



Co-supervised Pre-training of Pocket and Ligand

Zhangyang Gao, Cheng Tan, Jun Xia, and Stan Z. Li[✉]

Zhejiang University AI Lab, Research Center for Industries of the Future,
Westlake University, Hangzhou, China
{gaozhangyang,tancheng,xiajun,stan.zq.li}@westlake.edu.cn

Abstract. Can we inject the pocket-ligand complementarity knowledge into the pre-trained model and jointly learn their chemical space? Pre-training molecules and proteins have attracted considerable attention in recent years, while most of these approaches focus on learning one of the chemical spaces and lack the consideration of their complementarity. We propose a co-supervised pre-training (CoSP) framework to learn 3D pocket and ligand representations simultaneously. We use a gated geometric message passing layer to model 3D pockets and ligands, where each node's chemical features, geometric position, and direction are considered. To learn meaningful biological embeddings, we inject the pocket-ligand complementarity into the pre-training model via Chem-InfoNCE loss, cooperating with a chemical similarity-enhanced negative sampling strategy to improve the representation learning. Through extensive experiments, we conclude that CoSP can achieve competitive results in pocket matching, molecule property prediction, and virtual screening.

Keywords: AI for Science · Bioinformatics · Molecular Representation Learning · Graph Neural Networks

1 Introduction

Is there a pre-trained model that explores the chemical space of pockets and ligands while considering their complementarity? Recently, many deep learning methods have been proposed to understand the chemical space of protein pockets or drug molecules (or called ligands) and facilitate drug design in many aspects, e.g., finding hits for a novel target [59], repurposing ancient drugs for new targets [25, 57, 67], and searching for similar pockets and molecules [35, 46]. While these models have shown promising potential in learning separate pocket space or molecular space for specific tasks [17, 21, 31, 47, 71], jointly pre-training pockets and ligands considering their complementarity remains to be explored.

We propose **co-supervised pretraining (CoSP)** framework for understanding the joint chemical space of pockets and ligands. Taking the ligand as an example,

Z. Gao and C. Tan—Equal Contribution.

contrastive self-supervised pre-training [17, 49, 56] has yielded significant achievements in recent years. By identifying well-defined positive and negative ligand pairs via contrastive loss, the model can learn the underlying knowledge to facilitate downstream tasks. However, these self-supervised methods only capture data dependencies in the "self" domain while ignoring additional information from other complementary fields, such as bindable pockets. Meanwhile, previous studies [1, 5, 11, 37] have shown that pocket-ligand complementarity play a crucial role in determining molecular properties, since chemically similar ligands tend to bind to similar pockets. Inspired by this, we introduce cross-domain dependencies between pockets and ligands to improve molecular representation learning.

We propose gated geometric message passing (GGMP) layer to extract expressive bio-representations for 3D pockets and ligands. All bio-objects are treated as 3D graphs [20, 24] in that each node contains invariant chemical features (atomic number, etc.) and equivalent geometric features (position and direction). For each bio-object, we optimize the pairwise energy function [22], which considers both chemical features and geometric features via the gated operation. By minimizing the energy function, we derive the updating rules of position and direction vectors. Finally, we combine these rules with classical message passing, resulting in GGMP.

We introduce ChemInfoNCE loss to reduce the negative sampling bias [9, 39]. When applying contrastive learning, the false negative pairs that are actually positive will lead to performance degradation, called negative sampling bias. Chuang [9] assumes that the label distribution of the classification task is uniform and propose DebiasedInfoNCE to alleviate this problem. Considering the specificity of the molecules and extending the situation to continuous properties prediction (regression task), we introduce chemical similarity-enhanced negative ligand sampling. Interestingly, improving the sampling strategy is equivalent to modifying sample weights; thus, we provide a systematic understanding from the view of loss functions and propose ChemInfoNCE.

We evaluate our model on several downstream tasks, from pocket matching, molecule property prediction to virtual screening. Numerous experiments show that our approach can achieve competitive results on these tasks, suggesting that the pocket-ligand complementarity could improve biorepresentation learning.

2 Related Work

Motivation. Protein and molecule achieve their biological functions by binding to each other [7], thus exploring the protein-ligand complex help to improve the understanding of both proteins, molecules, and their interactions. To improve generalization and reduce complexity, we further consider local patterns about the protein pocket x and the bindable ligand \hat{x} . Taking (x, \hat{x}) as the positive pair, while (x, \hat{x}^-) as the negative pair, where \hat{x}^- cannot bind to x , we aim to pre-train a pocket model $f : x \mapsto \mathbf{h}$ and a ligand model $\hat{f} : \hat{x} \mapsto \hat{\mathbf{h}}$, such that the mutual information between \mathbf{h} and $\hat{\mathbf{h}}$ are maximized.

