DRUG-FEW: A FEW-SHOT LEARNING METHOD FOR TARGET-SPECIFIC DRUG VIRTUAL SCREENING WITH INTERACTION-INFORMED ADAPTATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Few-shot virtual screening (VS), which aims to identify active molecules for a target with only a few known ligands, is crucial for accelerating drug discovery in cases where experimental data are extremely limited. Traditional ligand-based few-shot learning methods rely on molecular similarity or latent embeddings to generalize from these limited examples, but they often fail to capture target-ligand interactions, leading to overlooked active molecules. Here, we present Drug-few, the first few-shot learning framework designed for strict target-specific VS that explicitly incorporates binding-relevant information. Drug-few introduces prompt tokens for rapid target-specific adaptation and incorporates lightweight adapter modules to refine pocket-ligand representations. These components are combined in a Gated Prompt Adapter (GPA), where the contribution of prompt tokens is dynamically modulated by interaction-aware signals. Extensive experiments on three benchmark datasets show that Drug-few consistently outperforms the zero-shot baseline, maintaining strong retrieval performance even when the target active molecules are dissimilar to the known ligands, demonstrating its ability to generalize to novel molecules.

1 INTRODUCTION

In drug discovery, virtual screening (VS) aims to prioritize candidate molecules for a given target from large compound libraries, accelerating the identification of potential bioactive compounds (Patel et al., 2021; Schneider, 2010; Sadybekov & Katritch, 2023). In certain screening scenarios, a target may be associated with a few known "hit" compounds, which are called ligands (Edwards & Owen, 2025; Li et al., 2025; Zhang et al., 2025). How to leverage such limited information to discover new active molecules across much larger chemical space is highly important (Eckmann et al., 2024). For instance, the orphan receptor GPR6, linked to Parkinson's disease, was only recently crystallized with two inverse agonists, yet the number of known ligands remains very limited, hindering novel modulator discovery (Barekatain et al., 2024). Consequently, there is a pressing need to develop strategies that can effectively tackle this challenging problem.

Traditional ligand-based drug discovery method relies on molecular fingerprints and similarity metrics, such as the Tanimoto coefficient, to rank candidates based on their structural resemblance to known ligands (Bajusz et al., 2015; Rogers & Hahn, 2010; Khan et al., 2016). With the rise of deep learning, few-shot ligand-based methods have recently emerged (Stanley et al., 2021; Li et al., 2025; Schimunek et al., 2023). Some meta-learning methods frame (Li et al., 2025) the problem as a series of tasks, each being a binary classification of active versus inactive compounds for a specific target, following a task-support-query paradigm. Others employ embedding- or metric-based approaches (Schimunek et al., 2023; Eckmann et al., 2024) to directly compare molecules in a learned latent space. In all cases, the focus is on generalizing molecular properties from limited examples rather than capturing the pocket-ligand interaction patterns. In other words, the molecules selected by these ligand-based models may be statistically similar to known ligands but do not necessarily bind effectively to the target (Sundar & Colwell, 2020; Scior et al., 2012). Essentially, these studies are aimed at molecular property prediction or classification, rather than addressing the practical challenges of target-specific VS.

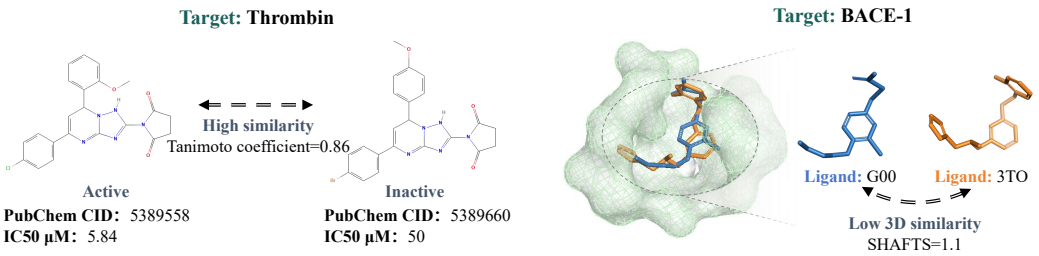


Figure 1: (a) Case with similar structure but different activity: Compound CID 5389558 and CID 5389660 exhibit nearly identical structures (Tanimoto coefficient = 0.86) but show markedly different thrombin inhibitory activities (IC50 μM = 5.84 μM vs. 50), representing a typical activity cliff. (b) Case with dissimilar structure but similar activity: Ligands G00 and 3TO both inhibit BACE-1, although their 3D similarity by SHAFTS is only 1.1.

Next, we provide concrete examples demonstrating these issues in practice. First, structurally similar compounds can exhibit strikingly different potencies, a phenomenon widely recognized as activity cliffs (Hu et al., 2013). For example, as shown in Figure 1a, compound CID 5389558 and CID 5389660 share high structural similarity (Tanimoto coefficient of 0.86), yet their thrombin inhibitory activities differ markedly: the former is active with an IC50 of 5.84 μM, whereas the latter is inactive with an IC50 greater than 50 μM. Second, ligands with very different scaffolds may still share comparable activity by adopting similar binding modes (Xu & Zou, 2021). As illustrated in Figure 1b, ligand G00 and ligand 3TO both inhibit BACE-1, but their 3D similarity score by SHAFTS is only 1.1, below the commonly used threshold of 1.2 that indicates significant 3D similarity. Notably, the RMSD between them is 1.42 Å, indicating that despite low molecular resemblance, the two ligands engage the target in similar ways. These examples demonstrate that relying solely on ligand similarity can both misclassify truly actives and overlook other actives, highlighting the necessity of explicitly modeling target-ligand interactions in practical VS.

In this work, we are the first to focus on few-shot learning for realistic target-specific virtual screening scenarios with binding-relevant information, where the goal is to identify active molecules for a given target with only a few available ligands. However, performing few-shot adaptation directly within conventional VS frameworks remains challenging. Approaches such as docking- or energy-based methods require modeling accurate docking conformations, which makes them computationally demanding and unsuitable when only a few known ligands are available for adaptation (Cai et al., 2024; Zhang et al., 2023b; McNutt et al., 2021). On the other hand, recent feature alignment based retrieval architectures (e.g., DrugCLIP (Gao et al., 2024a)) offer an elegant workaround. By embedding pockets and ligands into a shared space, they bypass explicit docking and affinity evaluation, enabling molecules to be ranked directly by feature similarity. In fact, this CLIP paradigm has already proven highly effective for few-shot learning in natural language processing (NLP) and computer vision (CV). Nevertheless, transferring such strategies to VS scenario is nontrivial, as each atom can critically influence binding, unlike image-text pairs, where local information is often redundant. Thus, an effective few-shot VS method requires not only modulation based on known ligands, but also sensitivity to fine-grained target-ligand interactions.

Building on these observations, we propose **Drug-few**, a composite adaptation strategy within the DrugCLIP paradigm for few-shot target-specific VS. In our design, prompt tokens are introduced to enable rapid target-specific adaptation from a few known ligands, while lightweight adapter modules are incorporated to refine pocket-molecule representations and capture subtle interaction features. Crucially, the two components are not simply combined; instead, we design a Gated Prompt Adapter (GPA) module, where the contribution of prompt tokens is dynamically modulated by interaction-aware signals, ensuring that adaptation is guided by binding-relevant information rather than generic similarity. In this way, Drug-few effectively balances efficiency and representational power, achieving robust few-shot VS even under high scaffold diversity.

Our contributions can be summarized as follows:

1.To our knowledge, Drug-few is the first few-shot learning method for strict target structure-aware VS that explicitly incorporates target-ligand interactions, moving beyond purely ligand-centric similarity.

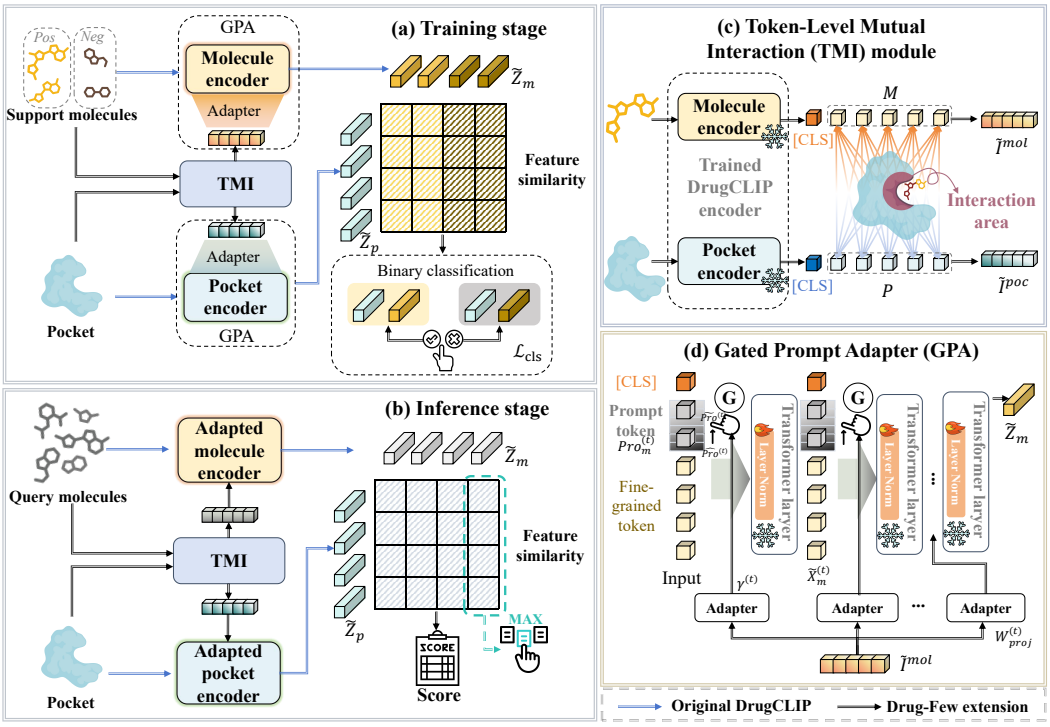


Figure 2: Overview of Drug-few. (a) Training stage. In the original DrugCLIP (blue arrows), pocket and molecules are encoded, and contrastive learning aligns their embeddings into a shared representation space, followed by feature similarity matching. In Drug-few, the modifications to the original DrugCLIP are indicated by black arrows. For a new target with only a few known molecules, the Gated Prompt Adapter (d) injects prompt tokens and lightweight adapters to condition the model on interaction-aware information, while the Token-Level Mutual Interaction module (c) further refines pocket-molecule representations to dynamically regulate the contribution of prompt tokens. (b) Inference stage. Query molecules are encoded in the same way, and their similarity to the adapted target representation is used to rank candidates for target-specific VS.

