's performance. This can be achieved by randomly shuffling the input sequence, allowing the model to focus solely on the molecular information itself without considering the positional information it carries.

The internal order of nodes. $\mathcal{G}_{\text{HMCG}}$ comprises n conformer nodes in total, corresponding to $n!$ possible sequences. Achieving order invariance by enabling the model to simultaneously process all these sequences is nearly impossible. Given that different sequences impact model performance to varying degrees, we propose that an optimal sequence exists from the model's perspective, one that cannot be captured through prior knowledge-based ordering. To address this, we hypothesize that each conformer has an inherent priority and allow the model to autonomously determine these priorities and dynamically adjust them based on different situations, thereby uncovering the optimal sequence in the model's view.

We use a Learnable Sorting Network $\mathcal{K}(\cdot)$ with a learnable parameter for each conformation to allow the model to prioritize the node h_i. Then we obtain the sorted order S_b while the initial order is S_a.

$$S_b = \mathbf{argsort}(\mathcal{K}(h_1),\mathcal{K}(h_2),\ldots,\mathcal{K}(h_n)), \tag{7}$$

where n is the number of the conformers. Due to the non-differentiability of argsort, we employed the Sinkhorn Operator [Cuturi, 2013] to dynamically update the sorting process. Initially, we obtain the transport matrix $\mathbf{Q}$.

$$\mathbf{Q}_{ij} = exp(-\frac{|S_a^i - S_b^j|}{\epsilon}), \tag{8}$$

where S_a^i and S_b^j represent the position values of conformers i and j, ϵ is a control factor. In the iterations, $\mathbf{Q}$ is updated as:

$$\mathbf{v} = \frac{1}{\mathbf{Q}^\top\mathbf{u}K}, \quad \mathbf{u} = \frac{1}{\mathbf{Q}\mathbf{v}K}, \tag{9}$$

where K is the size of $\mathbf{Q}$, which means the number of rows and columns, $u = \mathbf{1}_K$. The iterations stop when $\Delta(\mathbf{v}\mathbf{Q}^\top\mathbf{u}, \mathbf{1}_K/K) < \eta$. The max iteration number is depend on ϵ: typically, the smaller ϵ, the larger the number is needed to ensure that $\mathbf{v}\mathbf{Q}^\top\mathbf{u}$ is close to $\mathbf{1}_K/K$. Then we obtain the conformer representation $\mathbf{Z}$ by normalizing Q and multiplying with the conformation $\mathbf{H}$.

$$\mathbf{Z} = diag(\mathbf{u})\mathbf{Q}diag(\mathbf{v})\mathbf{H}, \tag{10}$$

where $diag(\cdot)$ means diagonal operation.

3.4 Alignment

After acquiring the multi-conformer representations $\mathbf{Z_a}$ and $\mathbf{Z_b}$ of molecule a and molecule b, the next step is to map them into the backbone LLM's hidden space using projectors f_{proj1}, f_{proj2}, where f_{proj1} and f_{proj2} share the same parameters. These projectors act as critical connectors, translating $\mathbf{Z_a}$ and $\mathbf{Z_b}$ into LLM-comprehensible encodings $\mathbf{M}_a$ and $\mathbf{M}_b$.

We instantiate f_{proj1} and f_{proj2} using trainable projection matrices, which share the same dimensionality as the word embedding space in the language model. More formally, the encodings can be expressed as:

$$\begin{aligned}\mathbf{M}_a &= [\mathbf{m}_1^a, \mathbf{m}_2^a, \ldots, \mathbf{m}_d^a] = f_{\text{proj1}}(\mathbf{Z_a}), \\ \mathbf{M}_b &= [\mathbf{m}_1^b, \mathbf{m}_2^b, \ldots, \mathbf{m}_d^b] = f_{\text{proj2}}(\mathbf{Z_b}),\end{aligned} \tag{11}$$

where d denotes the feature dimensionality of the LLM and m_i represents the embedded representation of a single aligned conformer.

3.5 Backbone LLM

DO-CoLM uses Galactica$_{1.3B}$ [Taylor *et al.*, 2022b], a decoder-only transformer based on the OPT framework, trained on scientific literature, as its backbone. It excels in interpreting molecular sequences. DO-CoLM leverages Galactica's reasoning to analyze interactions between two molecular sets and their information tokens. The prompt sequence $\mathbf{X}$ is as follows:

$$\mathbf{X} = \{\mathbf{P}, \mathbf{M}_a, \mathbf{M}_b\} = [\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n] \quad \text{s.t. } \mathbf{P} \sim \mathcal{P}, \quad (12)$$

where n denotes the total integrated input length, $\mathbf{P}$ represents the task-specific prompt, and $\mathcal{P}$ refers to a collection of manually designed prompts, each specifically crafted for the molecular interaction task $\mathbf{r}$. The details of the prompt will be discussed in Section 3.6. The generation process employs a causal mask to produce a response that encapsulates the key interactive properties, with a length of L:

$$\hat{\mathbf{X}} = [\hat{\boldsymbol{x}}_1, \hat{\boldsymbol{x}}_2, ..., \hat{\boldsymbol{x}}_L], \quad (13)$$

