MTPNet: Multi-Grained Target Perception for Unified Activity Cliff Prediction

Zishan Shu[1,3], **Yufan Deng**[1,3], **Hongyu Zhang**[1,3], **Zhiwei Nie**[1,2,3*] and **Jie Chen**[1,2,3*]
[1]School of Electronic and Computer Engineering, Peking University, Shenzhen, China
[2]Pengcheng Laboratory, Shenzhen, China
[3]AI for Science (AI4S)-Preferred Program, Peking University Shenzhen Graduate School, China
{zishanshu, dengyufan10, zhanghy}@stu.pku.edu.cn,
{zhiweiNie, jiechen2019}@pku.edu.cn,

Abstract

Activity cliff prediction is a critical task in drug discovery and material design. Existing computational methods are limited to handling single binding targets, which restricts the applicability of these prediction models. In this paper, we present the Multi-Grained Target Perception network (MTPNet) to incorporate the prior knowledge of interactions between the molecules and their target proteins. Specifically, MTPNet is a unified framework for activity cliff prediction, which consists of two components: Macro-level Target Semantic (MTS) guidance and Micro-level Pocket Semantic (MPS) guidance. By this way, MTPNet dynamically optimizes molecular representations through multi-grained protein semantic conditions. To our knowledge, it is the first time to employ the receptor proteins as guiding information to effectively capture critical interaction details. Extensive experiments on 30 representative activity cliff datasets demonstrate that MTPNet significantly outperforms previous approaches, achieving an average RMSE improvement of 18.95% on top of several mainstream GNN architectures. Overall, MTPNet internalizes interaction patterns through conditional deep learning to achieve unified predictions of activity cliffs, helping to accelerate compound optimization and design. Codes are available at: https://github.com/ZishanShu/MTPNet.

1 Introduction

In the field of drug discovery and design, Activity Cliffs (AC) refer to the phenomenon where minor structural changes in molecules lead to significant differences in biological activity [Van Tilborg *et al.*, 2022]. Studying ACs is crucial because even compounds with similar structures can exhibit drastically different biological activities, complicating the drug design process. Traditional computational methods primarily rely on molecular fingerprint comparison and similar techniques for activity cliff prediction [Consonni and Todeschini, 2010], but suffer from limited robustness [Wang *et al.*, 2017]. In recent years, GNN-based deep learning approaches have emerged as leaders in this field [Shin *et al.*, 2024; Meng *et al.*, 2024; Yang *et al.*, 2023; Zhu *et al.*, 2023], overcoming the limitations of traditional techniques. For instance, MoleBERT [Xia *et al.*, 2023b] integrates GNNs with pre-training to improve molecular representation accuracy, significantly improving predictive performance. Additionally, GNN-based models like ACGCN [Park *et al.*, 2022] and MolCLR [Wang *et al.*, 2022] have effectively captured complex structure-activity relationships, advancing the accuracy and applicability of activity cliff prediction.

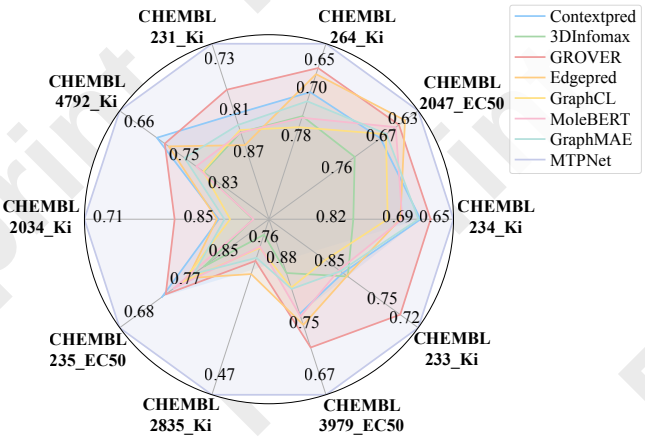

Figure 1: The overall RMSE performance comparison of various mainstream models across multiple activity cliff datasets. The radar chart shows that MTPNet significantly outperforms existing methods, achieving a 7.2% improvement over current SOTA models, highlighting the critical role of incorporating receptor protein information in enhancing activity cliff prediction performance. RMSE values are shown in reciprocal form to facilitate presentation.

However, existing methods primarily focus on modeling molecules themselves, overlooking the critical role of paired receptor proteins in chemical reactions, particularly in the context of activity cliff (AC) prediction. As shown in Figure 2, these methods face two major challenges: First, the insufficient use of protein features hampers accurate modeling of interactions between molecules and proteins, impacting the precision of AC predictions. Second, these methods struggle to generalize across various types of AC prediction tasks, constraining their applicability to different binding tar-

*Corresponding author.

gets. The latter, in particular, has become a significant bottleneck that hinders the widespread adoption of existing approaches. The fundamental principle of ACs suggests that even minor structural changes in molecules can lead to drastic shifts in biological activity, typically driven by complex interactions between ligands and receptor proteins. However, current methods fail to effectively capture these critical dynamic interaction characteristics. To address these challenges, we introduce the Multi-Grained Target Perception (MTP) module, which effectively integrates conditional information from receptor proteins and enhances the model's ability to perceive subtle structural changes. Specifically, the MTP module combines both macro and micro-level semantic guidance, enabling the identification of broad interaction patterns between molecules and the precise detection of small structural variations that result in significant differences in biological activity.

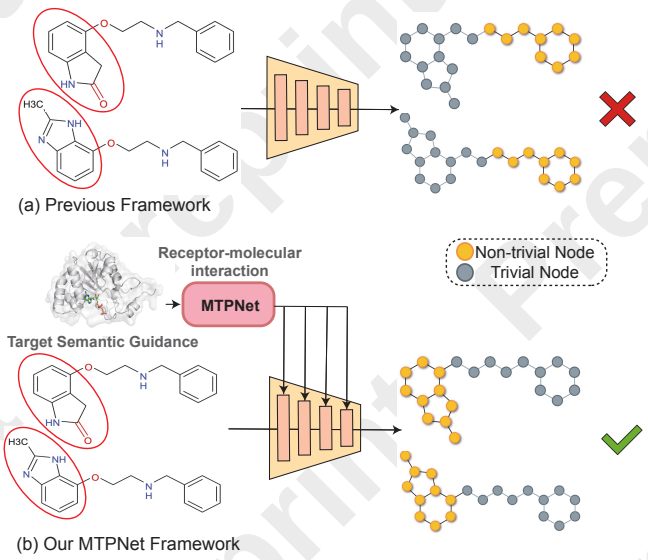

Figure 2: Motivation for incorporating receptor protein information in MTPNet. (a) The traditional methods, which only consider molecular features, focus on homogeneous and less significant parts (Motif Pattern). (b) MTPNet, by incorporating receptor protein information, better captures the interaction between molecules and receptors, focusing on the substitution differences in small molecules, thereby effectively revealing the underlying mechanisms of activity cliff formation.