2. We design a hybrid adaptation mechanism that incorporates prompt tokens guided by fine-grained interaction signals to enable efficient task-specific adaptation.

3. We conduct extensive experiments on three widely used VS benchmarks, and Drug-few consistently outperforms the zero-shot baseline. In particular, under the scaffold and feature-cluster splits, where the known actives are dissimilar to the query actives, Drug-few still maintains strong retrieval performance, showing its ability to identify novel molecules.

2 METHOD

2.1 OVERVIEW OF DRUG-FEW

Given a protein pocket p and a set of n candidate molecules $\{m_1, m_2, \ldots, m_n\}$, the goal of few-shot VS is to rank these candidate molecules according to their likelihood of binding to p. In practice, a few known binding molecules are provided for a given target. We refer to these as support molecules, while the molecules to be evaluated are query molecules.

As shown in Figure 2, Drug-few performs few-shot learning for training and inference using support set and query set, respectively. Both the pocket and molecules are projected into a shared embedding space via contrastive learning. To enable adaptation to the given pocket target with only a few known ligands, we introduce two key components on the base model DrugCLIP (Gao et al., 2024a): the Gated Prompt Adapter (GPA), which injects learnable prompt tokens and lightweight adapters; and the Token-Level Mutual Interaction (TMI) module, which computes fine-grained feature interactions between pockets and molecules.

2.2 BASE MODEL

Our method builds upon DrugCLIP (Gao et al., 2024a) and extends it to the few-shot virtual screening setting. Specifically, pockets and molecules are tokenized at the atomic level, where atom types form the token features and 3D spatial information is encoded as pairwise representations. The encoded inputs are processed separately by the Uni-Mol (Zhou et al., 2023) molecule encoder and pocket encoder, each with a backbone built from stacked Transformer layers, which integrates atomic embeddings with pairwise 3D distance encodings.

Specifically, the Pocket encoder $\phi_p(\cdot)$ and Molecule encoder $\phi_m(\cdot)$, together with projection heads $\sigma_p(\cdot)$ and $\sigma_m(\cdot)$, map inputs directly into a shared embedding space: $z_p = \sigma_p(\phi_p(p)), \quad z_m = \sigma_m(\phi_m(m))$. To align the two modalities, DrugCLIP adopt a contrastive objective. For a batch of N binding pairs, these pockets and molecules can be combined to form N^2 pairs $\{(p_i, m_j)\}_{i,j=1}^{N}$, for which the cosine similarity is computed as $S(p_i, m_j) = \frac{Z_p^{(i)\top} Z_m^{(j)}}{\|Z_p^{(i)}\| \cdot \|Z_m^{(j)}\|}$.

2.3 FEW-SHOT ADAPTATION FRAMEWORK

Building on the trained DrugCLIP dual-encoder framework, we propose a few-shot learning method to enable fine-grained adaptation of pocket-molecule interactions. This extension introduces two key components: Token-Level Mutual Interaction (TMI) module and Gated Prompt Adapter (GPA).

2.3.1 TOKEN-LEVEL MUTUAL INTERACTION (TMI) MODULE

As illustrated in Figure 2c, the tokenized pocket and molecule are first embedded using the frozen trained DrugCLIP encoders, namely the Molecule encoder and the Pocket encoder. This produces token-level representations. To focus on fine-grained contextual information, the [CLS] token is removed, yielding refined representations:

$$M \in \mathbb{R}^{B_m \times (L_m - 1) \times D}, \qquad P \in \mathbb{R}^{B_p \times (L_p - 1) \times D}, \tag{1}$$

where B_m and B_p denote the molecule and pocket batch sizes, L_m and L_p are the original token counts (including [CLS]), and D is the embedding dimension. We index the remaining token positions by $e \in \{1, \ldots, L_m - 1\}$ for molecules and $f \in \{1, \ldots, L_p - 1\}$ for pockets (so the second dimension of M, P is indexed by e, f respectively). Next, token-wise interactions are computed by element-wise multiplication for all pocket–molecule token pairs:

$$I_{b_m, b_p, e, f, d} \;=\; M_{b_m, e, d} \odot P_{b_p, f, d} \;\in\; \mathbb{R}^{B_m \times B_p \times (L_m - 1) \times (L_p - 1) \times D}, \tag{2}$$

where $d \in \{1, \ldots, D\}$ indexes embedding dimensions and $\odot$ denotes element-wise multiplication.

The token-wise interactions are then averaged over the token-pair dimensions while keeping the embedding channels:

$$\bar{I}_{b_m, b_p, d} \;=\; \frac{1}{(L_m - 1)(L_p - 1)} \sum_{e=1}^{L_m - 1} \sum_{f=1}^{L_p - 1} I_{b_m, b_p, e, f, d} \;\in\; \mathbb{R}^{B_m \times B_p \times D}. \tag{3}$$

Finally, obtain interaction-driven per-molecule and per-pocket representations by averaging across the opposite batch dimension:

$$\tilde{I}^{mol}_{b_m, d} \;=\; \frac{1}{B_p} \sum_{b_p=1}^{B_p} \bar{I}_{b_m, b_p, d} \quad \Rightarrow \quad \tilde{I}^{mol} \in \mathbb{R}^{B_m \times D}, \tag{4}$$

$$\tilde{I}^{poc}_{b_p, d} \;=\; \frac{1}{B_m} \sum_{b_m=1}^{B_m} \bar{I}_{b_m, b_p, d} \quad \Rightarrow \quad \tilde{I}^{poc} \in \mathbb{R}^{B_p \times D}. \tag{5}$$

Notably, the TMI module introduces no trainable parameters, keeping the few-shot adaptation stage lightweight and efficient.

Table 1: Few-shot virtual screening results on the DUD-E benchmark compared with zero-shot baseline under 2-, 4-, 8-, and 16-shot settings. **Bold** indicates better performance.

Sample	Method	AUROC (%)	BEDROC (%)	EF 0.5%	EF 1%	EF 5%
2	Zero-shot	80.67±0.03	49.85±0.07	39.28±0.08	32.15±0.07	10.65±0.01
	Drug-few	**82.35±0.04**	**52.13±0.05**	**40.90±0.23**	**33.65±0.08**	**11.10±0.02**
4	Zero-shot	80.74±0.01	49.79±0.03	39.75±0.03	32.41±0.02	10.66±0.01
	Drug-few	**84.78±0.24**	**56.20±0.07**	**43.95±0.13**	**36.80±0.62**	**11.89±0.06**
8	Zero-shot	80.73±0.01	49.45±0.07	40.53±0.06	32.87±0.07	10.66±0.01
	Drug-few	**88.54±0.37**	**61.95±0.43**	**48.79±0.15**	**41.66±0.23**	**13.20±0.12**
16	Zero-shot	80.76±0.04	49.11±0.04	42.79±0.12	33.78±0.01	10.70±0.03
	Drug-few	**93.16±0.11**	**71.70±0.13**	**58.17±0.34**	**50.22±0.21**	**15.40±0.08**

Table 2: Few-shot virtual screening results on the LIT-PCBA benchmark compared with zero-shot baseline under 2-, 4-, 8-, and 16-shot settings.

Sample	Method	AUROC (%)	BEDROC (%)	EF 0.5%	EF 1%	EF 5%
2	Zero-shot	57.30±0.32	6.29±0.33	8.71±0.91	**5.61±0.38**	**2.24±0.07**
	Drug-few	**58.04±0.72**	**6.43±0.52**	**8.86±1.56**	5.40±0.54	**2.24±0.32**
4	Zero-shot	57.52±0.09	6.36±0.26	**8.85±0.32**	**5.67±0.29**	2.30±0.14
	Drug-few	**59.88±0.74**	**6.74±0.46**	8.02±0.44	5.51±0.57	**2.61±0.17**
8	Zero-shot	57.22±0.75	5.97±1.06	8.24±2.74	5.57±1.47	2.18±0.22
	Drug-few	**61.11±1.36**	**6.64±0.51**	**8.69±2.13**	**5.81±0.57**	**2.27±0.17**
16	Zero-shot	58.19±2.12	7.46±3.13	11.21±6.38	6.68±3.36	2.46±0.70
	Drug-few	**67.03±1.72**	**9.90±3.91**	**12.41±8.34**	**8.16±3.91**	**3.69±1.09**

2.3.2 GATED PROMPT ADAPTER (GPA)

As shown in Figure 2d, GPA serves as a key adapter that inserts conditionally adapted prompt tokens into the original Transformer-stacked encoder. Its main function is twofold: first, to introduce learnable prompt tokens into the token-level input at every Transformer layer; and second, to adapt these prompt tokens using a lightweight adapter that incorporates interaction-informed signals derived from the TMI module. Together, this design enables the encoder to dynamically adjust its representations according to the specific pocket-molecule interaction. GPA adapts both the Pocket and Molecule encoders in same way; therefore, we describe the Molecule encoder as an example.

For each layer, L_{pro} prompt tokens $Pro_m^{(t)}$ are randomly initialized and inserted after the [CLS] token. Since the Uni-Mol (Zhou et al., 2023) encoder relies on both token and pairwise distance inputs, prompt tokens must also be defined in the pairwise space. Unlike real atoms, they lack 3D coordinates, and random assignment would add noise. To address this, we treat them in pairwise computation the same as [CLS]: only their pairwise bias is aligned with [CLS], while the token-level features remain randomly initialized. This design integrates prompt tokens into the pairwise representation without breaking geometric consistency. Lastly, $Pro_m^{(t)}$ are encoded to $\widehat{Pro}^{(t)}$.

To enable target-specific adaptation, a lightweight adapter $W_{proj}^{(t)}$ for t-th laryer takes the TMI-derived representation $\tilde{I}^{mol}$ as input and generates a scaling vector $\gamma^{(t)} \in \mathbb{R}^D$. This vector modulates the projected prompt embeddings $\widehat{Pro}^{(t)}$. through element-wise scaling, yielding the adapted features of prompt tokens: $\widetilde{Pro}^{(t)} = \widehat{Pro}^{(t)} \odot \gamma^{(t)}$, where $\odot$ denotes element-wise multiplication.