The training objective is to predict the target response from the input prompt $\mathbf{X}$. Specifically, the output for the i-th token, represented as $\hat{\boldsymbol{x}}_i$, is determined based on its preceding tokens as follows for $t \in (1, L)$:

$$p\left(\hat{\mathbf{X}}_{[1:t]}|\mathbf{X}\right) = \prod_{i=1}^{t} p\left(\hat{\boldsymbol{x}}_i|\mathbf{X}, \hat{\mathbf{X}}_{[1:i-1]}\right), \quad (14)$$

3.6 Prompt

For different tasks related to molecular interactions, we have designed distinct prompts. Here, *SMILES1* and *SMILES2* represents the SMILES notation, *ConEmb1* and *ConEmb2* represents multiple conformer tokens $\mathbf{M_a}$ and $\mathbf{M_b}$. For DDI tasks, the goal is to enable the LLM to leverage the conformational information of the two input molecules. By doing so, the LLM actively determines and provides an answer as to whether there is an interaction between the two molecules, drawing on their respective conformations to make judgment.

Prompt for DDI Tasks

Input Prompt	<*SMILES1*>, the information about its conformers is <*ConEmb1*>. <*SMILES2*>, the information about its conformers is <*ConEmb2*>. Do they have any side effects?
Target Response	Considering the conformer information of *molecule1* and the conformer information of *molecule2*, there is a side effect between molecule A and molecule B.

For SSI tasks, where LLMs are less adept, we design a prompt that encourages the model to consider molecular conformations during inference. Instead of a direct answer, the model predicts a range of possible interaction outcomes based on the conformations. It then synthesizes the data to estimate a specific value within that range.

Prompt for SSI Tasks

Input Prompt	<*SMILES1*>, the information about its conformers is <*ConEmb1*>. <*SMILES2*>, the information about its conformers is <*ConEmb2*>. What is the solvation Gibbs free energy of this pair of molecules?
Target Response	Considering the conformation of *molecule1* and *molecule2*, the solvation Gibbs free energy of these two molecules is above `2.5` and below `3.0`, so the accurate value is `2.768`.

4 Eeperimental Results

4.1 Experimental Setting

We evaluate DO-CoLM on well-established downstream molecule interaction tasks involving qualitative and quantitative analysis. Here we provide a overview of our experimental setup. Detailed descriptions are presented in the Appendix C.

Datasets. We employ 12 datasets across various domains such as DDI, SSI, and CSI. Specifically, we collect Drugbank (Version 5.0.3), ZhangDDI, ChChMiner, DeepDDI, TWOSIDES, Chromophore, MNSol, CompSol, Abraham, CombiSolv, FreeSolv, and CombiSolv-QM. Here, we process each molecule using RDKit to obtaining several conformations for each molecule. The hyperparameter experiment on the number of different conformations is shown in the Appendix E. Here, 10 conformations are used in subsequent experiments.

Baselines. For a comprehensive evaluation, we conduct various baseline methods encompassing distinct categories such as methods based on: GNNs, DL models other than GNN, and LLMs. Specifically, for DDI task, we employ IGIB-ISE [Zhang *et al.*, 2025], MHCADDI [Deac *et al.*, 2019], DeepDDI [Ryu *et al.*, 2018], SSI-DDI, CGIB, CMRL, MDF-SA-DDI [Lin *et al.*, 2022], DSN-DDI [Li *et al.*, 2023], MolTC [Fang *et al.*, 2024] as the baseline. For SSI and CSI tasks, we utilize D-MPNN [Vermeire and Green, 2021], SolvBert [Yu *et al.*, 2023], SMD [Meng *et al.*, 2023], CGIB, MMGNN [Du *et al.*, 2024], GEM [Fang *et al.*, 2022], GOVER [Rong *et al.*, 2020], Uni-Mol [Zhou *et al.*, 2023] as the baseline. Furthermore, all downstream tasks adopt LLM-based methods, such as Galactica, Chem T5 [Christofidellis *et al.*, 2023], MolT5, MolCA [Liu *et al.*, 2023] and MolTC as the baseline.

4.2 Evaluation Metrics

We employ prediction Accuracy (%) and AUC-ROC (Area Under the Receiver Operating Characteristic curve) as comparative metrics, while for quantitative tasks, MAE (Mean Absolute Error) and RMSE (Root Mean Square Error) are utilized as the metrics.

4.3 Experimental Results and Analysis

Due to the limitations of certain models on some datasets, we only present a subset of the results here. For detailed results on other datasets, please refer to Appendix D.