Building upon the MTP module, we further develop MTPNet, a unified framework for activity cliff prediction. MTPNet leverages the MTP module to incorporate receptor protein information into the molecule feature extraction process, enabling efficient predictions across multiple binding targets. By precisely modeling the interactions between molecules and proteins, MTPNet significantly outperforms existing models on datasets with single binding targets, demonstrating superior prediction performance and interpretability. Specifically, experimental results indicate that MTPNet improves RMSE by 7.2% compared to existing SOTA models. Furthermore, MTPNet achieves an Area Under the Curve (AUC) of 0.924, surpassing models such as Mole-BERT (AUC = 0.902) and MolCLR (AUC = 0.896), thereby highlighting its robust generalization capabilities and practical application value across multiple receptor-ligand systems. Moreover, plug-and-play evaluations reveal that with the MTP module, PCC shows an average improvement of 11.6%, R^2 improves by 17.8%, and RMSE improves by 19.0%, further demonstrating the module's seamless integrability and effectiveness.

In summary, the main contributions of this paper are summaried as follows:

- **Integration of Receptor Protein Perception:** To the best of our knowledge, MTPNet is the first conditional framework to incorporate receptor protein information into activity cliff prediction task, internalizing interaction patterns through progressive conditional deep learning.
- **Multi-Grained Target Perception Module:** We propose the Multi-Grained Target Perception (MTP) module, which dynamically optimizes molecular representations through Macro-level Target Semantic (MTS) guidance and Micro-level Pocket Semantic (MPS) guidance, effectively enhancing the predictive performance of the model by complementing the interaction patterns at different levels.
- **Unified Framework for Activity Cliff Prediction:** MTPNet provides a unified solution for activity cliff prediction across diverse receptor-ligand systems, offering applicability and interpretability.
- **Superior Predictive Performance:** Extensive experiments on 30 representative activity cliff datasets demonstrate that MTPNet achieves significant improvements in predictive performance, including an average RMSE reduction of 18.95% on top of several mainstream GNN architectures.

2 Related Work

Activity Cliff Prediction Methods: Traditional computational methods primarily rely on techniques such as molecular fingerprint comparisons to predict Activity Cliffs [Moriwaki *et al.*, 2018]. Early computational approaches utilized traditional machine learning methods like Support Vector Machines (SVM) [Vapnik, 2013] and Support Vector Regression (SVR) [Drucker *et al.*, 1996] for Activity Cliffs prediction. Although significant progress has been made in achieving high-throughput prediction results, these methods still suffer from limitations in robustness [Dong *et al.*, 2018], generalization [Butler *et al.*, 2018], and interpretability [Moriwaki *et al.*, 2018].

Recent advances in deep learning have significantly improved the prediction of activity cliffs, particularly through enhanced molecular feature representation [Xia *et al.*, 2023a]. Early studies, such as Iqbal et al. [Iqbal *et al.*, 2021], utilized Convolutional Neural Networks (CNNs) to capture spatial features of molecular structures, showcasing the potential of CNNs in molecular data processing. However, as the demand for more powerful molecular embeddings grew, Graph Neural Networks (GNNs) emerged as a superior alternative due to their ability to directly model molecular graphs [Du *et al.*, 2024], effectively capturing complex molecular interactions and learning rich embeddings [Xiang *et al.*, 2024; Zheng *et al.*, 2024; Nie *et al.*, 2024b; Wu, 2024; Nie *et*

al., 2024a]. For instance, the MoleBERT model [Xia *et al.*, 2023b], which integrates GNNs with pretraining techniques, has significantly improved predictive performance across multiple datasets. Park et al. [Park *et al.*, 2022] introduced the ACGCN model, addressing the information loss issues commonly associated with traditional fingerprint-based methods in structure-activity relationship (SAR) analysis. Additionally, MolCLR [Wang *et al.*, 2022] employs contrastive learning to pretrain molecular graphs, while Transformer-based models have further advanced the field. DeepAC [Chen *et al.*, 2022], a conditional Transformer model, leverages SMILES sequences and activity differences to predict ACs and generate novel compounds. GROVER [Rong *et al.*, 2020], which combines Transformer and GNN architectures, captures both local and global molecular features, achieving outstanding performance.

Protein Language Models: Protein language models (PLMs) have significantly advanced receptor protein feature extraction [Zhao *et al.*, 2023; Li *et al.*, 2023]. ESM [Lin *et al.*, 2023] uses self-supervised learning to pretrain on protein sequences, capturing rich features for downstream receptor function prediction, particularly on large-scale datasets. ProteinBERT [Brandes *et al.*, 2022], pretrained on large protein datasets, optimizes BERT for sequence modeling, showing high adaptability and efficiency. DeepProSite [Fang *et al.*, 2023] integrates sequence and structural information to improve receptor function prediction accuracy. DeepProtein [Xie and Fu, 2025] provides a versatile library and benchmark for protein sequence learning, enhancing predictions of function, localization, and interactions. SaProt [Su *et al.*, 2023] integrates sequence and structural data for precise feature extraction, excelling in ligand binding and function prediction.

3 Methodology

3.1 Preliminaries

Activity Cliff Definition: In the process of molecular binding with receptor proteins, the binding target describes the spatial arrangement and interaction patterns between molecules and protein receptors. Different binding targets can affect how molecules bind to receptor proteins, thereby influencing the training of the model. Therefore, understanding and handling different binding targets is crucial for improving the accuracy of activity cliff prediction models. In the task of activity cliff prediction, based on the diversity of different datasets and binding targets, we categorize binding targets into single binding target and multiple binding targets.