Table 1. Protein and molecule pre-training methods

Protein				Molecule			
Method	Data	Code	Year	Method	Data	Code	Year
CPCProt [33]	sequence	PyTorch	2020	FragNet [44]	SMILES	–	2021
Profile Prediction [48]	sequence	–	2020	MoCL [49]	graph	PyTorch	2021
ONTOPROTEIN [70]	sequence	PyTorch	2022	MPG [29]	graph	PyTorch	2021
CARP	sequence	–	2022	Grover [40]	graph	PyTorch	2020
GearNet	3D	–	2022	MICRO-Graph [71]	graph	–	2020
				CKGNN [17]	graph	–	2021
				MGSSL [73]	graph	PyTorch	2021
				MolCLR [56]	graph	PyTorch	2022
				3DInfomax [47]	graph+3D	PyTorch	2021
				GraphMVP [31]	graph+3D	–	2022
				GEM [16]	graph+3D	Paddle	2022

Equivalent 3D GNN. Extensive works have shown that 3D structural conformation can improve the quality of bio-representations with the help of equivalent message passing layer [4, 6, 10, 19, 41, 50]. Inspired by the energy analysis [20, 22], we propose a new **g**ated **g**eometric **m**essage **p**assing (GGMP) layer that consider not only the node position but also its direction, where the latter could indicate the location of pocket cavities and the angle of molecular bonds.

InfoNCE. The original InfoNCE is proposed by [36] to contrast semantically similar (positive) and dissimilar (negative) pairs of data points, such that the representations of similar pairs (x, \hat{x}) to be close, and those of dissimilar pairs (x, \hat{x}^-) to be more orthogonal. By default, the negative pairs are uniformly sampled from the data distribution. Therefore, false negative pairs will lead to significant performance drop. To address this issue, DebaisedInfoNCE [9] is proposed, which assumes that the label distribution of the classification task is uniform. Although DebaisedInfoNCE has achieved good results on image classification, it is not suitable for direct transfer to regression tasks, as the uniform distribution assumption is too strict. For bio-objects, we discard the above assumption, extend the situation to continuous attribute prediction, use fingerprint similarity to measure the probability of negative ligands, and propose ChemInfoNCE.

Self Bio Pre-training. Many pre-training methods have been proposed for a single protein or ligand domain, which can be classified as sequence-based, graph-based or structure-based. We summarize the **protein pre-training** models in Table.1. As for sequential models, CPCPort [33] maximizes the mutual information between predicted residues and context. Profile Prediction [48] suggests predicting MSA profile as a new pre-training task. OntoProtein [70] integrates GO (Gene Ontology) knowledge graphs into protein pre-training. While most of the sequence models rely on the transformer architecture, CARP [66] finds that CNNs can achieve competitive results with much fewer parameters and runtime costs. Recently, GearNet [74] explores the potential of 3D structural

pre-training from the perspective of masked prediction and contrastive learning. We also summarize the **molecule pre-training** models in Table.1. As for sequential models, FragNet [44] combines masked language model and multi-view contrastive learning to maximize the inner mutual information of the same SMILES and the agreement across augmented SMILES. Beyond SMILES, more approaches [17, 29, 40, 49, 56, 71, 73] tend to choose graph representation that can better model structural information. For example, Grover [40] integrates message passing and transformer architectures and pre-trains a super-large GNN using 10 million molecules. MICRO-Graph [71] and MGSSL [73] use motifs for contrastive learning. Considering the domain knowledge, MoCL [49] uses substructure substitution as a new data augmentation operation and predicts pairwise fingerprint similarities. Although these pre-training methods show promising results, they do not consider the 3D molecular conformations. To fill this gap, GraphMVP [31] and 3DInfomax [47] explore to maximize the mutual information between 3D and 2D views of the same molecule and achieve further performance improvements. Besides, GEM [16] proposes a geometry-enhanced graph neural network and pre-trains it via geometric tasks. For the pre-training of individual proteins or molecules, these methods demonstrate promising potential on various downstream tasks but ignore their complementarity.

Cross Bio Pre-training. In parallel with our study, Uni-Mol [76], probably the first pre-trained model that can handle both protein pockets and molecules, released the preprinted version. However, they pre-train the pockets and ligands separately without considering their interactions, whereas our approach differs in pre-training data, pre-training strategy, model structure and downstream tasks.