Baseline Model		Drugbank		ZhangDDI		ChChMiner		DeepDDI	
		Accuracy	AUC-ROC	Accuracy	AUC-ROC	Accuracy	AUC-ROC	Accuracy	AUC-ROC
GNN Based	IGIB-ISE	$95.16_{\pm 0.37}$	$98.83_{\pm 0.56}$	$88.80_{\pm 0.35}$	$94.75_{\pm 0.41}$	$94.97_{\pm 0.38}$	$97.84_{\pm 0.32}$	$96.24_{\pm 0.35}$	$98.52_{\pm 0.31}$
	SSI-DDI	$94.12_{\pm 0.32}$	$98.28_{\pm 0.29}$	$87.02_{\pm 0.32}$	$93.66_{\pm 0.38}$	$93.16_{\pm 0.21}$	$97.91_{\pm 0.12}$	$94.97_{\pm 0.23}$	$98.41_{\pm 0.30}$
	DSN-DDI	$94.95_{\pm 0.12}$	$99.01^{*}_{\pm 0.11}$	$87.55_{\pm 0.11}$	$94.83_{\pm 0.28}$	$84.20_{\pm 0.18}$	$94.05_{\pm 0.26}$	$95.44_{\pm 0.28}$	$98.02_{\pm 0.15}$
	CMRL	$94.93_{\pm 0.12}$	$98.76_{\pm 0.11}$	$87.68_{\pm 0.32}$	$94.58_{\pm 0.21}$	$94.25_{\pm 0.26}$	$98.34_{\pm 0.12}$	$96.27_{\pm 0.34}$	$98.97^{*}_{\pm 0.29}$
	CGIB	$94.67_{\pm 0.34}$	$98.53_{\pm 0.24}$	$87.82_{\pm 0.73}$	$93.98_{\pm 0.61}$	$94.35_{\pm 0.36}$	$98.45^{*}_{\pm 0.32}$	$96.13_{\pm 0.49}$	$97.98_{\pm 0.64}$
ML Based	DeepDDI	$93.05_{\pm 0.25}$	$98.36_{\pm 0.54}$	$83.60_{\pm 0.49}$	$91.23_{\pm 0.58}$	$90.44_{\pm 0.62}$	$95.83_{\pm 0.27}$	$92.49_{\pm 0.38}$	$98.12_{\pm 0.41}$
	MHCADDI	$79.30_{\pm 0.80}$	$86.23_{\pm 0.45}$	$77.66_{\pm 0.49}$	$87.04_{\pm 0.68}$	$84.56_{\pm 0.53}$	$90.01_{\pm 0.82}$	$87.71_{\pm 0.77}$	$88.64_{\pm 0.73}$
	MDF-SA-DDI	$93.96_{\pm 0.32}$	$97.59_{\pm 0.29}$	$86.69_{\pm 0.25}$	$94.13_{\pm 0.33}$	$93.74_{\pm 0.21}$	$98.12_{\pm 0.19}$	$94.89_{\pm 0.31}$	$97.74_{\pm 0.34}$
LLM Based	Galactica	$79.28_{\pm 0.34}$	$86.25_{\pm 0.34}$	$67.30_{\pm 0.56}$	$79.02_{\pm 0.59}$	$74.51_{\pm 0.43}$	$84.01_{\pm 0.67}$	$71.20_{\pm 0.42}$	$79.08_{\pm 0.40}$
	Chem T5	$85.93_{\pm 0.31}$	$92.07_{\pm 0.37}$	$72.44_{\pm 0.39}$	$89.51_{\pm 0.32}$	$80.99_{\pm 0.52}$	$85.35_{\pm 0.48}$	$75.78_{\pm 0.63}$	$84.52_{\pm 0.44}$
	MolCA	$87.85_{\pm 0.49}$	$94.01_{\pm 0.37}$	$68.51_{\pm 0.62}$	$88.33_{\pm 0.52}$	$90.17_{\pm 0.45}$	$93.04_{\pm 0.61}$	$82.97_{\pm 0.57}$	$89.03_{\pm 0.74}$
	MolT5	$89.59_{\pm 0.37}$	$93.18_{\pm 0.37}$	$76.89_{\pm 0.41}$	$87.62_{\pm 0.53}$	$80.97_{\pm 0.37}$	$90.18_{\pm 0.38}$	$89.22_{\pm 0.37}$	$94.02_{\pm 0.33}$
	MolTC	$95.89^{*}_{\pm 0.13}$	$98.87_{\pm 0.31}$	$89.39^{*}_{\pm 0.13}$	$95.55^{*}_{\pm 0.17}$	$95.59^{*}_{\pm 0.21}$	$98.06_{\pm 0.19}$	$96.70^{*}_{\pm 0.26}$	$98.85_{\pm 0.42}$
DO-CoLM (Ours)		$\mathbf{96.43}_{\pm 0.18}$	$\mathbf{99.13}_{\pm 0.31}$	$\mathbf{92.14}_{\pm 0.28}$	$\mathbf{96.03}_{\pm 0.12}$	$\mathbf{96.31}_{\pm 0.32}$	$\mathbf{98.52}_{\pm 0.16}$	$\mathbf{97.24}_{\pm 0.24}$	$\mathbf{99.12}_{\pm 0.31}$

Table 1: Comparative performance of various methods in qualitative interactive tasks. The best-performing methods are highlighted in bold, while the second-best methods are marked with * for emphasis.