Formally, let D be an activity cliff molecular dataset, where each x_i represents the input features of a molecular-receptor pair, and y_i represents the corresponding continuous property value, i.e., the change in compound potency values (ΔpK_i), reflecting the difference in assay-independent equilibrium constants (K_i). The input features x_i include the receptor protein features $x_i^{\text{pro}(m)}$ and the ligand molecule features $x_i^{\text{mol(m)}}$, which are fused using the Multi-Grained Target Perception (MTP) Module to capture the important interaction features between the protein and ligand.

Single Binding Target: In single binding target learning, it is assumed that all molecular-receptor pairs follow the same binding target, and the dataset is represented as:

$$D = \{(x_i^{\text{pro}}, x_i^{\text{mol}}, y_i)\}_{i=1}^{N} \tag{1}$$

The learning objective is to fit a mapping function:

$$f : (x_i^{\text{pro}}, x_i^{\text{mol}}) \rightarrow y_i \tag{2}$$

Multiple Binding Targets: In multiple binding target learning, it is assumed that the dataset contains multiple different and independent binding targets, and the dataset is represented as:

$$D = \bigcup_{m=1}^{M} D_m \tag{3}$$

where each $D_m = \{(x_i^{\text{pro}(m)}, x_i^{\text{mol}(m)}, y_i^{(m)})\}_{i=1}^{N_m}$ corresponds to the m-th binding target. The learning objective remains to fit a general mapping function:

$$f_m : (x_i^{\text{pro}(m)}, x_i^{\text{mol}(m)}) \rightarrow y_i^{(m)} \tag{4}$$

3.2 MTPNet

MTPNet is a receptor-aware framework designed to unify activity cliff prediction tasks across multiple binding targets, achieving efficient and flexible modeling. At its core lies the Multi-Grained Target Perception (MTP) Module, which employs a Multigranularity Protein Semantic Condition through Macro-level Target Semantic (MTS) guidance and Micro-level Pocket Semantic (MPS) guidance to dynamically capture the complex interaction patterns between receptors and ligands. In the MTPNet framework, molecules and target proteins are first embedded, where the ESM2 model is used to extract deep semantic information from proteins, and the Mole-BERT model is employed to capture key features of molecules. Subsequently, the MTP module alternates between global context integration and local structural refinement, progressively optimizing ligand feature representations and aligning them with target features. This approach significantly enhances the capability to model activity cliff phenomena.

MTPNet leverages MTP's global-local guidance mechanism to dynamically adapt to multiple binding targets, avoiding redundancy of separate models. This design seamlessly integrates global features F_{target} and local features F_{pocket} into the optimization of ligand representations F_{mol}, unifying the feature modeling process across multiple binding targets. By fusing receptor protein and ligand features and modeling multiple binding targets through MTP, MTPNet not only addresses the challenges of activity cliff prediction in multi-binding target scenarios but also provides a highly efficient and accurate framework for diverse receptor-ligand systems. This makes MTPNet a unified and flexible paradigm for activity cliff modeling.

Multi-Grained Target Perception (MTP) Module

The Multi-Grained Target Perception (MTP) Module employs a Multigranularity Protein Semantic Condition to dynamically align ligand and target features through global and localized interaction modeling. This strategy combines two