3 Methodology

3.1 Co-Supervised Pre-training Framework

We propose the **co-supervised pretraining** (CoSP) framework, as shown in Figure.1, to explore the joint chemical space of protein pockets and ligands, where the methodological innovations include:

1. We propose the gated geometric message passing layer to model 3D pockets and ligands.
2. We establish a co-supervised pre-training framework to learn pocket and ligand representations.
3. We introduce ChemInfoNCE with improved negative sampling guided by chemical similarity.
4. We evaluate the model on pocket matching, molecule property prediction, and virtual-screening tasks.

3.2 Geometric Representation

We introduce the unified data representation and neural network for modeling 3D pockets and ligands. We use structures collected from the BioLip dataset [64] as pretraining data for developing CoSP_{base} model. Further, we use augmented pretraining data with CrossDock dataset [18], resulting in the CoSP_{large} model. In downstream tasks where ligand conformations are not provided, we generate 3D conformations using MMFF [52] (if successful) or their 2D conformations (if failed).

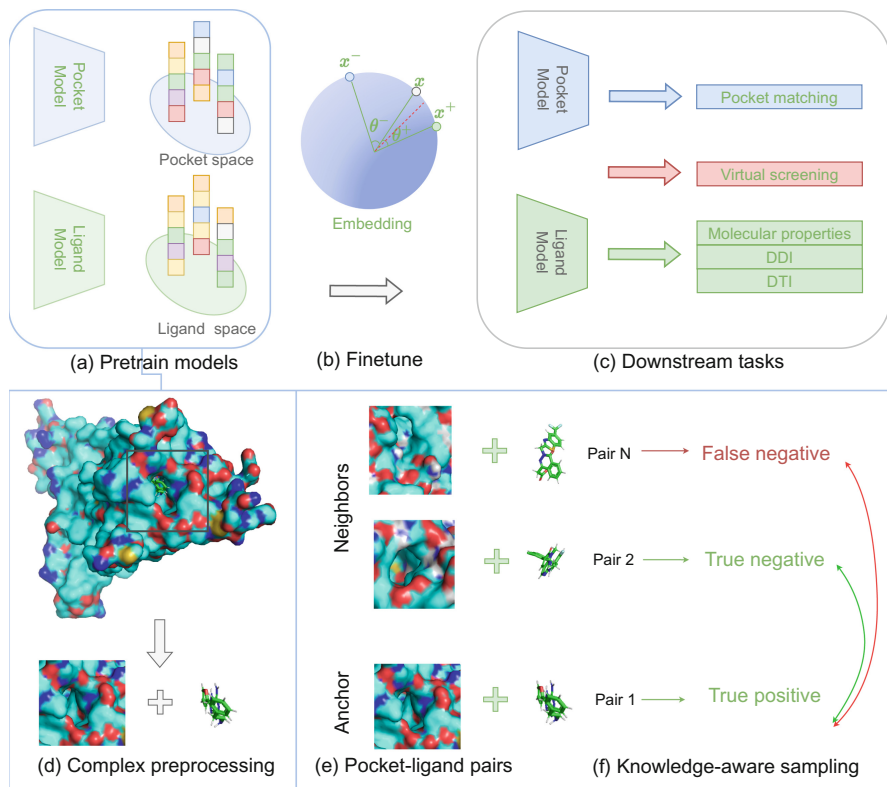


Fig. 1. Overview of CoSP. We contrast bound pocket-ligand pairs with unbound ones to learn the complementarity-aware chemical embeddings. We extract positive pocket-ligand pairs (e) from the protein-ligand complexes (d), and augment pos/neg relations of complexes via ligand similarity (f). We pretrain the model on BioLip dataset (a), followed by finetuning (b) and evaluation (c) on different tasks.

Pocket and Ligand Graph. We represent bio-object as graph $\mathcal{G}(X, \mathcal{V}, \mathcal{E})$, consisting of coordinate matrix $X \in \mathbb{R}^{n,3}$, node features $\mathcal{V} \in \mathbb{R}^{n,d_f}$, and edge features $\mathcal{E} \in \mathbb{R}^{n,d_e}$, where n , d_f and d_e represent the number of node, node

features dimension and edge features dimension. For pockets, the graph nodes include amino acids within 10 Å to the ligand, X constrains the position of C_α of residues, on which we construct \mathcal{E} via k-nn algorithm. For molecules, the graph nodes include all ligand atoms except Hs, X constrains the atom positions, and we use the molecular bonds as \mathcal{E} .