Baseline Model		FreeSolv		Abraham		CompSol		CombiSolv	
		MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
GNN Based	MMGNN	$0.538_{\pm 0.031}$	$0.904_{\pm 0.028}$	$0.197_{\pm 0.009}$	$0.396_{\pm 0.012}$	$0.170^{*}_{\pm 0.005}$	$0.302_{\pm 0.005}$	$0.186_{\pm 0.004}$	$0.390^{*}_{\pm 0.015}$
	D-MPNN	$0.712_{\pm 0.015}$	$1.229_{\pm 0.029}$	$0.494_{\pm 0.013}$	$0.715_{\pm 0.027}$	$0.208_{\pm 0.007}$	$0.372_{\pm 0.008}$	$0.492_{\pm 0.014}$	$0.905_{\pm 0.055}$
	GEM	$0.608_{\pm 0.019}$	$1.198_{\pm 0.052}$	$0.254_{\pm 0.006}$	$0.537_{\pm 0.008}$	$0.205_{\pm 0.007}$	$0.343_{\pm 0.005}$	$0.295_{\pm 0.009}$	$0.773_{\pm 0.019}$
	CGIB	$0.550_{\pm 0.011}$	$0.922_{\pm 0.057}$	$0.261_{\pm 0.008}$	$0.531_{\pm 0.009}$	$0.174_{\pm 0.005}$	$0.313_{\pm 0.004}$	$0.231_{\pm 0.004}$	$0.395_{\pm 0.010}$
ML Based	GOVER	$0.646_{\pm 0.026}$	$1.094_{\pm 0.045}$	$0.367_{\pm 0.009}$	$0.635_{\pm 0.017}$	$0.187_{\pm 0.006}$	$0.381_{\pm 0.015}$	$0.413_{\pm 0.017}$	$0.738_{\pm 0.035}$
	SolvBert	$0.592_{\pm 0.031}$	$1.054_{\pm 0.054}$	$0.498_{\pm 0.008}$	$0.687_{\pm 0.015}$	$0.191_{\pm 0.010}$	$0.348_{\pm 0.008}$	$0.436_{\pm 0.018}$	$0.709_{\pm 0.021}$
	Uni-Mol	$0.576_{\pm 0.061}$	$1.009_{\pm 0.072}$	$0.365_{\pm 0.008}$	$0.612_{\pm 0.024}$	$0.197_{\pm 0.002}$	$0.345_{\pm 0.003}$	$0.269_{\pm 0.005}$	$0.671_{\pm 0.017}$
	SMD	$0.609_{\pm 0.037}$	$1.211_{\pm 0.036}$	$0.398_{\pm 0.022}$	$0.645_{\pm 0.037}$	$0.197_{\pm 0.006}$	$0.349_{\pm 0.007}$	$0.656_{\pm 0.012}$	$1.013_{\pm 0.031}$
LLM Based	Galactica	$0.892_{\pm 0.011}$	$1.398_{\pm 0.067}$	$0.651_{\pm 0.008}$	$1.058_{\pm 0.016}$	$0.597_{\pm 0.008}$	$0.862_{\pm 0.008}$	$0.841_{\pm 0.021}$	$1.456_{\pm 0.039}$
	Chem T5	$0.812_{\pm 0.036}$	$1.357_{\pm 0.057}$	$0.639_{\pm 0.010}$	$0.915_{\pm 0.017}$	$0.447_{\pm 0.008}$	$0.739_{\pm 0.010}$	$0.892_{\pm 0.015}$	$1.308_{\pm 0.024}$
	MolCA	$0.771_{\pm 0.035}$	$1.292_{\pm 0.041}$	$0.584_{\pm 0.007}$	$0.887_{\pm 0.011}$	$0.477_{\pm 0.008}$	$0.726_{\pm 0.023}$	$0.638_{\pm 0.043}$	$1.095_{\pm 0.037}$
	MolT5	$0.715_{\pm 0.047}$	$1.115_{\pm 0.075}$	$0.553_{\pm 0.009}$	$0.842_{\pm 0.006}$	$0.486_{\pm 0.003}$	$0.701_{\pm 0.008}$	$0.673_{\pm 0.031}$	$1.114_{\pm 0.029}$
	MolTC	$0.497^{*}_{\pm 0.013}$	$0.694^{*}_{\pm 0.042}$	$0.194^{*}_{\pm 0.011}$	$0.389^{*}_{\pm 0.010}$	$0.179_{\pm 0.006}$	$0.295^{*}_{\pm 0.004}$	$0.183^{*}_{\pm 0.004}$	$0.452_{\pm 0.008}$
DO-CoLM (Ours)		$\mathbf{0.469}_{\pm 0.014}$	$\mathbf{0.649}_{\pm 0.030}$	$\mathbf{0.175}_{\pm 0.008}$	$\mathbf{0.372}_{\pm 0.012}$	$\mathbf{0.165}_{\pm 0.007}$	$\mathbf{0.281}_{\pm 0.003}$	$\mathbf{0.172}_{\pm 0.005}$	$\mathbf{0.385}_{\pm 0.012}$

Table 2: Comparative performance of various methods in quantitative interactive tasks. The best-performing methods are highlighted in bold, while the second-best methods are marked with * for emphasis.

Qualitative Prediction Performance: Table 1 presents the comparative performance of DO-CoLM and various baseline methods on four widely-used chemical datasets. The results clearly demonstrate the consistent superiority of DO-CoLM over existing approaches across all datasets. For instance, on the ZhangDDI dataset, our model achieves a 3% improvement in accuracy compared to MolTC. This can be attributed to the incorporation of our molecular conformational relationship capture mechanism and the heterograph neural network, which effectively integrates conformational information. These enhancements enable the model to capture the 3D structural information and latent conformational interactions that are often overlooked in 1D SMILES and 2D molecular graphs. In addition, the DO-CoLM in our model reorders molecular token nodes from the model's perspective, allowing for a better understanding of molecular modality data and resulting in superior reasoning performance.