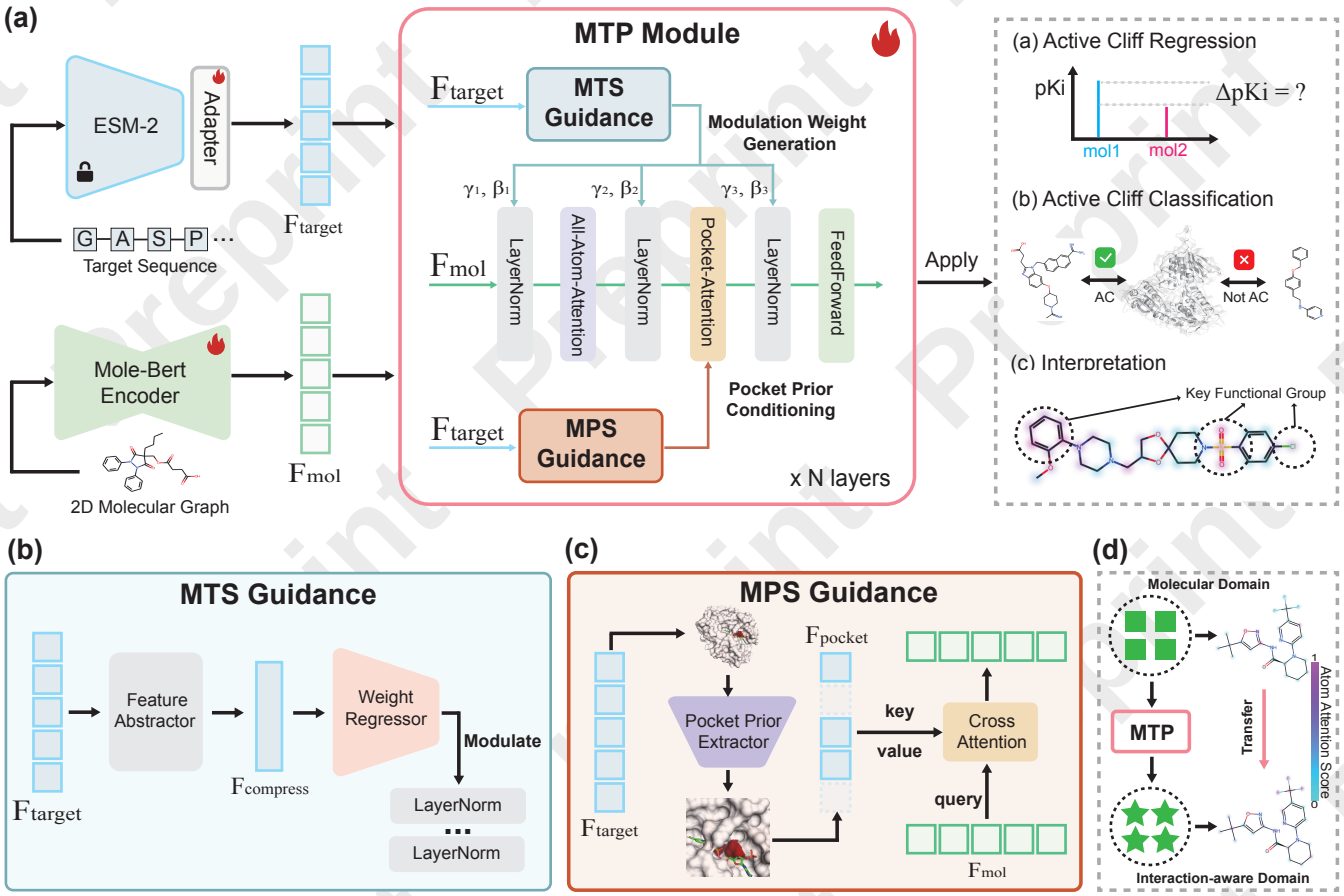

Figure 3: The overview of MTPNet. (a) The complete workflow of MTPNet, beginning with the embedding of molecules and target proteins, followed by the MTP module, and concluding with three key applications: activity cliff regression, activity cliff classification, and molecular interpretation. (b) Detailed architecture of the Macro-level Target Semantic (MTS) Guidance, which captures global receptor semantic information. (c) Detailed architecture of the Micro-level Pocket Semantic (MPS) Guidance, which focuses on local receptor-ligand interactions within binding pockets. (d) The transformation of the molecular domain into an interaction-aware domain after passing through the MTP module, highlighting its physicochemical significance.

complementary mechanisms: Macro-level Target Semantic (MTS) Guidance and Micro-level Pocket Semantic (MPS) Guidance. In the MTP module, the input ligand feature matrix $F_{\text{mol}} \in \mathbb{R}^{m \times d}$ and target protein feature matrix $F_{\text{target}} \in \mathbb{R}^{n \times d}$are iteratively refined layer by layer through global and localized semantic modeling to optimize F_{mol}.

Thus, the layer-by-layer adjustment process is essentially a recursive optimization of F_{mol}, while target features F_{target} provide global and localized semantic guidance alternately through MTS and MPS. By alternating between these two mechanisms and using attention mechanisms, the MTP module dynamically captures critical molecular-receptor interaction patterns underlying activity cliffs.

In the MTP module, the outputs of MTS (see Eq.11) and MPS (see Eq.13) are alternately used within a stacked attention framework, enabling progressive updates to the ligand representation. This process is described as follows:

First, the MTS Guidance generates the initial optimized ligand features by aligning the ligand representation F_{mol} with the global context of the receptor F_{target}. This alignment is achieved through a self-attention mechanism (SA) combined with conditionally normalized ligand features, as defined in Eq.11. The initial ligand representation is computed as:

$$F_{\text{mol}}^{(0)} = \Phi_{\text{MTS}}(F_{\text{mol}}, F_{\text{target}}) \tag{5}$$

At the l-th layer, the ligand features are refined by alternately incorporating localized receptor information from the MPS Guidance. Using the cross-attention mechanism, the MPS Guidance integrates the ligand features F_{mol} with pocket features F_{pocket}, as defined in Eq.13. The ligand representation at each layer is updated iteratively as:

$$F_{\text{mol}}^{(l)} = F_{\text{mol}}^{(l-1)} + \Phi_{\text{MPS}}(F_{\text{mol}}^{(l-1)}, F_{\text{target}}) \tag{6}$$

In each iteration, the refined ligand features $F_{\text{mol}}^{(l)}$ undergo a further transformation that includes a two-layer feedforward

network with dropout and ReLU activation. This process ensures that the updated features effectively capture the dynamic receptor-ligand interaction patterns.

Finally, after L stacked layers of iterative refinement, the MTP module outputs the fully optimized ligand features Φ_{MTP}, which comprehensively integrate both global (MTS) and localized (MPS) receptor information (see Eq.7).

$$\Phi_{\text{MTP}}(F_{\text{mol}}, F_{\text{target}}) = \Phi_{\text{MTS}}(F_{\text{mol}}, F_{\text{target}}) + \sum_{l=1}^{L} \Phi_{\text{MPS}}(F_{\text{mol}}^{(l-1)}, F_{\text{target}}) \quad (7)$$

The MTP module provides a comprehensive framework for global and localized semantic optimization. Within the multi-layer stacked attention structure, the MTP module combines the global and localized feature extraction capabilities of MTS and MPS, effectively addressing the challenges posed by multi-binding target scenarios while efficiently integrating receptor protein and ligand molecule features. This approach not only captures subtle molecular structural changes that lead to activity cliffs but also significantly enhances the performance of molecular property prediction.

Macro-level Target Semantic (MTS) Guidance

The Macro-level Target Semantic (MTS) Guidance aims to extract dynamic semantic information from the global context of receptor proteins to align ligand features F_{mol} with target features F_{target}, facilitating the learning of the mapping f_m from x to y. MTS uses the Feature Abstractor to facilitate the extraction of global semantic information from the receptor features. The Feature Abstractor compresses the receptor features to capture their key semantic characteristics. Specifically, it performs average pooling on the input target features F_{target}, reducing the feature dimensions while extracting global semantic information, thus providing simplified.

$$F_{\text{compress}} = \text{AvgPool}(F_{\text{target}}) \quad (8)$$

Then, the global semantic embedding is passed through a Weight Regressor to generate dynamic conditional weights (γ_i, β_i), which are used to adjust the distribution of ligand features. The Weight Regressor module is responsible for generating these weights through a linear transformation, with the core strategy being the linear transformation of global semantic embeddings. This allows the model to learn how to adjust the feature distribution at each layer. Specifically, the Weight Regressor generates a set of dynamic conditional weights based on the input global semantic information, which are then used to modulate the ligand features via Adaptive Layer Normalization (AdaLN), providing dynamic adaptation to different inputs and enhancing the model's flexibility and performance.

$$\gamma_1, \beta_1, \gamma_2, \beta_2, \gamma_3, \beta_3 = \text{Linear}(F_{\text{compress}}) \quad (9)$$

In each layer, the ligand features F_{mol} are refined using Adaptive Layer Normalization (AdaLN), incorporating the conditional weights γ_i and β_i generated by the Weight Regressor to dynamically adjust the feature distribution:

$$\text{LN}_i^{\text{cond}} = \text{LN}_i(\gamma = \gamma_i, \beta = \beta_i) \quad (10)$$

The conditionally normalized ligand features are passed into a self-attention (SA) module to extract global contextual representations of the ligand.

$$\Phi_{\text{MTS}}(F_{\text{mol}}, F_{\text{target}}) = \text{SA}(F_{\text{mol}} \mid \text{LN}_i = \text{LN}_i^{\text{cond}}) \quad (11)$$

Through this process, the Macro-level Target Semantic (MTS) Guidance dynamically adjusts ligand features to align them with the global semantic context of the receptor. This alignment ensures that the ligand representation captures the global semantic patterns induced by the receptor, thereby supporting further optimization of ligand features and improving the extraction of protein-ligand interaction features.