Gated Geometric Massage Passing. From layer t to $t + 1$, we use the gated geometric massage passing (GGMP) layer to update 3D graph representations, i.e., $[\mathbf{v}_i^{t+1}, \mathbf{x}_i^{t+1}, \mathbf{n}_i^{t+1}] = \text{GGMP}(\mathbf{v}_i^t, \mathbf{x}_i^t, \mathbf{n}_i^t)$, where \mathbf{n}_i is the direction vector. For molecules, \mathbf{n}_i points to the negative neighborhood center of node i ; for pockets, \mathbf{n}_i indicates the position of protein caves. Given 3D conformations, we minimize the pairwise energy function E :

$$E(X, F, \mathcal{E}) = \sum_{(i,j) \in \mathcal{E}} u(\mathbf{v}_i, \mathbf{v}_j, \mathbf{e}_{ij}) g(\langle \mathbf{n}_i, \mathbf{n}_j \rangle, d_{ij}^2) \quad (1)$$

where $d_{ij}^2 = \|\mathbf{x}_i - \mathbf{x}_j\|^2$, both chemical energy $u(\cdot)$ and geometric energy $g(\cdot)$ are considered. By calculating the gradients of \mathbf{x}_i and \mathbf{n}_i , we obtain their updating rules:

$$\begin{aligned} -\frac{\partial E(X, F, \mathcal{E})}{\partial \mathbf{x}_i} &= -\sum_{j \in \mathcal{N}_i} 2u_{ij} \frac{\partial g_{ij}}{\partial d_{ij}^2} (\mathbf{x}_i - \mathbf{x}_j) \\ &\approx \sum_{j \in \mathcal{N}_i} u(\mathbf{v}_i, \mathbf{v}_j, \mathbf{e}_{ij}) \phi_x(d_{ij}^2, \langle \mathbf{n}_i, \mathbf{n}_j \rangle) (\mathbf{x}_i - \mathbf{x}_j) \end{aligned} \quad (2)$$

$$\begin{aligned} -\frac{\partial E(X, F, \mathcal{E})}{\partial \mathbf{n}_i} &= -\sum_{j \in \mathcal{N}_i} u_{ij} \frac{\partial g_{ij}}{\partial \langle \mathbf{n}_i, \mathbf{n}_j \rangle} \mathbf{n}_j \\ &\approx \sum_{j \in \mathcal{N}_i} u(\mathbf{v}_i, \mathbf{v}_j, \mathbf{e}_{ij}) \phi_n(d_{ij}^2, \langle \mathbf{n}_i, \mathbf{n}_j \rangle) \mathbf{n}_j \end{aligned} \quad (3)$$

Note that ϕ_x and ϕ_n are the approximation of $\frac{\partial g_{ij}}{\partial d_{ij}^2}$ and $\frac{\partial g_{ij}}{\partial \langle \mathbf{n}_i, \mathbf{n}_j \rangle}$. Combining graph message passing, we propose the GGMP layer:

$$\mathbf{m}_{ij} = \phi_m(\mathbf{v}_i^t, \mathbf{v}_j^t, \mathbf{e}_{ij}) \quad (4)$$

$$\mathbf{g}_{ij} = \phi_g(d_{ij}^2, \langle \mathbf{n}_i^t, \mathbf{n}_j^t \rangle) \quad (5)$$

$$\mathbf{h}_i^{t+1} = \phi_h(\mathbf{h}_i^t, \sum_{j \in \mathcal{N}_i} \mathbf{m}_{ij} \mathbf{g}_{ij}) \quad (6)$$

$$\mathbf{x}_i^{t+1} = \mathbf{x}_i^t + \lambda \sum_{j \in \mathcal{N}_i} u(\mathbf{m}_{ij}) \phi_x(\mathbf{g}_{ij}) (\mathbf{x}_i^t - \mathbf{x}_j^t) \quad (7)$$

$$\mathbf{n}_i^{t+1} = \mathbf{n}_i^t + \lambda \sum_{j \in \mathcal{N}_i} u(\mathbf{m}_{ij}) \phi_n(\mathbf{g}_{ij}) \mathbf{n}_j^t \quad (8)$$

where ϕ_* and u are approximated by neural networks, λ is a hyperparameter, and $\mathbf{n}_i^0 = -\sum_{j \in \mathcal{N}(i)} \mathbf{x}_j^0 / \|\sum_{j \in \mathcal{N}(i)} \mathbf{x}_j^0\|$.