After being modulated, the processed features of prompt tokens are inserted immediately after the [CLS] token. At the t-th layer, the molecule representation is thus formulated as

$$\tilde{X}_m^{(t)} = [\tilde{x}_{[\text{cls}]}; \widetilde{Pro}^{(t)}; \tilde{x}_1, \ldots, \tilde{x}_{L_m-1}]. \tag{6}$$

The subsequent computation then follows the process of the original encoder layer in Uni-Mol Zhou et al. (2023). Similarly, the same procedure is applied within the Pocket encoder, ensuring a consistent representation framework for both molecules and pockets. After passing through the GPA-adapted encoders, the molecule and pocket features are updated to $\tilde{Z}_m$ and $\tilde{Z}_p$, respectively.

Table 3: Few-shot virtual screening results on the DEKOIS 2.0 benchmark compared with zero-shot baseline under 2-, 4-, 8-, and 16-shot settings.

Sample	Method	AUROC (%)	BEDROC (%)	EF 0.5%	EF 1%	EF 5%
2	Zero-shot	79.12±0.04	49.28±0.59	20.12±0.06	18.39±0.08	9.13±0.03
	Drug-few	**80.61±0.11**	**52.25±0.12**	**20.83±0.07**	**19.33±0.08**	**9.65±0.02**
4	Zero-shot	79.01±0.07	48.48±0.12	20.60±0.18	18.88±0.10	9.08±0.02
	Drug-few	**82.66±0.05**	**55.32±0.52**	**23.03±0.56**	**21.28±0.30**	**10.41±0.08**
8	Zero-shot	79.21±0.31	47.19±0.32	22.44±0.19	20.30±0.09	9.15±0.07
	Drug-few	**87.22±0.42**	**61.14±0.92**	**27.69±1.10**	**26.35±1.59**	**12.05±0.06**
16	Zero-shot	78.75±0.27	43.59±0.34	26.91±0.06	23.50±0.13	9.34±0.07
	Drug-few	**91.98±0.14**	**66.62±0.36**	**37.78±0.29**	**35.17±0.30**	**14.12±0.07**

Table 4: Few-shot virtual screening results on the CASF-2016 benchmark compared with zero-shot baseline under 2-, 4-, 8-, and 16-shot settings.

Sample	Method	AUROC (%)	BEDROC (%)	EF 0.5%	EF 1%	EF 5%
2	zero-shot	85.93±0.01	67.75±0.01	36.62±0.01	34.52±0.01	12.93±0.00
	Drug-few	**88.71±0.18**	**71.30±0.13**	**37.74±0.28**	**35.72±0.11**	**13.98±0.09**
4	zero-shot	85.93±0.00	67.70±0.02	36.73±0.00	34.59±0.01	12.93±0.00
	Drug-few	**91.32±0.34**	**74.03±1.53**	**38.78±0.83**	**36.83±0.85**	**14.55±0.12**
8	zero-shot	85.92±0.01	67.62±0.02	36.95±0.01	34.74±0.01	12.93±0.00
	Drug-few	**94.04±0.41**	**78.72±1.31**	**40.32±1.19**	**38.27±1.16**	**15.77±0.26**
16	zero-shot	85.94±0.02	67.49±0.03	37.38±0.01	35.09±0.00	12.94±0.01
	Drug-few	**97.34±0.46**	**87.12±0.32**	**42.65±0.34**	**41.69±0.29**	**17.54±0.10**

2.4 Training and Inference Stages

As shown in Pseudocode 1, the framework follows a two-stage process of training and inference. For each new target, the original DrugCLIP weights are first reloaded, then fine-tuned on the corresponding support set, and the updated model is subsequently used for inference on the query set.

2.4.1 Training Stage

Given a support set of molecules $\{m_k^s\}$, a query set $\{m_r^q\}$, and a pocket p, the inputs are first mapped to token-level embeddings and enriched with TMI-derived features. Subsequently, layer-wise prompt tokens are introduced through the GPA, producing adapted embeddings for the k-th molecule and the corresponding pocket, denoted as $\tilde{Z}_{m^s}^{(k)}$ and $\tilde{Z}_p$, respectively. Next, the similarity between each pocket-molecule pair is computed as $\tilde{S}_k = \frac{\tilde{Z}_p^\top \tilde{Z}_{m^s}^{(k)}}{\|\tilde{Z}_p\| \cdot \|\tilde{Z}_{m^s}^{(k)}\|}$.

Correspondingly, a binary label vector Y_k indicates the actives for pocket p. The classification loss summarizing the overall training objective is then defined as:

$$\mathcal{L}_{\text{cls}} = \frac{1}{N^s} \sum_{k=1}^{N^s} \text{BCE}(\tilde{S}_k, Y_k), \tag{7}$$

where $\text{BCE}(\cdot, \cdot)$ denotes the binary cross-entropy function and N^s is the number of support molecules. During optimization, only the GPA parameters, including prompt embeddings, projection matrices, adapters, and Transformer layer normalization, are updated.

2.4.2 Inference Stage

During inference, query molecules and the pocket are processed in the same manner as during training, including the application of GPA for adaptation. Specifically, for the r-th query molecule $\{m_r^q\}$ and the pocket p, the similarity score is computed as $\tilde{S}_r = \frac{\tilde{Z}_p^\top \tilde{Z}_{m^q}^{(r)}}{\|\tilde{Z}_p\| \cdot \|\tilde{Z}_{m^q}^{(r)}\|}$. Accordingly, the prediction for each query molecule is given by similarity scores. Finally, the query molecules are ranked according to their predicted scores $\tilde{S}_r$.

Table 5: Few-shot virtual screening results on the DUD-E benchmark under the scaffold split compared with zero-shot baseline (2-, 4-, 8-, and 16-shot settings).

Sample	Method	AUROC (%)	BEDROC (%)	EF 0.5%	EF 1%	EF 5%
2	Zero-shot	80.58±0.10	49.61±0.17	41.22±0.34	32.73±0.15	10.61±0.04
	Drug-few	**82.36±0.14**	**51.93±0.12**	**43.31±0.37**	**34.31±0.24**	**11.11±0.04**
4	Zero-shot	80.41±0.08	48.87±0.19	41.96±0.19	32.98±0.17	10.54±0.03
	Drug-few	**84.44±0.21**	**54.84±0.25**	**46.74±0.58**	**37.24±0.12**	**11.72±0.04**
8	Zero-shot	80.32±0.19	48.39±0.21	43.46±0.07	33.68±0.09	10.54±0.06
	Drug-few	**87.87±0.49**	**60.07±0.89**	**52.96±0.73**	**42.40±0.49**	**13.04±0.17**
16	Zero-shot	78.89±0.47	46.83±0.51	44.83±0.34	33.97±0.56	10.41±0.12
	Drug-few	**91.60±0.50**	**67.34±0.68**	**61.12±0.67**	**49.44±0.89**	**14.86±0.14**

Table 6: Few-shot virtual screening results on the DUD-E benchmark under the feature-cluster split compared with zero-shot baseline (2-, 4-, 8-, and 16-shot settings).

Sample	Method	AUROC (%)	BEDROC (%)	EF 0.5%	EF 1%	EF 5%
2	Zero-shot	80.77±0.01	49.94±0.01	39.34±0.01	32.21±0.03	10.68±0.01
	Drug-few	**82.16±0.08**	**51.98±0.10**	**40.72±0.14**	**33.56±0.10**	**11.06±0.02**
4	Zero-shot	80.87±0.01	49.93±0.02	39.82±0.05	32.54±0.02	10.71±0.01
	Drug-few	**84.29±0.27**	**55.18±0.38**	**43.24±0.37**	**36.31±0.36**	**11.73±0.02**
8	Zero-shot	81.01±0.03	49.91±0.06	40.71±0.05	33.09±0.07	10.76±0.01
	Drug-few	**87.27±0.16**	**58.33±1.50**	**46.00±1.60**	**39.05±1.02**	**12.52±0.17**
16	Zero-shot	81.37±0.03	49.75±0.06	43.15±0.04	34.34±0.05	10.86±0.02
	Drug-few	**92.66±0.56**	**64.66±3.83**	**51.81±3.05**	**44.51±2.81**	**14.62±0.47**

3 EXPERIMENTS

We evaluate Drug-few on three widely used target-specific virtual screening benchmarks, DUD-E (Mysinger et al., 2012), LIT-PCBA (Tran-Nguyen et al., 2020), DEKOIS 2.0 (Bauer et al., 2013), and CASF-2016 (Su et al., 2018). Each benchmark provides a collection of protein targets along with their corresponding active and inactive compounds. For more details of the benchmarks, please refer to Appendix C. For few-shot evaluation, we randomly sample a small number of actives (2, 4, 8, or 16) together with an equal number of randomly selected inactives to construct the support set for model training. Subsequently, the trained model is applied to the remaining compounds, where the goal is to prioritize actives at the top of the ranked list. To measure performance, we adopt established early recognition metrics in VS, including AUROC, BEDROC, and Enrichment Factor (EF). Finally, all experiments are conducted on an RTX 4090 GPU with an SGD optimizer, and full implementation details, including training procedures and hyperparameters, are provided in Appendix B.

It is important to emphasize that Drug-few is the first few-shot learning framework explicitly designed for target-specific VS. In contrast, existing few-shot drug discovery methods typically address molecular property prediction (e.g., classifying compounds as active or inactive) without considering the protein target, making direct comparison inappropriate. Therefore, we benchmark Drug-few against its zero-shot counterpart, i.e., the base DrugCLIP model directly applied to candidate ranking without adaptation, and the results are summarized in Section 3.1. Moreover, we compare Drug-few with the ligand-only few-shot method in Section 3.2. Furthermore, to highlight the ability of Drug-few to uncover novel actives beyond simple similarity, we perform evaluations under scaffold and feature-cluster splits (Section 3.3). For a fair comparison with classic CLIP-based few-shot learning methods, we include such baselines in the ablation study. All ablation studies are presented in Section 3.4, followed by visualization analyses in Section 3.5.

3.1 COMPARISON WITH ZERO-SHOT DRUGCLIP

In this section, we compare Drug-few with DrugCLIP (zero-shot). For each target, we randomly sample a subset of actives and the same number of inactives as the support set for training, while the remaining compounds are used for evaluation. Each experiment is repeated three times, and the

mean performance with standard deviations is reported in Table 1, Table 2, Table 3, and Table 4 for DUD-E, LIT-PCBA, DEKOIS 2.0, and CASF-2016, respectively.