Quantitative Prediction Performance: Table 2 presents the performance of DO-CoLM on quantitative tasks. The data indicates that LLM-based models generally perform slightly worse than GNN-based and ML-based models. However, DO-CoLM has maintained a leading position in quantitative analysis tasks, which are typically challenging for LLM models. DO-CoLM achieved an RMSE of 0.385, representing nearly a 15% improvement over the second-best LLM-based model, MolTC, which had an RMSE of 0.452. Furthermore, DO-CoLM demonstrated significant performance gains over the best GNN-based model, MMGNN. For instance, on the FreeSolv dataset, DO-CoLM reduced the RMSE by nearly 28%. These improvements can be attributed to DO-CoLM's integration of 3D molecular conformation information, which more accurately captures molecular interactions, as well as chain-of-thought reasoning and its unique design considerations for input order from the model's perspective.

Datasets	Metrics	DO-CoLM				MolTC			
		EP	IP	EDRate (%)	IDRate (%)	EP	IP	RDRate (%)	IDRate (%)
Drugbank	Accuracy	$96.16_{\pm 0.19}$	$96.33_{\pm 0.18}$	**0.27**	**0.11**	$84.43_{\pm 0.17}$	$94.03_{\pm 0.13}$	**12.07**	**1.93**
	AUC-ROC	$98.89_{\pm 0.30}$	$99.02_{\pm 0.31}$	**0.24**	**0.10**	$87.62_{\pm 0.42}$	$96.77_{\pm 0.31}$	**11.37**	**2.12**
ZhangDDI	Accuracy	$92.01_{\pm 0.27}$	$92.12_{\pm 0.28}$	**0.14**	**0.02**	$80.15_{\pm 0.16}$	$88.04_{\pm 0.13}$	**10.33**	**1.05**
	AUC-ROC	$95.76_{\pm 0.14}$	$95.95_{\pm 0.12}$	**0.15**	**0.03**	$85.64_{\pm 0.19}$	$94.37_{\pm 0.21}$	**10.41**	**1.19**
ChChMiner	Accuracy	$96.06_{\pm 0.27}$	$96.14_{\pm 0.32}$	**0.25**	**0.14**	$82.78_{\pm 0.17}$	$93.78_{\pm 0.21}$	**13.34**	**1.89**
	AUC-ROC	$98.49_{\pm 0.16}$	$98.42_{\pm 0.15}$	**0.27**	**0.10**	$85.42_{\pm 0.18}$	$96.03_{\pm 0.19}$	**12.89**	**2.07**
DeepDDI	Accuracy	$97.05_{\pm 0.23}$	$97.14_{\pm 0.24}$	**0.12**	**0.09**	$87.62_{\pm 0.31}$	$95.46_{\pm 0.26}$	**9.45**	**1.30**
	AUC-ROC	$99.01_{\pm 0.32}$	$99.03_{\pm 0.31}$	**0.11**	**0.10**	$89.76_{\pm 0.37}$	$96.95_{\pm 0.42}$	**9.20**	**1.81**

Table 3: Comparison of results for perturbations in overall external order (EP) and internal multimodal node order (IP). EDRate reflects the performance decrease due to external order perturbations, while IDRate shows the decrease due to internal multimodal node order perturbations.

4.4 Sequence perturbation performance

In this section, we test the model's robustness by introducing perturbations to the input data order. Table 3 presents the experimental results of DO-CoLM and MolTC under input sequence perturbations. Since input order inherently exhibits asymmetry in quantitative prediction tasks, this evaluation primarily focuses on qualitative prediction tasks. We introduced two types of input perturbations for the experiments: overall molecular input order perturbation and internal molecular conformation node order perturbation. As shown in Table 3, DO-CoLM demonstrates significant robustness in qualitative prediction tasks, with overall molecular input order perturbation having almost no impact on its performance, while MolTC exhibits significant performance degradation. For example, on the DrugBank dataset, MolTC's accuracy drops from 95.89% to 84.43%, a decrease of 12.07%, and AUC-ROC decreases by 11.37%. Under internal molecular conformation node order perturbations, DO-CoLM remains stable, while MolTC experiences an average decline of 1.7%. This is due to the design of the AOAM, which enables the model to adaptively handle different input sequences, resulting in stable performance.

4.5 Ablation Study

Table 4 presents the results of the ablation experiments, where **w/o DCR** indicates the removal of the DCRCM. **w/o AO** represents the exclusion of AOAM, replaced by three alternative ordering methods: random order, degree-based order, and fixed order. By removing the DCRCM, we analyzed the impact of molecular conformational information interaction on model performance. Additionally, by comparing model performance under random, fixed, and degree-based orders, we validated that AOAM identifies the optimal sequence from the model's perspective. The experimental results demonstrate that each module significantly improves model performance. The removal of DCRCM has the most noticeable impact, causing nearly a 10% decrease in performance on quantitative tasks. Furthermore, replacing AOAM with other prior knowledge-based ordering methods leads to a certain degree of performance degradation. We also performed an ablation study on the conformations, as detailed in the Appendix E.