Micro-level Pocket Semantic (MPS) Guidance

The Micro-level Pocket Semantic (MPS) Guidance focuses on the receptor's binding pocket region to capture local interaction patterns between receptors and ligands. By integrating the binding pocket features F_{pocket} with ligand features F_{mol}, the MPS Guidance refines the mapping f_m from x to y, enhancing the model's understanding of localized receptor-ligand interactions.

Specifically, for the target features $F_{\text{target}} \in \mathbb{R}^{n \times d}$ and ligand features $F_{\text{mol}} \in \mathbb{R}^{m \times d}$, the binding pocket features $F_{\text{pocket}} \in \mathbb{R}^{p \times d}$ are extracted from F_{target} using Pocket Prior Extractor (e.g., Cavity Plus [Xu *et al.*, 2018]). These pocket features are then combined with the ligand features F_{mol} through a cross-attention mechanism to achieve a deep interaction.

In the cross-attention mechanism, the molecule features F_{mol} are used to compute the query vector Q_{mol}, while the ligand features F_{mol} and pocket features F_{pocket} are concatenated to generate the key vector K_{pocket} and value vector V_{pocket}. The attention weights are computed using the scaled dot-product attention mechanism:

$$\text{Attention}(Q_{mol}, K_{pocket}, V_{pocket}) = \text{Softmax}\left(\frac{Q_{mol}K_{pocket}^{\top}}{\sqrt{d_k}}\right)V_{pocket} \quad (12)$$

where $Q_{mol} = W_q F_{\text{mol}}, K_{pocket} = W_k F_{\text{pocket}}, V_{pocket} = W_v F_{\text{pocket}}$ and $W_q, W_k, W_v \in \mathbb{R}^{d \times d}$ are learnable linear transformation matrices that extract the corresponding representations from the ligand and pocket features.

Through this process, the MPS Guidance effectively integrates ligand features F_{mol} and pocket features F_{pocket}, capturing the fine-grained interaction patterns between the receptor binding pocket and the ligand. By focusing on relevant molecular substructures, it enhances the model's sensitivity to subtle changes in molecular activity.

$$\Phi_{\text{MPS}}(F_{\text{mol}}, F_{\text{target}}) = \text{CrossAttention}(F_{\text{mol}}, F_{\text{pocket}}) \quad (13)$$

The MPS Guidance provides localized semantic alignment, complementing the global contextual representations generated by the Macro-level Target Semantic (MTS) Guidance,

ensuring a comprehensive understanding of receptor-ligand interactions.

4 Experiments

4.1 Data and Experimental Setups

All evaluations in this section are conducted on datasets from MoleculeACE (Activity Cliff Estimation) [Van Tilborg *et al.*, 2022], an open-access benchmarking platform available on GitHub at https://github.com/molML/MoleculeACE. This platform provides over 35,000 molecules distributed across 30 macromolecular targets, each corresponding to a separate dataset. Among these, 12 datasets contain fewer than 1,000 molecules in the training set, making MoleculeACE particularly suitable for evaluating model performance in low-data regimes.

In our experiments, we use the MTPNet framework, which integrates the Multi-Grained Target Perception (MTP) Module to dynamically model receptor-ligand interactions. MTP leverages both macro-level receptor context and micro-level binding pocket features to refine ligand representations, enabling precise activity cliff estimation. This receptor-aware design ensures that MTPNet can effectively capture the complex interplay between ligands and receptors, even in scenarios with limited data.

4.2 Performance Evaluation and Comparison

We first compared MTPNet with various machine learning (ML) and deep learning (DL) baseline algorithms on the MoleculeACE dataset, focusing on evaluating the prediction performance of activity cliff molecules. The specific results are presented in Figure 1 and Appendix. The analysis results indicate that MTPNet outperforms all baseline methods across the 30 activity cliff datasets, achieving the lowest RMSE values. Notably, MTPNet exhibits an average improvement of 18.95%, which is significantly higher than the performance gains achieved by current state-of-the-art (SOTA) pretraining methods. This finding emphasizes the importance of integrating receptor protein information in effectively addressing activity cliff tasks.

To validate the robustness and effectiveness of the MTP module, we conducted an evaluation on 30 activity cliff datasets within MoleculeACE, focusing on its impact as a plug-and-play enhancement for baseline models such as GCN [Kipf and Welling, 2017], GAT [Veličković *et al.*, 2017], GIN [Xu *et al.*, 2019], GraphTrans [Wu *et al.*, 2021], MolCLR [Wang *et al.*, 2022], and Mole-BERT [Xia *et al.*, 2023b]. As shown in Table 1, integrating the MTP module resulted in consistent improvements, with PCC increasing by 11.6%, R^2 by 17.8%, and RMSE decreasing by 19.0%. Additionally, we conducted a "scale-up" experiment by expanding the baseline models to match the parameter size of their MTP-augmented counterparts without incorporating the MTP module. The results indicate that while parameter scaling (e.g., increasing GCN's parameter size from 1.11M to 3.17M) slightly reduced RMSE from 0.950 to 0.915, the improvement was limited and still fell far short of the performance achieved by incorporating the MTP module (RMSE further reduced to 0.744). These findings demonstrate that the MTP module excels in capturing multi-level semantic information and structural nuances between molecules and receptors, far surpassing the benefits of merely increasing model size, thereby highlighting the unique strengths of the MTP module design.