Compared with zero-shot, our few-shot learning approach consistently outperforms across nearly all metrics, showing that the model can extract meaningful binding information from few examples without overfitting. As the number of support examples increases, the performance of our method steadily improves, which further validates its capacity to leverage additional supervision effectively. It is worth noting that when more actives are sampled for the support set, the proportion of actives remaining in the evaluation set becomes smaller, making the ranking task more challenging. Specifically, on the DUD-E benchmark, our method surpasses the zero-shot baseline by 12.40% in AUROC and 22.59% in BEDROC under the 16-shot setting, which demonstrates a clear advantage in learning pocket–ligand interaction rules from limited data. Similarly, on the LIT-PCBA benchmark, Drug-few also outperforms the zero-shot baseline in nearly all AUROC and BEDROC metrics, with only a small number of EF scores showing comparable values. Furthermore, under the most limited 2-shot setting, our framework achieves a 2.97% BEDROC improvement on the DEKOIS 2.0 benchmark and a 3.55% improvement on CASF-2016, highlighting its robustness even with extremely scarce supervision.

3.2 Comparison with ligand-only few-shot method

Most existing few-shot virtual screening methods are ligand-only approaches that rely solely on SMILES strings or molecular similarity and do not incorporate any form of target-specific information. Because our framework explicitly integrates 3D pocket representations as part of the target modeling, a direct comparison with ligand-only methods is inherently less fair due to the fundamentally different problem formulation.

To provide a more meaningful reference, we additionally evaluate the most advanced ligand-based few-shot model, FS-CAP (Eckmann et al., 2024), which represents the current state of the art among ligand-similarity-based methods. Using the official implementation and pretrained weights, we run three independent trials for FS-CAP on exactly the same support and query splits on DUD-E benchmark used by Drug-few, and report the averaged performance and standard deviations. This setup ensures a controlled and strictly comparable evaluation protocol. Table 7 summarizes the results under 2-, 4-, 8-, and 16-shot settings. Across all metrics and shot configurations, Drug-few consistently and substantially outperforms FS-CAP. These results demonstrate the clear advantage of incorporating explicit 3D pocket information in few-shot target-specific virtual screening.

Table 7: Comparison between FS-CAP and Drug-few under 2-, 4-, 8-, and 16-shot settings on DUD-E benchmark. Results are averaged over three runs.

Sample	Method	AUROC (%)	BEDROC (%)	EF 0.5%	EF 1%	EF 5%
2	FS-CAP	70.80±0.06	19.89±0.10	12.95±0.18	11.42±0.21	5.77±0.02
	Drug-few	**82.35±0.04**	**52.13±0.05**	**40.90±0.23**	**33.65±0.08**	**11.10±0.02**
4	FS-CAP	70.83±0.10	19.50±0.27	12.76±0.41	11.23±0.16	5.72±0.04
	Drug-few	**84.78±0.24**	**56.20±0.07**	**43.95±0.13**	**36.80±0.62**	**11.89±0.06**
8	FS-CAP	70.89±0.12	19.27±0.25	12.67±0.23	11.25±0.16	5.74±0.06
	Drug-few	**88.54±0.37**	**61.95±0.43**	**48.79±0.15**	**41.66±0.23**	**13.20±0.12**
16	FS-CAP	70.78±0.04	18.89±0.17	12.90±21.23	11.28±12.27	5.73±0.05
	Drug-few	**93.16±0.11**	**71.70±0.13**	**58.17±0.34**	**50.22±0.21**	**15.40±0.08**

3.3 Evaluation under similarity-controlled splits

To further assess whether Drug-few can discover actives beyond simple similarity rules, we conduct experiments on the DUD-E benchmark under two complementary split strategies. The first, referred to as the scaffold split, ensures that the scaffolds of support actives do not overlap with those of the query actives. This setting reflects a traditional measure of chemical similarity based on two-dimensional structural cores and prevents the model from exploiting trivial scaffold overlap. The second, termed the feature-cluster split, leverages molecular representations from the pre-trained Uni-Mol (Zhou et al., 2023) encoder. Here, all actives for each target are clustered in the learned embedding space, and support and query actives are sampled from different clusters, thereby enforcing dissimilarity in both geometric structures and chemical properties. While the scaffold split eval-

Table 8: Ablation studies on interaction-aware adaptation, adapter placement and prompt tokens inserting placement. **Bold** indicates best performance.

Sample	Method	AUROC (%)	BEDROC (%)	EF 0.5%	EF 1%	EF 5%
2	w/o interaction-aware	72.94±3.04	14.47±3.72	9.15±2.75	7.80±2.19	4.69±0.91
	Post-hoc adapter	61.44±0.61	7.66±0.14	4.78±0.24	4.13±0.06	2.62±0.07
	Shallow	81.83±0.14	51.52±0.18	40.04±0.39	33.27±0.11	10.94±0.07
	Drug-few	**82.35±0.04**	**52.13±0.05**	**40.90±0.23**	**33.65±0.08**	**11.10±0.02**
4	w/o interaction-aware	76.97±2.26	20.26±5.40	13.17±4.01	11.45±3.20	6.23±1.32
	Post-hoc adapter	72.33±0.08	16.30±0.31	10.52±0.18	9.37±0.24	5.03±0.08
	Shallow	84.15±0.11	54.97±0.11	42.82±0.09	36.04±0.08	11.75±0.08
	Drug-few	**84.78±0.24**	**56.20±0.07**	**43.95±0.13**	**36.80±0.62**	**11.89±0.06**
8	w/o interaction-aware	78.06±4.59	21.51±6.97	14.71±5.23	12.61±4.44	6.50±1.65
	Post-hoc adapter	82.99±0.23	34.73±0.75	25.17±0.63	21.59±0.47	8.99±0.17
	Shallow	87.12±0.20	59.37±0.24	47.11±0.45	39.92±0.18	12.68±0.09
	Drug-few	**88.54±0.37**	**61.95±0.43**	**48.79±0.15**	**41.66±0.23**	**13.20±0.12**
16	w/o interaction-aware	88.22±1.86	41.56±5.78	32.41±5.13	26.96±3.86	10.95±1.21
	Post-hoc adapter	89.39±0.29	51.45±0.81	41.68±0.85	34.56±0.51	12.22±0.19
	Shallow	90.82±0.46	66.11±1.11	54.71±1.01	46.33±1.04	14.18±0.15
	Drug-few	**93.16±0.11**	**71.70±0.13**	**58.17±0.34**	**50.22±0.21**	**15.40±0.08**

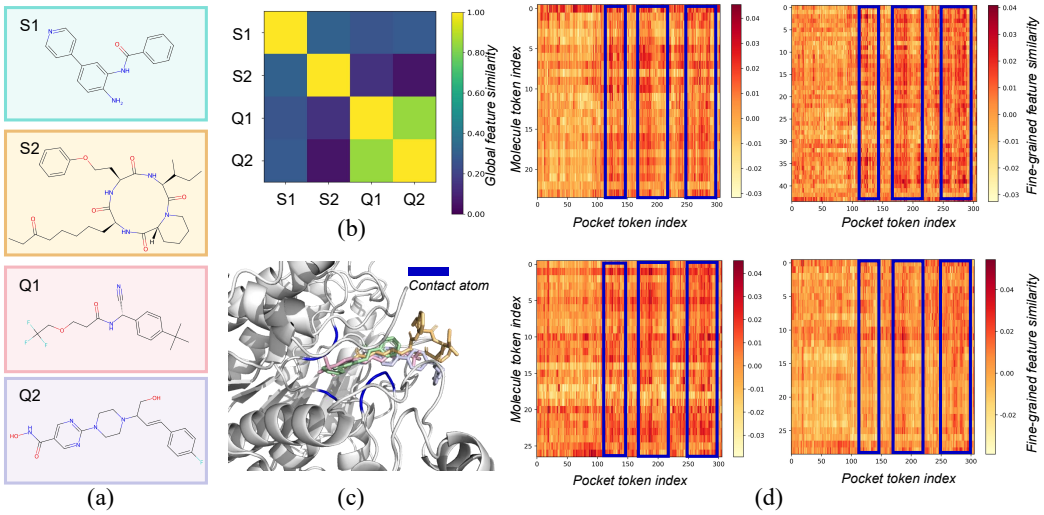


Figure 3: Visualization analysis on the target HDAC2 under the 2-shot setting with feature-cluster split sampling. (a) 2D structures of all support actives (S1, S2) and query actives (Q1, Q2). (b) Global feature similarity heatmap among S1, S2, Q1, and Q2, with darker colors indicating lower similarity. (c) Docked conformations of the four molecules with HDAC2. Molecular colors correspond to the backgrounds in (a), and blue regions on the protein indicate contact atoms with four ligands. (d) Fine-grained molecule–pocket attention maps for S1, S2, Q1, and Q2. Blue boxes highlight pocket regions with strong interactions with the molecules.

uates generalization across distinct structural backbones, the feature-cluster split captures broader molecular diversity.

Results are presented in Tables 5 and 6. Across both split strategies, Drug-few consistently outperforms the zero-shot baseline, achieving performance comparable to that under random splits. These results indicate that Drug-few does not rely on molecule similarity alone. Rather, it effectively captures target-specific interaction patterns, enabling the prioritization of novel actives even when support and query sets are deliberately made dissimilar at either the scaffold or representation level. This shows the potential practical utility of Drug-few in discovering structurally and chemically novel active molecules, which is the central challenge in target-specific VS.

3.4 Ablation study

To better understand the contribution of each component in our framework, we perform a series of ablation studies on the DUD-E benchmark. All experiments are performed on the same datasets described in Section 3.1, and additional studies are in Appendix H.

Interaction-aware adaptation: We remove the TMI module that controls the prompt tokens with pocket-ligand interaction information. In this setting, the model simply inserts randomly initialized prompt tokens after the [CLS] token in the input of each Transformer layer, similar to a direct adaptation of VPT-style prompting in CLIP few-shot learning (Jia et al., 2022). As shown in the "Interaction-aware adaptation" row of Table 8, this setting results in extremely high variance across runs, and the mean performance is even lower than the zero-shot baseline. This indicates that naively migrating CLIP-style prompting harms model stability and accuracy. In contrast, our interaction-guided adaptation provides a more reliable inductive bias, ensuring that the model learns from meaningful pocket-molecule interaction information.