Dataset	Metric	w/o DCR	w/o AO		
			Random	Deg	Fixed
DDI	Accuracy	$1.22_{\pm 0.14}$	$0.63_{\pm 0.07}$	$0.44_{\pm 0.04}$	$0.56_{\pm 0.05}$
	Rate (↓)	1.51 %	0.77 %	0.54 %	0.69 %
	AUCROC	$1.67_{\pm 0.32}$	$0.78_{\pm 0.08}$	$0.55_{\pm 0.07}$	$0.62_{\pm 0.09}$
	Rate (↓)	1.78 %	0.83 %	0.58 %	0.66 %
SSI	MAE	$0.017_{\pm 0.004}$	$0.012_{\pm 0.002}$	$0.007_{\pm 0.003}$	$0.010_{\pm 0.004}$
	Rate (↑)	7.43 %	5.24 %	3.06 %	4.37 %
	RMSE	$0.025_{\pm 0.007}$	$0.016_{\pm 0.003}$	$0.011_{\pm 0.004}$	$0.012_{\pm 0.005}$
	Rate (↑)	9.43 %	6.03 %	3.72 %	4.52 %
CSI Abs.	MAE	$1.07_{\pm 0.11}$	$0.61_{\pm 0.03}$	$0.48_{\pm 0.04}$	$0.51_{\pm 0.05}$
	Rate (↑)	8.04 %	4.58 %	3.61 %	3.84 %
	RMSE	$1.58_{\pm 0.20}$	$0.72_{\pm 0.06}$	$0.60_{\pm 0.05}$	$0.65_{\pm 0.04}$
	Rate (↑)	10.18 %	4.64 %	3.87 %	4.18 %
CSI Emis.	MAE	$2.07_{\pm 0.17}$	$1.05_{\pm 0.14}$	$0.74_{\pm 0.10}$	$0.81_{\pm 0.11}$
	Rate (↑)	11.12 %	5.64 %	3.97 %	4.36 %
	RMSE	$2.89_{\pm 0.28}$	$1.47_{\pm 0.12}$	$1.03_{\pm 0.11}$	$1.10_{\pm 0.15}$
	Rate (↑)	12.34 %	6.27 %	4.40 %	4.71 %
CSI Life.	MAE	$0.074_{\pm 0.003}$	$0.026_{\pm 0.004}$	$0.016_{\pm 0.002}$	$0.020_{\pm 0.003}$
	Rate (↑)	10.70 %	3.76 %	2.31 %	2.89 %
	RMSE	$0.095_{\pm 0.010}$	$0.034_{\pm 0.008}$	$0.023_{\pm 0.007}$	$0.029_{\pm 0.010}$
	Rate (↑)	11.26 %	4.02 %	2.73 %	3.43 %

Table 4: Ablation Study of Removing DCRCM (w/o DCR) and Replacing AOAM (w/o AO) with other ordering methods.

5 Conclusion

In this paper, we propose **DO-CoLM**, which incorporates the Dynamic Conformational Relationship Capture Module and the Adaptive Order Adjustment Module to address the challenges of inadequate molecular 3D information and dynamics, as well as sensitivity to input order in existing LM-based MRL models. Experiments across twelve diverse datasets in various domains demonstrate the superiority of our approach over current GNN and LLM-based baselines. This advancement establishes a new standard for integrating multimodal data in LLM-based MRL. The limitations are further discussed in Appendix F.

Ethical Statement

There are no ethical issues.

Acknowledgments

This work was partially supported by the Project of Stable Support for Youth Team in Basic Research Field, CAS (YSBR-005), Anhui Science Foundation for Distinguished Young Scholars (No.1908085J24), Natural Science Foundation of China (No.62072427), Jiangsu Natural Science Foundation (No. BK20191193).

Contribution Statement

Zhuo Chen is the first author; Wenjie Du and Yang Wang are corresponding authors.

References

[Alonso *et al.*, 2006] Hernan Alonso, Andrey A Bliznyuk, and Jill E Gready. Combining docking and molecular dynamic simulations in drug design. *Medicinal research reviews*, 26(5):531–568, 2006.

[Andrade *et al.*, 2009] Carolina H Andrade, Kerly FM Pasqualoto, Elizabeth I Ferreira, and Anton J Hopfinger. Rational design and 3d-pharmacophore mapping of 5-thiourea-substituted α-thymidine analogues as mycobacterial tmpk inhibitors. *Journal of chemical information and modeling*, 49(4):1070–1078, 2009.

[Berglund *et al.*, 2023] Lukas Berglund, Meg Tong, Max Kaufmann, Mikita Balesni, Asa Cooper Stickland, Tomasz Korbak, and Owain Evans. The reversal curse: Llms trained on" a is b" fail to learn" b is a". *arXiv preprint arXiv:2309.12288*, 2023.

[Christofidellis *et al.*, 2023] Dimitrios Christofidellis, Giorgio Giannone, Jannis Born, Ole Winther, Teodoro Laino, and Matteo Manica. Unifying molecular and textual representations via multi-task language modelling. *arXiv preprint arXiv:2301.12586*, 2023.

[Cuturi, 2013] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, 26, 2013.

[Deac *et al.*, 2019] Andreea Deac, Yu-Hsiang Huang, Petar Veličković, Pietro Liò, and Jian Tang. Drug-drug adverse effect prediction with graph co-attention. *arXiv preprint arXiv:1905.00534*, 2019.

[Du *et al.*,] Wenjie Du, Shuai Zhang, Zhaohui Cai, Zhiyuan Liu, Junfeng Fang, Jianmin Wang, and Yang Wang. Molecular merged hypergraph neural network for explainable solvation free energy prediction. *Research*, 0(ja).