Model	PCC ↑	R^2 ↑	RMSE ↓	Param. ↓
$GCN_{ICLR2017}$	0.711	0.501	0.950	1.11M
w/ Scale up	0.737	0.541	0.915	3.17M
w/ MTP	**0.837**	**0.693**	**0.744**	**3.83M**
$GAT_{ICLR2018}$	0.706	0.497	0.956	6.69M
w/ Scale up	0.715	0.506	0.945	9.85M
w/ MTP	**0.828**	**0.683**	**0.756**	**9.42M**
$GIN_{ICLR2019}$	0.718	0.509	0.941	2.25M
w/ Scale up	0.731	0.535	0.922	5.02M
w/ MTP	**0.826**	**0.674**	**0.767**	**4.97M**
$GraphTrans_{NIPS2021}$	0.719	0.515	0.936	3.43M
w/ Scale up	0.751	0.559	0.897	6.88M
w/ MTP	**0.828**	**0.676**	**0.765**	**6.15M**
$MolCLR_{NMI2023}$	0.715	0.504	0.946	2.04M
w/ Scale up	0.724	0.526	0.929	4.53M
w/ MTP	**0.822**	**0.668**	**0.774**	**4.76M**
$Mole\text{-}BERT_{ICLR2023}$	0.721	0.504	0.947	2.34M
w/ Scale up	0.742	0.546	0.906	4.78M
w/ MTP	**0.845**	**0.703**	**0.733**	**5.07M**

Table 1: Ablation results of baseline models under layer scale-up versus MTP Module augmentation.

In addition to these evaluations, we also assessed the performance of MTPNet in classification tasks. Specifically, we conducted experiments on the CYP3A4 dataset [Rao *et al.*, 2022] (see Table 2), which includes activity cliff data of Cytochrome P450 3A4 inhibitors/substrates experimentally measured by Veith et al. (2009) [Veith *et al.*, 2009], comprising 3,626 active inhibitors/substrates and 5,496 inactive compounds. The results show that the MTPNet achieved an AUC of 0.924, significantly outperforming baseline models like Mole-BERT (AUC = 0.902) and MolCLR (AUC = 0.896). These results demonstrate that MTPNet excels not only in regression tasks but also in classification tasks, effectively capturing activity cliffs and showcasing its potential for drug discovery and molecular activity prediction.

Model	AUC
$GCN_{ICLR2017}$	0.766
$GAT_{ICLR2018}$	0.773
$SemiMol_{IJCAI2024}$	0.857
$GraphTrans_{NIPS2021}$	0.890
$Mole\text{-}BERT_{ICLR2023}$	0.902
$MolCLR_{NMI2023}$	0.896
Ours	**0.924**

Table 2: Comparison experiments on CYP3A4 dataset.

4.3 Ablation Study

To evaluate the contribution of each component in MTPNet, we conducted an ablation study (see Table 3) comparing model performance with and without the Adaptive LayerNorm (AdaLN) and Cross Attention (CA) modules. Remov-

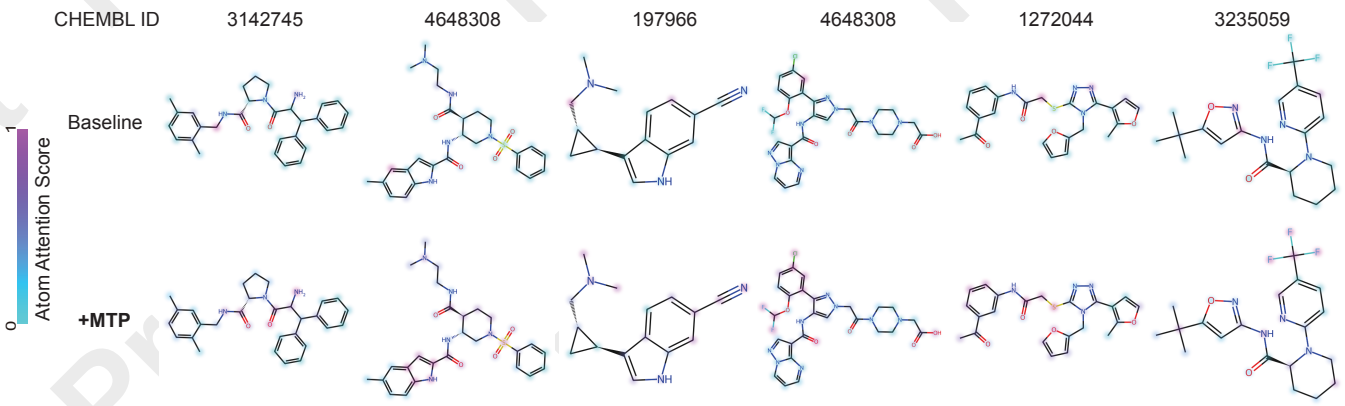

Figure 4: Visualization of Atom Attention Scores for Molecules in CHEMBL v29. Figure 4 compares the atom attention distributions between the baseline model and the MTP module for six representative molecules (CHEMBL IDs: 3142745, 4648308, 197966, 4648308, 1272044, and 3235059) from the CHEMBL v29 dataset. The attention scores are color-coded, where purple indicates high attention and blue indicates low attention, as shown in the color bar on the left. The MTP module demonstrates a significantly more focused and meaningful attention distribution, effectively identifying key functional groups (e.g., amino groups, carbonyl groups) and specific chemical bonds (e.g., double bonds, triple bonds) that play critical roles in activity cliff phenomena. In contrast to the baseline model's dispersed attention patterns, the MTP module achieves more focused and precise attention, effectively capturing key molecular interactions and uncovering the chemical principles behind activity cliff phenomena.

ing AdaLN is equivalent to disabling Macro-level Target Semantic (MTS) guidance, while removing CA corresponds to disabling Micro-level Pocket Semantic (MPS) guidance. The results show that removing either the AdaLN or CA module leads to a decline in PCC, R^2, and RMSE, which underscores the critical role of MTS guidance and MPS guidance in capturing global receptor semantics and localized receptor–ligand interactions.

Model	PCC ↑	R^2 ↑	RMSE ↓
w/o MTS & MPS	0.737	0.526	0.917
w/o MTS	0.830	0.680	0.760
w/o MPS	0.826	0.676	0.766
Ours	**0.845**	**0.703**	**0.733**

Table 3: Ablation results on the MTP Module sub-components (MTS and MPS).