Adapter placement: We evaluate the effect of placing a lightweight adapter layer only at the end of the encoder, corresponding to the "post-hoc adapter" design commonly used in CLIP few-shot transfer (Gao et al., 2024b). As shown in Table 8, this design exhibits clear overfitting under the 2-, 4-, 8-shot settings, with performance metrics even falling below the zero-shot baseline. This suggests that simply applying a post-hoc adjustment to the final features is insufficient to capture fine-grained pocket-ligand interactions, and may even introduce adverse effects.

Prompt tokens inserting placement: We compare "shallow" prompt tokens (inserted only before the input of the first layer) with "deep" prompt tokens (inserted before the input of every layer). Table 8 shows "deep" consistently outperforms "shallow", improving BEDROC from 66.11% to 71.70% in 16-shot setting, demonstrating the necessity of distributing prompt tokens across all layers.

3.5 Visualization

As shown in Figure 3, we visualize results of target HDAC2 under the 2-shot feature-cluster setting. All support actives (S1, S2) and two selected query actives (Q1, Q2) are shown in Figure 3a. As shown in Figure 3b, the global similarity between support actives and query actives are weak. In contrast, the fine-grained attention maps (Figure 3d) reveal consistent focus on the same pocket regions (blue boxes), which align with contact atoms in the docked conformations (Figure 3c), confirming chemically meaningful interactions. Despite low global similarity, Drug-few ranks Q1 and Q2 in the top 0.06% and 0.15% of candidates. These results highlight the importance of interaction-aware representations for early active recognition and demonstrate Drug-few's ability to prioritize novel molecules from few known actives. Additional case studies are provided in Appendix F.

4 Conclusion

In this study, we present Drug-few, the first framework that explicitly incorporates target structural information for target-specific few-shot virtual screening. By incorporating a small set of learnable parameters into the DrugCLIP backbone, it rapidly captures target–ligand interaction patterns and refines their representations. Drug-few consistently outperforms the zero-shot counterpart and maintains strong performance under scaffold and feature-cluster splits, demonstrating robustness to dissimilar molecules. Importantly, it enables the discovery of novel actives and provides a practical approach for identifying new actives for a given target when experimental data are extremely limited. By establishing this framework, Drug-few opens a new path for applying few-shot learning to target-specific virtual screening and lays the foundation for future methods to further advance early-stage drug discovery.

Ethics Statement

This work does not involve human subjects, sensitive personal data, or dual-use concerns. All datasets employed are publicly available and widely used in prior research, with appropriate licenses and citations. We ensure that no proprietary or confidential data are included, and our methods are intended solely for advancing scientific research in computational biology and machine learning.

To the best of our knowledge, this work raises no ethical issues related to fairness, bias, privacy, or security.

Reproducibility Statement

We provide detailed descriptions of our model architectures, hyperparameters, training protocols and inference process in the main body and the Appendix. All experiments are conducted using publicly available datasets, and we will release our code and trained models to ensure full reproducibility.

References

Dávid Bajusz, Anita Rácz, and Károly Héberger. Why is tanimoto index an appropriate choice for fingerprint-based similarity calculations? *Journal of cheminformatics*, 7(1):20, 2015.

Mahta Barekatain, Linda C Johansson, Jordy H Lam, Hao Chang, Anastasiia V Sadybekov, Gye Won Han, Joseph Russo, Joshua Bliesath, Nicola L Brice, Mark BL Carlton, et al. Structural insights into the high basal activity and inverse agonism of the orphan receptor gpr6 implicated in parkinson's disease. *Science signaling*, 17(865):eado8741, 2024.

Matthias R Bauer, Tamer M Ibrahim, Simon M Vogel, and Frank M Boeckler. Evaluation and optimization of virtual screening workflows with dekois 2.0–a public library of challenging docking benchmark sets. *Journal of chemical information and modeling*, 53(6):1447–1462, 2013.

Michael Brocidiacono, Paul Francoeur, Rishal Aggarwal, Konstantin I Popov, David Ryan Koes, and Alexander Tropsha. Bigbind: learning from nonstructural data for structure-based virtual screening. *Journal of Chemical Information and Modeling*, 64(7):2488–2495, 2023.

Heng Cai, Chao Shen, Tianye Jian, Xujun Zhang, Tong Chen, Xiaoqi Han, Zhuo Yang, Wei Dang, Chang-Yu Hsieh, Yu Kang, et al. Carsidock: a deep learning paradigm for accurate protein–ligand docking and screening based on large-scale pre-training. *Chemical Science*, 15(4):1449–1471, 2024.

Peter Eckmann, Jake Anderson, Rose Yu, and Michael K Gilson. Ligand-based compound activity prediction via few-shot learning. *Journal of Chemical Information and Modeling*, 64(14):5492–5499, 2024.

Aled M Edwards and Dafydd R Owen. Protein–ligand data at scale to support machine learning. *Nature Reviews Chemistry*, pp. 1–12, 2025.

Bowen Gao, Bo Qiang, Haichuan Tan, Yinjun Jia, Minsi Ren, Minsi Lu, Jingjing Liu, Wei-Ying Ma, and Yanyan Lan. Drugclip: Contrasive protein-molecule representation learning for virtual screening. *Advances in Neural Information Processing Systems*, 36, 2024a.

Peng Gao, Shijie Geng, Renrui Zhang, Teli Ma, Rongyao Fang, Yongfeng Zhang, Hongsheng Li, and Yu Qiao. Clip-adapter: Better vision-language models with feature adapters. *International Journal of Computer Vision*, 132(2):581–595, 2024b.

Thomas A Halgren, Robert B Murphy, Richard A Friesner, Hege S Beard, Leah L Frye, W Thomas Pollard, and Jay L Banks. Glide: a new approach for rapid, accurate docking and scoring. 2. enrichment factors in database screening. *Journal of medicinal chemistry*, 47(7):1750–1759, 2004.

Ye Hu, Gerald M Maggiora, and Jürgen Bajorath. Activity cliffs in pubchem confirmatory bioassays taking inactive compounds into account. *Journal of computer-aided molecular design*, 27(2): 115–124, 2013.

Menglin Jia, Luming Tang, Bor-Chun Chen, Claire Cardie, Serge Belongie, Bharath Hariharan, and Ser-Nam Lim. Visual prompt tuning. In *European conference on computer vision*, pp. 709–727. Springer, 2022.

Asad U Khan et al. Descriptors and their selection methods in qsar analysis: paradigm for drug design. *Drug discovery today*, 21(8):1291–1302, 2016.

Ruifeng Li, Wei Liu, Xiangxin Zhou, Mingqian Li, Qiang Zhang, Hongyang Chen, and Xuemin Lin. Contextual representation anchor network to alleviate selection bias in few-shot drug discovery. *arXiv preprint arXiv:2410.20711*, 2024.

Ruifeng Li, Mingqian Li, Wei Liu, Yuhua Zhou, Xiangxin Zhou, Yuan Yao, Qiang Zhang, and Hongyang Chen. Unimatch: Universal matching from atom to task for few-shot drug discovery. *arXiv preprint arXiv:2502.12453*, 2025.

Andrew T McNutt, Paul Francoeur, Rishal Aggarwal, Tomohide Masuda, Rocco Meli, Matthew Ragoza, Jocelyn Sunseri, and David Ryan Koes. Gnina 1.0: molecular docking with deep learning. *Journal of cheminformatics*, 13(1):43, 2021.

Michael M Mysinger, Michael Carchia, John J Irwin, and Brian K Shoichet. Directory of useful decoys, enhanced (dud-e): better ligands and decoys for better benchmarking. *Journal of medicinal chemistry*, 55(14):6582–6594, 2012.

Ankit R Patel, Hitesh B Patel, Shailesh K Mody, Ratn Deep Singh, Vaidehi N Sarvaiya, Sanjay H Vaghela, and Sheen Tukra. Virtual screening in drug discovery. *Journal of Veterinary Pharmacology and Toxicology*, 20(2):1–9, 2021.

Xiaoliang Qian, Bin Ju, Ping Shen, Keda Yang, Li Li, and Qi Liu. Meta learning with attention based fp-gnns for few-shot molecular property prediction. *ACS omega*, 9(22):23940–23948, 2024.

David Rogers and Mathew Hahn. Extended-connectivity fingerprints. *Journal of chemical information and modeling*, 50(5):742–754, 2010.

Anastasiia V Sadybekov and Vsevolod Katritch. Computational approaches streamlining drug discovery. *Nature*, 616(7958):673–685, 2023.

Johannes Schimunek, Philipp Seidl, Lukas Friedrich, Daniel Kuhn, Friedrich Rippmann, Sepp Hochreiter, and Günter Klambauer. Context-enriched molecule representations improve few-shot drug discovery. In *The Eleventh International Conference on Learning Representations*, 2023.

Gisbert Schneider. Virtual screening: an endless staircase? *Nature Reviews Drug Discovery*, 9(4): 273–276, 2010.

Thomas Scior, Andreas Bender, Gary Tresadern, José L Medina-Franco, Karina Martínez-Mayorga, Thierry Langer, Karina Cuanalo-Contreras, and Dimitris K Agrafiotis. Recognizing pitfalls in virtual screening: a critical review. *Journal of chemical information and modeling*, 52(4):867–881, 2012.

Russell Spitzer and Ajay N Jain. Surflex-dock: Docking benchmarks and real-world application. *Journal of computer-aided molecular design*, 26:687–699, 2012.

Megan Stanley, John F Bronskill, Krzysztof Maziarz, Hubert Misztela, Jessica Lanini, Marwin Segler, Nadine Schneider, and Marc Brockschmidt. Fs-mol: A few-shot learning dataset of molecules. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*, 2021.

Minyi Su, Qifan Yang, Yu Du, Guoqin Feng, Zhihai Liu, Yan Li, and Renxiao Wang. Comparative assessment of scoring functions: the casf-2016 update. *Journal of chemical information and modeling*, 59(2):895–913, 2018.

Vikram Sundar and Lucy Colwell. Attribution methods reveal flaws in fingerprint-based virtual screening. *arXiv preprint arXiv:2007.01436*, 2020.