[Du *et al.*, 2024] Wenjie Du, Shuai Zhang, Di Wu, Jun Xia, Ziyuan Zhao, Junfeng Fang, and Yang Wang. MMGNN: A molecular merged graph neural network for explainable solvation free energy prediction. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI 2024, Jeju, South Korea, August 3-9, 2024*, pages 5808–5816. ijcai.org, 2024.

[Edwards *et al.*, 2021] Carl Edwards, ChengXiang Zhai, and Heng Ji. Text2mol: Cross-modal molecule retrieval with natural language queries. In *EMNLP (1)*, pages 595–607. Association for Computational Linguistics, 2021.

[Edwards *et al.*, 2022] Carl Edwards, Tuan Lai, Kevin Ros, Garrett Honke, Kyunghyun Cho, and Heng Ji. Translation between molecules and natural language. *arXiv preprint arXiv:2204.11817*, 2022.

[Fang *et al.*, 2022] Xiaomin Fang, Lihang Liu, Jieqiong Lei, Donglong He, Shanzhuo Zhang, Jingbo Zhou, Fan Wang, Hua Wu, and Haifeng Wang. Geometry-enhanced molecular representation learning for property prediction. *Nature Machine Intelligence*, 4(2):127–134, 2022.

[Fang *et al.*, 2024] Junfeng Fang, Shuai Zhang, Chang Wu, Zhengyi Yang, Zhiyuan Liu, Sihang Li, Kun Wang, Wenjie Du, and Xiang Wang. Moltc: Towards molecular relational modeling in language models. *arXiv preprint arXiv:2402.03781*, 2024.

[Fu *et al.*, 2020] Tianfan Fu, Cao Xiao, and Jimeng Sun. Core: Automatic molecule optimization using copy & refine strategy. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 638–645, 2020.

[Jablonka *et al.*, 2024] Kevin Maik Jablonka, Philippe Schwaller, Andres Ortega-Guerrero, and Berend Smit. Leveraging large language models for predictive chemistry. *Nature Machine Intelligence*, 6(2):161–169, 2024.

[Lee *et al.*, 2023] Namkyeong Lee, Dongmin Hyun, Gyoung S. Na, Sungwon Kim, Junseok Lee, and Chanyoung Park. Conditional graph information bottleneck for molecular relational learning. In *ICML*, volume 202 of *Proceedings of Machine Learning Research*, pages 18852–18871. PMLR, 2023.

[Li *et al.*, 2023] Zimeng Li, Shichao Zhu, Bin Shao, Xiangxiang Zeng, Tong Wang, and Tie-Yan Liu. Dsn-ddi: an accurate and generalized framework for drug–drug interaction prediction by dual-view representation learning. *Briefings in Bioinformatics*, 24(1):bbac597, 2023.

[Li *et al.*, 2024] Sihang Li, Zhiyuan Liu, Yanchen Luo, Xiang Wang, Xiangnan He, Kenji Kawaguchi, Tat-Seng Chua, and Qi Tian. 3d-molm: Towards 3d molecule-text interpretation in language models. In *ICLR*, 2024.

[Liang *et al.*, 2023] Youwei Liang, Ruiyi Zhang, Li Zhang, and Pengtao Xie. Drugchat: Towards enabling chatgpt-like capabilities on drug molecule graphs. *ArXiv*, abs/2309.03907, 2023.

[Lin *et al.*, 2022] Shenggeng Lin, Yanjing Wang, Lingfeng Zhang, Yanyi Chu, Yatong Liu, Yitian Fang, Mingming Jiang, Qiankun Wang, Bowen Zhao, Yi Xiong, et al. Mdf-sa-ddi: predicting drug–drug interaction events based on multi-source drug fusion, multi-source feature fusion and transformer self-attention mechanism. *Briefings in Bioinformatics*, 23(1):bbab421, 2022.

[Liu *et al.*, 2023] Zhiyuan Liu, Sihang Li, Yanchen Luo, Hao Fei, Yixin Cao, Kenji Kawaguchi, Xiang Wang, and Tat-Seng Chua. Molca: Molecular graph-language modeling

with cross-modal projector and uni-modal adapter. *arXiv preprint arXiv:2310.12798*, 2023.

[Meng *et al.*, 2023] Fanwang Meng, Hanwen Zhang, Juan Samuel Collins-Ramirez, and Paul W. Ayers. Something for nothing: Improved solvation free energy prediction with learning. 2023.

[Morris *et al.*, 2019] Christopher Morris, Martin Ritzert, Matthias Fey, William L Hamilton, Jan Eric Lenssen, Gaurav Rattan, and Martin Grohe. Weisfeiler and leman go neural: Higher-order graph neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 4602–4609, 2019.

[Mortier *et al.*, 2015] Jérémie Mortier, Christin Rakers, Marcel Bermudez, Manuela S Murgueitio, Sereina Riniker, and Gerhard Wolber. The impact of molecular dynamics on drug design: applications for the characterization of ligand–macromolecule complexes. *Drug discovery today*, 20(6):686–702, 2015.