4.4 Interpretation

MTPNet, by integrating global and local semantic guidance, accurately captures the interaction patterns between key functional groups and chemical bonds within molecules, thereby uncovering the chemical essence of activity cliff phenomena. Figure 4 shows that the MTP module assigns significant attention to key functional groups such as amino groups (NH2), carbonyl groups (C=O), sulfonyl groups (O=S=O), carboxyl groups (COOH), and halogen groups, as well as to specific chemical bonds such as double and triple bonds. Its attention distribution is far superior to the dispersed attention of baseline models. These functional groups and chemical bonds play critical roles in molecular binding affinity and activity changes with receptors. For instance, amino groups influence molecular activity by forming hydrogen bonds, sulfonyl groups regulate molecular solubility and receptor-binding stability due to their strong polarity, and halogen groups and carboxyl groups significantly impact molecular behavior by modulating hydrophobicity and acidity, respectively. For molecules containing both key functional groups and specific chemical bonds, the MTP module generally assigns greater attention to functional groups than to chemical bonds. This observation supports an important principle in chemistry: functional groups are more critical than chemical bonds in determining molecular properties and reactivity. By capturing this pattern, the MTP module demonstrates strong interpretability, even reflecting fundamental chemical principles. In summary, the MTP module significantly enhances the interpretability of activity cliff predictions, accurately identifying key activity sites within molecules and revealing the core influence of functional groups and chemical bonds on molecular behavior. This has profound implications for the study of chemical reaction mechanisms and protein-ligand binding rules, while also offering novel perspectives and approaches for understanding the causes of activity cliffs and exploring complex receptor-ligand interaction mechanisms.

5 Conclusion

By introducing the interactions of molecules and their target proteins as the guidance, MTPNet achieves unified prediction across diverse downstream tasks related to activity cliffs. Through internalizing interaction patterns at different granularities, MTPNet comprehensively outperforms other methods on 30 representative datasets. In addition, when MTPNet is adopted as a plug-in, the prediction performance of multiple mainstream GNN architectures is significantly improved, showing ideal usability and robustness. Looking to the future, MTPNet is expected to achieve more efficient hit-to-lead optimization to accelerate drug design.

Ethical Statement

There are no ethical issues.

Acknowledgments

This work was supported in part by the Shenzhen Medical Research Funds in China (No. B2302037), Natural Science Foundation of China (No. 61972217, 32071459, 62176249, 62006133, 62271465), and AI for Science (AI4S)-Preferred Program, Peking University Shenzhen Graduate School, China.

Contribution Statement

This work was a collaborative effort by all contributing authors. Zishan Shu, Yufan Deng, and Hongyu Zhang made equal contributions to this study and are designated as co-first authors. Zhiwei Nie and Jie Chen, serving as the corresponding authors, are responsible for all communications related to this manuscript.

References

[Brandes *et al.*, 2022] Nadav Brandes, Dan Ofer, Yam Peleg, Nadav Rappoport, and Michal Linial. Proteinbert: a universal deep-learning model of protein sequence and function. *Bioinformatics*, 38(8):2102–2110, 2022.

[Butler *et al.*, 2018] Keith T. Butler, Daniel W. Davies, Hugh Cartwright, Olexandr Isayev, and Aron Walsh. Machine learning for molecular and materials science. *Nature*, pages 547–555, July 2018.

[Chen *et al.*, 2022] Hengwei Chen, Martin Vogt, and Jürgen Bajorath. Deepac – conditional transformer-based chemical language model for the prediction of activity cliffs formed by bioactive compounds. *Digital Discovery*, 1(6):898–909, 2022.

[Consonni and Todeschini, 2010] Viviana Consonni and Roberto Todeschini. *Molecular Descriptors*, pages 29–102. January 2010.

[Dong *et al.*, 2018] Jie Dong, Ning-Ning Wang, Zhi-Jiang Yao, Lin Zhang, Yan Cheng, Defang Ouyang, Ai-Ping Lu, and Dong-Sheng Cao. Admetlab: a platform for systematic admet evaluation based on a comprehensively collected admet database. *Journal of Cheminformatics*, 10(1), December 2018.

[Drucker *et al.*, 1996] H. Drucker, C. J. Burges, L. Kaufman, A. Smola, and V. Vapnik. Support vector regression machines. In *Advances in Neural Information Processing Systems*, volume 9, 1996.

[Du *et al.*, 2024] Wenjie Du, Shuai Zhang, Jun Xia Di Wu, 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*, pages 5808–5816, 2024.

[Fang *et al.*, 2023] Y. Fang, Y. Jiang, L. Wei, Q. Ma, Z. Ren, Q. Yuan, and D. Q. Wei. Deepprosite: structure-aware protein binding site prediction using esmfold and pretrained language model. *Bioinformatics*, 39(12):btad718, December 2023.

[Iqbal *et al.*, 2021] J. Iqbal, M. Vogt, and J. Bajorath. Prediction of activity cliffs on the basis of images using convolutional neural networks. *Journal of Computer-Aided Molecular Design*, pages 1–8, 2021.

[Kipf and Welling, 2017] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017.

[Li *et al.*, 2023] Peiying Li, Yongchang Liu, Shikui Tu, and Lei Xu. Glpocket: A multi-scale representation learning approach for protein binding site prediction. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, pages 4821–4828, 2023.

[Lin *et al.*, 2023] Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, Allan dos Santos Costa, Maryam Fazel-Zarandi, Tom Sercu, Salvatore Candido, and Alexander Rives. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023.

[Meng *et al.*, 2024] Ziqiao Meng, Liang Zeng, Zixing Song, Tingyang Xu, Peilin Zhao, and Irwin King. Towards geometric normalization techniques in se (3) equivariant graph neural networks for physical dynamics simulations. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, pages 5981–5989, 2024.

[Moriwaki *et al.*, 2018] Hirotomo Moriwaki, Yu-Shi Tian, Norihito Kawashita, and Tatsuya Takagi. Mordred: a molecular descriptor calculator. *Journal of Cheminformatics*, December 2018.

[Nie *et al.*, 2024a] Zhiwei Nie, Daixi Li, Jie Chen, Fan Xu, Yutian Liu, Jie Fu, Xudong Liu, Zhennan Wang, Yiming Ma, Kai Wang, et al. Hunting for peptide binders of specific targets with data-centric generative language models. *bioRxiv*, 2024.