Viet-Khoa Tran-Nguyen, Célien Jacquemard, and Didier Rognan. Lit-pcba: an unbiased data set for machine learning and virtual screening. *Journal of chemical information and modeling*, 60(9): 4263–4273, 2020.

Oleg Trott and Arthur J Olson. Autodock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, 31(2):455–461, 2010.

Xianjin Xu and Xiaoqin Zou. Dissimilar ligands bind in a similar fashion: a guide to ligand binding-mode prediction with application to celpp studies. *International Journal of Molecular Sciences*, 22(22):12320, 2021.

Keqiong Zhang, Zhiran Fan, Qilong Wu, Jianfeng Liu, and Sheng-You Huang. Improved prediction of drug–protein interactions through physics-based few-shot learning. *Journal of Chemical Information and Modeling*, 2025.

Xiangying Zhang, Haotian Gao, Haojie Wang, Zhihang Chen, Zhe Zhang, Xinchong Chen, Yan Li, Yifei Qi, and Renxiao Wang. Planet: a multi-objective graph neural network model for protein–ligand binding affinity prediction. *Journal of Chemical Information and Modeling*, 64(7):2205–2220, 2023a.

Xujun Zhang, Odin Zhang, Chao Shen, Wanglin Qu, Shicheng Chen, Hanqun Cao, Yu Kang, Zhe Wang, Ercheng Wang, Jintu Zhang, et al. Efficient and accurate large library ligand docking with karmadock. *Nature Computational Science*, 3(9):789–804, 2023b.

Liangzhen Zheng, Jingrong Fan, and Yuguang Mu. Onionnet: a multiple-layer intermolecular-contact-based convolutional neural network for protein–ligand binding affinity prediction. *ACS omega*, 4(14):15956–15965, 2019.

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. In *The Eleventh International Conference on Learning Representations*, 2023.

Kaiyang Zhou, Jingkang Yang, Chen Change Loy, and Ziwei Liu. Conditional prompt learning for vision-language models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 16816–16825, 2022a.

Kaiyang Zhou, Jingkang Yang, Chen Change Loy, and Ziwei Liu. Learning to prompt for vision-language models. *International Journal of Computer Vision*, 130(9):2337–2348, 2022b.

A Related work

Virtual screening: Virtual screening (VS) aims to identify the most relevant candidate compounds from large libraries that are likely to bind a target pocket. Traditional methods (Trott & Olson, 2010; Halgren et al., 2004; Spitzer & Jain, 2012) rely on docking software to generate and score interaction conformations, which is computationally intensive and sensitive to scoring-function accuracy. Deep learning approaches (Zheng et al., 2019; Zhang et al., 2023a; Cai et al., 2024; Zhang et al., 2023b; Brocidiacono et al., 2023) accelerate this process by predicting docking poses or binding affinities with deep models, but they still require large amounts of labeled data that are rarely available. To overcome these limitations, recent studies reformulate VS as an embedding-retrieval task. For example, DrugCLIP (Gao et al., 2024a) aligns protein pocket and ligand features through contrastive learning, enabling candidate ranking by similarity in a shared embedding space. This design preserves target-awareness while avoiding costly docking or affinity labels, and its transferability makes it naturally compatible with few-shot settings, similar to how CLIP supports data-efficient learning in NLP and CV (Zhou et al., 2022b;a; Jia et al., 2022).

Few-shot learning for drug discovery: In drug discovery, a growing line of work explores few-shot learning paradigms. Early approaches were predominantly ligand-centric, using molecular fingerprints (e.g., Morgan fingerprints) or clustering to find new molecules similar to known ligands (Rogers & Hahn, 2010; Khan et al., 2016). With the rise of deep learning, representation-based methods emerged. Some employ neural networks or fused fingerprints to encode molecules, followed by similarity comparison or prototype aggregation between support and query compounds (Eckmann et al., 2024; Schimunek et al., 2023; Stanley et al., 2021). Another direction applies meta-learning to enable cross-task transferability (Li et al., 2025; Qian et al., 2024; Li et al., 2024). Despite their advances, these studies remain ligand-only formulations: tasks are constructed at the molecule level, and predictions rely on embedding similarity or prototype matching without explicitly modeling protein-ligand interactions. These ligand-based methods limits their ability to discover novel actives and may lead to practical issues such as activity cliffs. Strictly speaking, they resemble few-shot molecular property prediction or classification rather than strict VS, where the task is to identify active ligands for a specific target. Addressing this gap requires frameworks that explicitly incorporate target information into the few-shot setting, which motivates our work.

B Details of training and inference stages

All training and inference experiments are conducted on the trained DrugCLIP (Gao et al., 2024a) backbone. For each target, the model weights are reloaded independently to ensure that few-shot episodes do not interfere with each other. This design is consistent with the goal of building a target-specific framework, rather than training a global classifier to distinguish actives from inactives across all targets. As a result, each training and inference process focuses on a single target.

Two types of support and query sets sampling strategies are used. In random sampling (Section 3.1), we select N pairs of actives and inactives as the N-shot support set, and the remaining compounds serve as the query set. The model is trained on the support set and evaluated on the query set, where the objective is to rank actives ahead of inactives, reflecting realistic virtual screening scenarios. In similarity-controlled splits (Section 3.3), we adopt scaffold and feature-cluster splits. Scaffold split ensures that actives in the support and query sets do not share scaffolds. Feature-cluster split ensures separation according to molecular feature clusters derived from the Uni-Mol (Zhou et al., 2023) encoder.

All experiments are run on a single NVIDIA RTX 4090 GPU. To make the approach applicable when the number of known samples per target is uncertain, a unified training configuration is used across all few-shot settings. The model is trained for 10 epochs with a learning rate of 0.01 using the SGD optimizer. The projection matrix for prompt tokens is initialized with Kaiming initialization. The hyperparameters are summarized in Table 9.

Table 9: Hyperparameter settings

Hyperparameter	**Value**
Number of molecular prompt tokens	5
Number of pocket prompt tokens	3
Dimension of projection matrix (W_{proj})	512
W_{proj} hidden dimension	64
Number of transformer layers	15

C Details of benchmark datasets

DUD-E (Mysinger et al., 2012) includes 102 protein targets with an average of 224 annotated actives per target, each paired with property-matched decoys that differ in topology. LIT-PCBA (Tran-Nguyen et al., 2020) covers 15 targets with 7,844 active compounds and over 407,381 unique inactives derived from PubChem assays, providing a more challenging and realistic evaluation. DEKOIS 2.0 (Bauer et al., 2013) consists of 81 protein targets across diverse families, where each target includes 40 curated actives and 1,200 drug-like decoys designed to rigorously test VS performance.

D Custom CLIP few-shot strategies

To better understand the design choices in few-shot adaptation for CLIP-based models, we categorize existing methods into two main strategies: prompt insertion before the encoder and adapter-based tuning after the encoder. The following sections briefly describe each strategy and relate them to Drug-few.

D.1 Prompt insertion before the encoder

This mode inserts learnable prompt tokens or vectors directly into the model input before the encoder. Inspired by textual prompt-based fine-tuning in NLP (e.g., CoOp (Zhou et al., 2022b), CoCoOp (Zhou et al., 2022a)), the prompt vectors are optimized using downstream task data while keeping the main encoder largely frozen. In multi-modal settings, visual prompt tuning (VPT) (Jia et al., 2022) similarly introduces learnable tokens or pixel-level prompt tokens before the visual encoder. The key idea is to guide the model to adapt to new tasks by modifying the input representation without changing the backbone features. In our experiments, the "Interaction-aware adaptation" ablation corresponds to this mode: when we remove the interaction module and insert random prompt tokens in each Transformer layer (VPT-style), the model shows high variance and even lower mean performance than zero-shot, highlighting that naive prompt insertion without task-relevant guidance is unstable.

D.2 Prompt insertion after the encoder

In this mode, a small adapter module is added after the encoder, and only this module is fine-tuned on the downstream task (Gao et al., 2024b). The encoder outputs are projected through the adapter, which captures task-specific adjustments while preserving the pre-trained knowledge of the backbone. This approach is efficient and reduces the risk of overfitting, as the main encoder weights remain fixed and only a small set of parameters is updated. In our "Adapter placement" ablation, placing a lightweight adapter only at the end of the encoder corresponds to this mode. The results show clear overfitting in few-shot settings, with performance sometimes falling below zero-shot, indicating that post-hoc adapters alone cannot capture fine-grained pocket-ligand interactions effectively.

In contrast, Drug-few combines interaction-aware prompt insertion with adapter-based adjustments in a coordinated manner. By conditioning prompt tokens on pocket-ligand interactions and leveraging adapters effectively, the method consistently improves performance across almost all targets, achieving both stability and fine-grained adaptation that neither naive pre-encoder prompting nor post-hoc adapters alone can provide.

E PSEUDO CODE

Pseudo codes for the training vs. the inference stage are listed in Algorithm 1.

Algorithm 1 Training vs. Inference Stage

Training Stage

1: Reload DrugCLIP weights
2: **for** each batch of $\{m_k^s\}$ **do** and target p
3: Encoding $\{m_k^s\}$ and p
4: Compute TMI features: $\tilde{I}^{mol}, \tilde{I}^{poc}$
5: Apply GPA:

$$\tilde{Z}_m = \text{GPA}(\phi_p(m^s))$$
$$\tilde{Z}_p = \text{GPA}(\phi_p(p))$$

6: Compute similarity score:

$$\tilde{S}_k = \frac{\tilde{Z}_p^\top \tilde{Z}_{m^s}^{(k)}}{\|\tilde{Z}_p\| \cdot \|\tilde{Z}_{m^s}^{(k)}\|}$$

7: Compute loss:

$$\mathcal{L}_{\text{cls}} = \frac{1}{N^s} \sum_{k=1}^{N^s} \text{BCE}(\tilde{S}_k, Y_k)$$

8: Update the adapted parameters
9: **end for**

Inference Stage

1: Load weights from training stage for p
2: **for** each batch of $\{m_r^q\}$ **do** and target p
3: Encoding $\{m_r^q\}$ and p
4: Compute TMI features:$\tilde{I}^{mol}, \tilde{I}^{poc}$
5: Apply GPA:

$$\tilde{Z}_m = \text{GPA}(\phi_p(m^q))$$
$$\tilde{Z}_p = \text{GPA}(\phi_p(p))$$

6: Compute similarity score:

$$\tilde{S}_r = \frac{\tilde{Z}_p^\top \tilde{Z}_{m^q}^{(r)}}{\|\tilde{Z}_p\| \cdot \|\tilde{Z}_{m^q}^{(r)}\|}$$

7: **end for**
8: Rank molecules by $\tilde{S}$
9:

F CASE STUDY

To better understand how the Drug-few adapts to new query molecules, we further analyze the representations of molecules and pockets on the target CASP3. Since our framework introduces prompt tokens into both the Molecule encoder and the Pocket encoder, and these prompt tokens are modulated by interaction-aware signals, the pocket representation dynamically shifts depending on the paired molecule. This mechanism enables the model to capture fine-grained interaction patterns, rather than treating the pocket as a fixed feature.