[Nyamabo *et al.*, 2021] Arnold K Nyamabo, Hui Yu, and Jian-Yu Shi. Ssi–ddi: substructure–substructure interactions for drug–drug interaction prediction. *Briefings in Bioinformatics*, 22(6):bbab133, 2021.

[Qi *et al.*, 2017] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652–660, 2017.

[Rong *et al.*, 2020] Yu Rong, Yatao Bian, Tingyang Xu, Weiyang Xie, Ying Wei, Wenbing Huang, and Junzhou Huang. Self-supervised graph transformer on large-scale molecular data. In *NeurIPS*, 2020.

[Ryu *et al.*, 2018] Jae Yong Ryu, Hyun Uk Kim, and Sang Yup Lee. Deep learning improves prediction of drug–drug and drug–food interactions. *Proceedings of the national academy of sciences*, 115(18):E4304–E4311, 2018.

[Sagawa and Kojima, 2023] Tatsuya Sagawa and Ryosuke Kojima. Reaction5: a large-scale pre-trained model towards application of limited reaction data. *arXiv preprint arXiv:2311.06708*, 2023.

[Sizhe Liu *et al.*, 2024] Sizhe Sizhe Liu, Jun Xia, Lecheng Zhang, Yuchen Liu, Yue Liu, Wenjie Du, Zhangyang Gao, Bozhen Hu, Cheng Tan, Stan Z Li, et al. Flexmol: A flexible toolkit for benchmarking molecular relational learning. *Advances in Neural Information Processing Systems*, 37:35454–35467, 2024.

[Skjærven *et al.*, 2011] Lars Skjærven, Nathalie Reuter, and Aurora Martinez. Dynamics, flexibility and ligand-induced conformational changes in biological macromolecules: a computational approach. *Future Medicinal Chemistry*, 3(16):2079–2100, 2011.

[Tan *et al.*, 2024] Zhijie Tan, Xu Chu, Weiping Li, and Tong Mo. Order matters: Exploring order sensitivity in multimodal large language models. *arXiv preprint arXiv:2410.16983*, 2024.

[Tang *et al.*, 2024] Xiangru Tang, Andrew Tran, Jeffrey Tan, and Mark B. Gerstein. Mollm: A unified language model for integrating biomedical text with 2d and 3d molecular representations. *bioRxiv*, 2024.

[Taylor *et al.*, 2022a] Ross Taylor, Marcin Kardas, Guillem Cucurull, Thomas Scialom, Anthony Hartshorn, Elvis Saravia, Andrew Poulton, Viktor Kerkez, and Robert Stojnic. Galactica: A large language model for science. *arXiv preprint arXiv:2211.09085*, 2022.

[Taylor *et al.*, 2022b] Ross Taylor, Marcin Kardas, Guillem Cucurull, Thomas Scialom, Anthony Hartshorn, Elvis Saravia, Andrew Poulton, Viktor Kerkez, and Robert Stojnic. Galactica: A large language model for science. *arXiv preprint arXiv:2211.09085*, 2022.

[Vermeire and Green, 2021] Florence H Vermeire and William H Green. Transfer learning for solvation free energies: From quantum chemistry to experiments. *Chemical Engineering Journal*, 418:129307, 2021.

[Vinyals *et al.*, 2015] Oriol Vinyals, Samy Bengio, and Manjunath Kudlur. Order matters: Sequence to sequence for sets. *arXiv preprint arXiv:1511.06391*, 2015.

[Wu *et al.*, 2022] Fang Wu, Shuting Jin, Yinghui Jiang, Xurui Jin, Bowen Tang, Zhangming Niu, Xiangrong Liu, Qiang Zhang, Xiangxiang Zeng, and Stan Z Li. Pretraining of equivariant graph matching networks with conformation flexibility for drug binding. *Advanced Science*, 9(33):2203796, 2022.

[Yu *et al.*, 2023] Jiahui Yu, Chengwei Zhang, Yingying Cheng, Yun-Fang Yang, Yuan-Bin She, Fengfan Liu, Weike Su, and An Su. Solvbert for solvation free energy and solubility prediction: a demonstration of an nlp model for predicting the properties of molecular complexes. *Digital Discovery*, 2(2):409–421, 2023.

[Zhang *et al.*, 2025] Shuai Zhang, Junfeng Fang, Xuqiang Li, hongxin xiang, ALAN XIA, Ye Wei, Wenjie Du, and Yang Wang. Iterative substructure extraction for molecular relational learning with interactive graph information bottleneck. In *The Thirteenth International Conference on Learning Representations*, 2025.

[Zheng *et al.*, 2025] Yizhen Zheng, Huan Yee Koh, Jiaxin Ju, Anh TN Nguyen, Lauren T May, Geoffrey I Webb, and Shirui Pan. Large language models for scientific discovery in molecular property prediction. *Nature Machine Intelligence*, pages 1–11, 2025.

[Zhong *et al.*, 2024] Yujie Zhong, Guangming Li, Jianxin Yang, et al. Learning motif-based graphs for drug–drug interaction prediction via local–global self-attention. *Nature Machine Intelligence*, 6:1094–1105, 2024.

[Zhou *et al.*, 2023] Gengmo Zhou, Zhifeng Gao, Qiankun Ding, Hang Zheng, Hongteng Xu, Zhewei Wei, Linfeng Zhang, and Guolin Ke. Uni-mol: a universal 3d molecular representation learning framework. 2023.