[Nie *et al.*, 2024b] Zhiwei Nie, Hongyu Zhang, Hao Jiang, Yutian Liu, Xiansong Huang, Fan Xu, Yonghong Tian, Jie Chen, and Wen-Bin Zhang. Multi-purpose enzyme-substrate interaction prediction with progressive conditional deep learning. *Research Square PREPRINT*, 2024.

[Park *et al.*, 2022] Junhui Park, Gaeun Sung, SeungHyun Lee, SeungHo Kang, and ChunKyun Park. Acgcn: Graph convolutional networks for activity cliff prediction between matched molecular pairs. *Journal of Chemical Information and Modeling*, 62(10):2341–2351, 2022.

[Rao *et al.*, 2022] Jiahua Rao, Shuangjia Zheng, Yutong Lu, and Yuedong Yang. Quantitative evaluation of explainable graph neural networks for molecular property prediction. *Patterns*, 3(12):100628, 2022.

[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. *arXiv: Biomolecules,arXiv: Biomolecules*, June 2020.

[Shin *et al.*, 2024] Dong-Hee Shin, Young-Han Son, Deok-Joong Lee, Ji-Wung Han, and Tae-Eui Kam. Dynamic many-objective molecular optimization: Unfolding complexity with objective decomposition and progressive optimization. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, pages 6026–6034, 2024.

[Su *et al.*, 2023] Jin Su, Chenchen Han, Yuyang Zhou, Junjie Shan, Xibin Zhou, and Fajie Yuan. Saprot: Protein language modeling with structure-aware vocabulary, October 2023. bioRxiv 2023.10.01.560349.

[Van Tilborg *et al.*, 2022] Derek Van Tilborg, Alisa Alenicheva, and Francesca Grisoni. Exposing the limitations of molecular machine learning with activity cliffs. *Journal of chemical information and modeling*, 62(23):5938–5951, 2022.

[Vapnik, 2013] Vladimir Vapnik. *The Nature of Statistical Learning Theory*. Springer Science & Business Media, 2013.

[Veith *et al.*, 2009] Henrike Veith, Noel Southall, Ruili Huang, Tim James, Darren Fayne, Natalia Artemenko, Min Shen, James Inglese, Christopher P. Austin, David G. Lloyd, and et al. Comprehensive characterization of cytochrome p450 isozyme selectivity across chemical libraries. *Nature Biotechnology*, 27(11):1050–1055, 2009.

[Veličković *et al.*, 2017] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.

[Wang *et al.*, 2017] Yanli Wang, Stephen H. Bryant, Tiejun Cheng, Jiyao Wang, Asta Gindulyte, Benjamin A. Shoemaker, Paul A. Thiessen, Siqian He, and Jian Zhang. Pubchem bioassay: 2017 update. *Nucleic Acids Research*, pages D955–D963, January 2017.

[Wang *et al.*, 2022] Y. Wang, J. Wang, Z. Cao, et al. Molecular contrastive learning of representations via graph neural networks. *Nature Machine Intelligence*, 4:279–287, 2022.

[Wu *et al.*, 2021] Zhanghao Wu, Paras Jain, Matthew Wright, Azalia Mirhoseini, Joseph E Gonzalez, and Ion Stoica. Representing long-range context for graph neural networks with global attention. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 13266–13279. Curran Associates, Inc., 2021.

[Wu, 2024] Fang Wu. A semi-supervised molecular learning framework for activity cliff estimation. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, pages 6080–6088, 2024.

[Xia *et al.*, 2023a] Jun Xia, Lecheng Zhang, Xiao Zhu, Yue Liu, Zhangyang Gao, Bozhen Hu, Cheng Tan, Jiangbin Zheng, Siyuan Li, and Stan Z. Li. Understanding the limitations of deep models for molecular property prediction: Insights and solutions. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Advances in Neural Information Processing Systems*, volume 36, pages 64774–64792. Curran Associates, Inc., 2023.

[Xia *et al.*, 2023b] Jun Xia, Chengshuai Zhao, Bozhen Hu, Zhangyang Gao, Cheng Tan, Yue Liu, Siyuan Li, and Stan Z. Li. Mole-bert: Rethinking pre-training graph neural networks for molecules. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023.

[Xiang *et al.*, 2024] Hongxin Xiang, Shuting Jin, Jun Xia, Man Zhou, Jianmin Wang, Li Zeng, and Xiangxiang Zeng. An image-enhanced molecular graph representation learning framework. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, pages 6107–6115, 2024.

[Xie and Fu, 2025] Jiaqing Xie and Tianfan Fu. Deepprotein: Deep learning library and benchmark for protein sequence learning. *Bioinformatics*, May 2025.

[Xu *et al.*, 2018] Y. Xu, S. Wang, Q. Hu, S. Gao, X. Ma, W. Zhang, Y. Shen, F. Chen, L. Lai, and J. Pei. Cavityplus: a web server for protein cavity detection with pharmacophore modelling, allosteric site identification and covalent ligand binding ability prediction. *Nucleic Acids Research*, 46(W1):W374–W379, 2018.

[Xu *et al.*, 2019] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *International Conference on Learning Representations*, 2019.

[Yang *et al.*, 2023] Xixi Yang, Li Fu, Yafeng Deng, Yuansheng Liu, Dongsheng Cao, and Xiangxiang Zeng. Gpmo: Gradient perturbation-based contrastive learning for molecule optimization. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, pages 4940–4948, 2023.

[Zhao *et al.*, 2023] Ziyuan Zhao, Peisheng Qian, Xulei Yang, Zeng Zeng, Cuntai Guan, Wai Leong Tam, and Xiaoli Li. Semignn-ppi: Self-ensembling multi-graph neural network for efficient and generalizable protein-protein interaction prediction. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, pages 4984–4992, 2023.

[Zheng *et al.*, 2024] Yan Zheng, Song Wu, Junyu Lin, Yazhou Ren, Jing He, Xiaorong Pu, and Lifang He. Cross-view contrastive fusion for enhanced molecular property prediction. In *Proccedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, 2024.

[Zhu *et al.*, 2023] Yiheng Zhu, Zhenqiu Ouyang, Ben Liao, Jialu Wu, Yixuan Wu, Chang-Yu Hsieh, Tingjun Hou, and Jian Wu. Molhf: A hierarchical normalizing flow for molecular graph generation. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, pages 5002–5010, 2023.