Figure 4 provides a comparative visualization between the zero-shot baseline and Drug-few. Each pocket-molecule pair is represented by a triangle (molecule) and a star (pocket) of the same color. The dashed gray lines connect the pairs identified by the zero-shot DrugCLIP, while the red lines correspond to pairs produced by Drug-few. It can be observed that Drug-few consistently pulls active molecules closer to their corresponding pocket representation, effectively improving the alignment of pairs.

On the right side of Figure 4, we visualize four representative active molecules. The percentages indicate their predicted rank positions among all molecules for the CASP3 target. Compared to the zero-shot model, Drug-few significantly improves the rankings of these actives, moving them from the tail (e.g., 31.5%) toward the top (e.g., 1.48%). This highlights the model's ability to discover structurally diverse actives that would otherwise be overlooked.

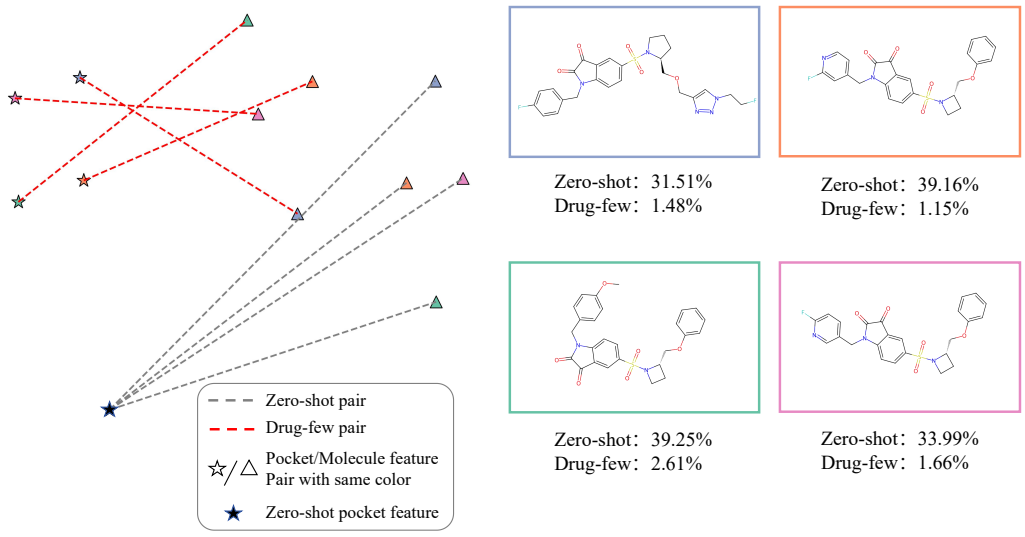
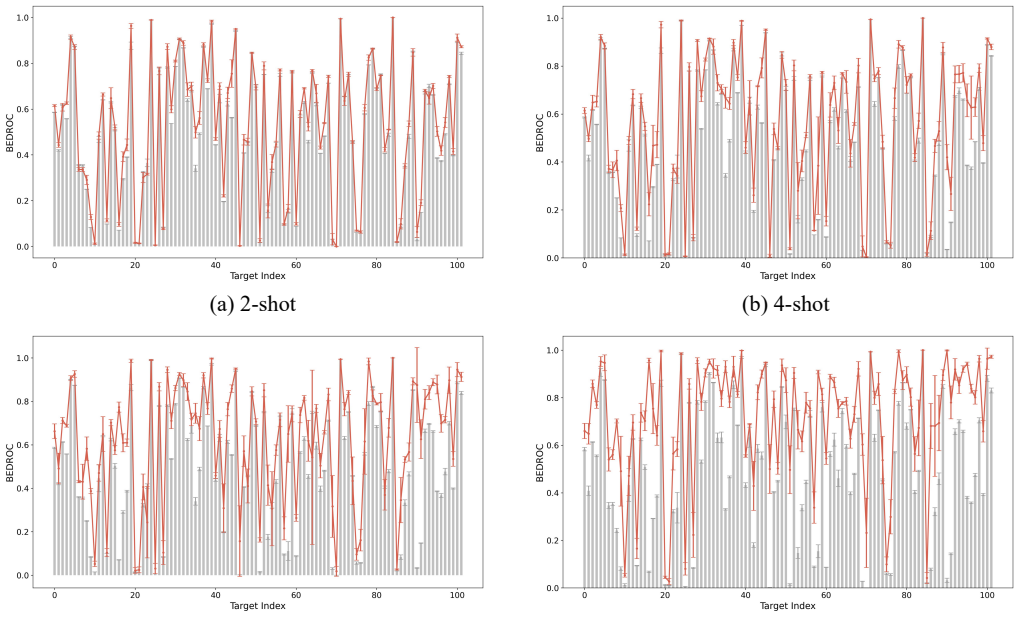


Figure 4: Case study on the CASP3 target. Left: stars denote pocket features and triangles denote molecular features; pairs are shown in the same color. Gray dashed lines indicate zero-shot pairings, while red dashed lines show Drug-few pairings. Right: visualization of four active molecules, with percentages representing their predicted ranking among all compounds.

G PER-TARGET PERFORMANCE ANALYSIS

As shown in Figure 5, we perform a per-target analysis comparing our method with the zero-shot approach described in Section 4.1. We report the performance of 2-, 4-, 8-, and 16-shot settings on each target on the DUD-E benchmark. In the plots, the red line represents our method, while the gray bars correspond to the zero-shot results. Drug-few consistently outperforms the zero-shot baseline across most targets. Moreover, the performance gap increases as the number of shots grows, indicating that the improvement is not driven by a few well-performing samples but occurs broadly across nearly all targets.



(a) 2-shot (b) 4-shot

(c) 8-shot (d) 16-shot

Figure 5: Results of 2-, 4-, 8-, and 16-shot experiments on individual targets are shown in panels (a-d). The red line represents Drug-few, and the gray bars indicate zero-shot results. Error bars represent the mean and standard deviation over three independent samplings.

H Ablation Study

H.1 Prompt token numbers

In Appendix B, we set the number of molecule tokens (M) to 5 and pocket tokens (P) to 3. We further investigate the effect of different token combinations (M/P = 1/3/5/7) on model performance. Table 10 reports the results of Drug-few on the DUD-E benchmark using random splits, averaged over three independent runs. The results indicate that performance is relatively consistent across all six M/P token combinations. This suggests that Drug-few is robust to the choice of token counts, and the selected configuration (M=5, P=3) provides the best overall performance. Across different few-shot sizes, we observe the expected trend: as the size of the support set increases (2, 4, 8, 16), performance improves consistently for all token combinations.

H.2 Layer norm tuning

We further assess the impact of freezing vs. unfreezing the layer normalization parameters in the Transformer. When all normalization layers remain frozen, model performance degrades compared to the setting where these parameters are trainable. As shown in Table 11, updating the layer norms yields a clear improvement. This gain can be attributed to the relatively small scale of the Uni-Mol encoder (Zhou et al., 2023), where modest adjustments to normalization parameters enhance adaptation without destabilizing training.

H.3 Interaction mechanism

To evaluate the contribution of our proposed interaction mechanism, we conduct an ablation study by replacing it with a lightweight cross-attention module. As shown in Table 12, Drug-few consistently outperforms the cross-attention variant across the 2-, 4-, 8-, and 16-shot settings, especially in the BEDROC and EF metrics. For example, under the 16-shot setting, BEDROC improves from 56.75% to 71.70% and EF1% increases from 37.28 to 50.22. These metrics reflect the model's ability to prioritize active compounds, which is critical for practical virtual screening.

Furthermore, the cross-attention module exhibits higher variance, indicating reduced stability. In contrast, our interaction mechanism provides stronger and more consistent performance without

Table 10: Ablation results of different molecule (M) and pocket (P) token combinations (M/P = 1, 3, 5, 7) on the DUD-E benchmark across 2-, 4-, 8-,and 16-shot. **Bold** and underline denote the best and second results across all settings.

M/P token	Sample	AUROC (%)	BEDROC (%)	EF 0.5%	EF 1%	EF 5%
51	2	82.32±0.08	52.02±0.08	40.80±0.14	33.65±0.05	11.10±0.02
	4	84.85±0.12	55.98±0.19	**43.97±0.33**	36.83±0.25	11.89±0.04
	8	88.55±0.36	61.67±0.59	48.73±0.44	41.43±0.43	13.20±0.09
	16	92.81±0.58	70.55±0.97	57.20±0.76	49.25±0.76	15.25±0.17
55	2	**82.40±0.05**	52.12±0.05	40.87±0.08	33.62±0.11	11.10±0.02
	4	84.78±0.23	55.92±0.17	43.79±0.19	36.77±0.12	11.89±0.04
	8	88.28±0.38	61.62±0.63	48.65±0.48	41.43±0.48	13.14±0.13
	16	93.03±0.39	71.52±0.04	**58.18±0.01**	50.31±0.28	15.32±0.12
57	2	82.35±0.04	**52.19±0.11**	**41.00±0.22**	**33.76±0.19**	11.11±0.02
	4	**84.88±0.27**	55.80±0.03	43.93±0.12	36.71±0.05	11.87±0.02
	8	88.51±0.60	62.15±1.19	48.74±1.27	**41.81±0.76**	13.24±0.21
	16	93.14±0.31	69.50±1.29	56.84±0.82	48.39±1.07	15.18±0.13
13	2	82.35±0.01	52.14±0.05	40.92±0.05	33.70±0.08	11.10±0.01
	4	84.56±0.07	56.00±0.12	44.09±0.14	36.93±0.20	11.89±0.02
	8	88.36±0.69	61.83±0.95	48.93±0.92	41.64±0.63	13.20±0.20
	16	92.91±0.03	71.12±0.23	57.63±0.51	49.86±0.48	15.22±0.05
33	2	82.35±0.04	52.17±0.04	40.93±0.22	33.71±0.06	**11.12±0.02**
	4	84.71±0.13	55.91±0.12	43.92±0.10	**36.95±0.18**	11.84±0.03
	8	**88.77±0.45**	**62.40±1.08**	**49.23±1.40**	41.56±0.49	**13.29±0.23**
	16	93.10±0.43	71.65±0.51	57.93±0.72	**50.32±0.46**	15.34±0.11
73	2	82.37±0.09	52.04±0.10	40.78±0.23	33.64±0.15	11.09±0.01
	4	84.75±0.25	55.89±0.27	43.74±0.44	36.84±0.18	11.84±0.06
	8	88.45±0.29	61.68±0.53	48.55±0.49	41.45±0.53	13.15±0.10
	16	93.10±0.26	71.05±0.14	57.15±0.43	49.73±0.13	15.36±0.09
Drug-few(53)	2	82.35±0.04	52.13±0.05	40.90±0.23	33.65±0.08	11.10±0.02
	4	84.78±0.24	**56.20±0.07**	43.95±0.13	36.80±0.62	**11.89±0.06**
	8	88.54±0.37	61.95±0.43	48.79±0.15	41.66±0.23	13.20±0.12
	16	**93.16±0.11**	**71.70±0.13**	58.17±0.34	50.22±0.21	**15.40±0.08**

Table 11: Ablation study on layer norm tuning.

Sample	Method	AUROC (%)	BEDROC (%)	EF 0.5%	EF 1%	EF 5%
2	w/o norm	81.86±0.02	51.33±0.05	40.32±0.07	33.07±0.09	10.97±0.02
	Drug-few	**82.35±0.04**	**52.13±0.05**	**40.90±0.23**	**33.65±0.08**	**11.10±0.02**
4	w/o norm	83.65±0.11	54.02±0.25	42.70±0.11	35.25±0.15	11.56±0.05
	Drug-few	**84.78±0.24**	**56.20±0.07**	**43.95±0.13**	**36.80±0.62**	**11.89±0.06**
8	w/o norm	87.47±0.07	60.30±0.58	47.84±0.31	40.64±0.58	12.85±0.12
	Drug-few	**88.54±0.37**	**61.95±0.43**	**48.79±0.15**	**41.66±0.23**	**13.20±0.12**
16	w/o norm	92.27±0.21	68.46±0.11	55.67±0.14	47.78±0.13	14.88±0.12
	Drug-few	**93.16±0.11**	**71.70±0.13**	**58.17±0.34**	**50.22±0.21**	**15.40±0.08**

adding any parameters or increasing model complexity, thus avoiding the instability associated with standard cross-attention. This ablation study validates the effectiveness and robustness of our interaction design.

H.4 Ligand-only prototype ranking

To investigate whether the performance improvement of Drug-few is solely due to better ligand representation, we conduct an ablation study using a ligand-similarity based prototype method. In this experiment, we first train the original Drug-few model using the standard training procedure. During evaluation, we freeze all model parameters and completely remove the pocket branch, so that the model no longer uses any pocket information or interaction module. The features of active support ligands are encoded using the frozen molecule encoder, averaged to form a prototype, and query molecules are ranked solely by their similarity to this prototype.

Table 12: Ablation study comparing the cross-attention variant with our proposed interaction mechanism on the DUD-E benchmark.

Sample	Method	AUROC (%)	BEDROC (%)	EF 0.5%	EF 1%	EF 5%
2	cross	**83.60±0.18**	**54.44±0.48**	**42.63±0.18**	**35.22±0.46**	**11.58±0.07**
	Drug-few	82.35±0.04	52.13±0.05	40.90±0.23	33.65±0.08	11.10±0.02
4	cross	**89.22±0.18**	**56.90±4.62**	42.63±3.68	36.64±3.70	**12.88±0.47**
	Drug-few	84.78±0.24	56.20±0.07	**43.95±0.13**	**36.80±0.62**	11.89±0.06
8	cross	**90.99±1.59**	56.12±6.87	42.19±6.06	36.36±5.12	**13.27±1.03**
	Drug-few	88.54±0.37	**61.95±0.43**	**48.79±0.15**	**41.66±0.23**	13.20±0.12
16	cross	**93.38±0.28**	56.75±1.30	42.94±1.32	37.28±1.00	14.38±0.14
	Drug-few	93.16±0.11	**71.70±0.13**	**58.17±0.34**	**50.22±0.21**	**15.40±0.08**

Table13 summarizes the results on the DUD-E benchmark. The ligand-similarity prototype method consistently underperforms Drug-few across all metrics, particularly in early enrichment and BEDROC. This confirms that Drug-few's performance gain arises from effectively capturing pocket-ligand interactions rather than merely improving ligand representations.

Table 13: Comparison of Drug-few with a ligand-similarity prototype method on the DUD-E benchmark.

Sample	Method	AUROC (%)	BEDROC (%)	EF 0.5%	EF 1%	EF 5%
2	prototype	80.23±1.14	45.94±3.33	35.57±2.66	29.25±2.32	10.05±0.62
	Drug-few	**82.35±0.04**	**52.13±0.05**	**40.90±0.23**	**33.65±0.08**	**11.10±0.02**
4	prototype	83.08±0.52	50.95±0.53	39.43±0.20	33.18±0.41	11.07±0.09
	Drug-few	**84.78±0.24**	**56.20±0.07**	**43.95±0.13**	**36.80±0.62**	**11.89±0.06**
8	prototype	84.81±0.36	51.51±0.31	40.42±0.76	34.36±0.41	11.38±0.02
	Drug-few	**88.54±0.37**	**61.95±0.43**	**48.79±0.15**	**41.66±0.23**	**13.20±0.12**
16	prototype	87.06±0.14	55.09±0.19	45.19±0.18	38.21±0.25	12.33±0.06
	Drug-few	**93.16±0.11**	**71.70±0.13**	**58.17±0.34**	**50.22±0.21**	**15.40±0.08**

I POTENTIAL DATA LEAKAGE

To evaluate whether potential data leakage affects the reliability of our conclusions, we perform an additional study using a version of DrugCLIP retrained after removing any proteins that overlap with the test set. Table 14 summarizes the results on the DUD-E benchmark. Even after filtering, our method consistently outperforms the zero-shot baseline of this retrained DrugCLIP across all sample sizes, indicating that its effectiveness remains robust and is not attributable to unintended data leakage.

Table 14: Performance of Drug-few using a version of DrugCLIP retrained after removing any proteins that overlap with the test set.

Sample	Method	AUROC (%)	BEDROC (%)	EF 0.5%	EF 1%	EF 5%
2	zero-shot	74.82±0.03	32.59±0.05	25.43±0.06	20.57±0.06	7.61±0.01
	Drug-few	**78.53±0.22**	**38.07±0.34**	**28.92±0.19**	**24.25±0.23**	**8.76±0.07**
4	zero-shot	74.84±0.02	32.50±0.02	25.77±0.04	20.70±0.02	7.62±0.01
	Drug-few	**82.53±0.06**	**44.12±0.39**	**34.27±0.49**	**28.41±0.34**	**9.94±0.10**
8	zero-shot	74.82±0.05	32.29±0.03	26.25±0.18	21.01±0.01	7.63±0.02
	Drug-few	**88.82±0.40**	**55.91±0.56**	**43.02±0.30**	**37.02±0.26**	**12.54±0.12**
16	zero-shot	74.91±0.13	32.05±0.12	27.69±0.12	21.66±0.08	7.65±0.03
	Drug-few	**93.60±0.26**	**65.47±1.26**	**52.44±1.52**	**45.05±1.09**	**14.95±0.15**

J SENSITIVITY ANALYSIS OF TMI MODULE TO BATCH SIZE

To evaluate the sensitivity of the TMI module to inference batch size and batch composition, we conduct experiments on the DUD-E benchmark with batch sizes of 8, 16, 32, and 64. As shown in Table 15, the performance remains largely consistent across all batch sizes, with negligible differences compared to our default batch size of 32. These results indicate that the TMI module is robust to variations in batch size.

Table 15: Performance of Drug-few under different inference batch sizes on the DUD-E benchmark.

Sample	Method (Batch Size)	AUROC (%)	BEDROC (%)	EF 0.5%	EF 1%	EF 5%
2	8	82.34±0.04	52.13±0.05	40.89±0.24	33.65±0.09	11.10±0.01
	16	82.35±0.04	52.13±0.05	40.89±0.23	33.66±0.08	11.10±0.01
	64	82.35±0.04	52.13±0.05	40.90±0.23	33.65±0.08	11.10±0.01
	Drug-few (32)	82.35±0.04	52.13±0.05	40.90±0.23	33.65±0.08	11.10±0.02
4	8	84.78±0.24	56.20±0.07	43.97±0.14	36.99±0.29	11.89±0.07
	16	84.78±0.24	56.20±0.07	43.95±0.14	37.00±0.29	11.89±0.06
	64	84.78±0.24	56.20±0.07	43.96±0.12	37.01±0.28	11.89±0.06
	Drug-few (32)	84.78±0.24	56.20±0.07	43.95±0.13	36.80±0.62	11.89±0.06
8	8	88.53±0.37	61.95±0.43	48.78±0.16	41.66±0.22	13.19±0.12
	16	88.54±0.37	61.95±0.43	48.78±0.17	41.66±0.23	13.20±0.12
	64	88.54±0.37	61.96±0.43	48.79±0.15	41.66±0.24	13.20±0.12
	Drug-few (32)	88.54±0.37	61.95±0.43	48.79±0.15	41.66±0.23	13.20±0.12
16	8	93.16±0.10	71.69±0.13	58.20±0.34	50.22±0.20	15.39±0.08
	16	93.16±0.11	71.70±0.13	58.20±0.33	50.25±0.24	15.39±0.08
	64	93.16±0.11	71.71±0.12	58.18±0.35	50.22±0.20	15.39±0.09
	Drug-few (32)	93.16±0.11	71.70±0.13	58.17±0.34	50.22±0.21	15.40±0.08

K LLM USAGE STATEMENT

For this study, LLMs are used solely to polish the manuscript. They do not contribute to conceptual development, experimental procedures, data analysis, or interpretation of results. All research-related ideas and technical content are entirely produced and verified by the authors.