3.3 Contrastive Loss

In contrastive learning, the biased negative sampling impairs model performance by sampling false negative data during training. Previous methods [9, 39] address this problem with the assumption that false-negative samples are uniformly distributed under the classification setting. We propose chemical knowledge-based sampling to better address this issue, where fingerprint similarity is used to measure the probability of negative ligands. Interestingly, the change in sampling distribution is equivalent to the design of a weighted loss, and we provide a comprehensive understanding from the perspective of contrastive loss.

Uni-contrastive Loss. Given the pocket $x \sim p$, we draw positive ligands \hat{x}^+ from the distribution \hat{p}_x^+ of bindable molecules and negative ligands $\{\hat{x}_i^-\}_{i=1}^N$ from the distribution \hat{q} of non-bindable ones. By default, the positive ligands are determined by the pocket-ligand complexes, while negative ones are uniformly sampled from the ligand sets. We use pocket model f and ligand model \hat{f} to learn the latent representations \mathbf{h} , $\hat{\mathbf{h}}^+$ and $\{\hat{\mathbf{h}}_i^-\}_{i=1}^N$, where the proxy task is to maximize the positive similarity $s^+(\mathbf{h}, \hat{\mathbf{h}}^+)$ against the negative similarities $s_i^-(\mathbf{h}, \hat{\mathbf{h}}_i^-)$, $i = 1, 2, \dots$, resulting in:

$$L_{\text{Uni}} = \mathbb{E}_{\substack{x \sim p, \hat{x}^+ \sim \hat{p}_x^+ \\ \{\hat{x}_i^-\}_{i=1}^N \sim \hat{q}}} \left[\log \left(1 + \frac{Q}{N} \sum_{i=1}^N \frac{s_i^-(\mathbf{h}, \hat{\mathbf{h}}_i^-)}{s^+(\mathbf{h}, \hat{\mathbf{h}}^+)} \right) \right] \quad (9)$$

where Q and N are constants. For each data sample x , the gradients contributed to s^+ and s_i^- are:

$$\frac{\partial L}{\partial s^+} = \frac{1}{1 + \sum_{i=1}^N s_i^- / s^+} \sum_{i=1}^N \frac{\partial s_i^- / s^+}{\partial s^+} \quad (10)$$

$$\frac{\partial L}{\partial s_i^-} = \frac{1}{1 + \sum_{i=1}^N s_i^- / s^+} \frac{\partial s_i^- / s^+}{\partial s_i^-} \quad (11)$$

The L_{Uni} provides balanced gradient to positive and negative samples, i.e., $\frac{\partial L}{\partial s^+} = \sum_i \frac{\partial L}{\partial s_i^-}$. One can verify that InfoNCE is the special case of L_{Uni} by setting $s^+(\mathbf{h}, \hat{\mathbf{h}}^+) = e^{\gamma \mathbf{h}^T \mathbf{h}^+}$ and $s_i^-(\mathbf{h}, \hat{\mathbf{h}}_i^-) = e^{\gamma \mathbf{h}^T \mathbf{h}_i^-}$.

DebiasedInfoNCE. Uniformly sampling negative ligands from the data distribution \hat{q} could mistaken positive samples as negative ones. Denote $h(\cdot)$ as the labeling function, [9] suggests to draw negative samples from the real negative distribution $\hat{q}_x^-(\hat{x}^-) = p(\hat{x}^- | h(\hat{x}^-) \neq h(x))$. To handle the $\{h(\hat{x}^-) \neq h(x)\}$ event, the joint distribution $p(\hat{x}, c) = p(\hat{x}|c)p(c)$ over data \hat{x} and label c is considered. Assume the class probability $p(c) = \tau^+$ is uniform, and let $\tau^- = 1 - \tau^+$ be the probability of observing any different class, \hat{q} could be decomposed as $\tau^- \hat{q}_x^-(\hat{x}^-) + \tau^+ \hat{q}_x^+(\hat{x}^-)$. Therefore, $\hat{q}_x^- = (\hat{q} - \tau^+ \hat{q}_x^+) / \tau^-$, and the DebiasedInfoNCE is:

$$L_{\text{Debiased}} = \mathbb{E}_{\substack{x \sim p, \hat{x}^+ \sim \hat{p}_x^+, \\ \{\hat{x}_i^-\}_{i=1}^N \sim \hat{q}_x}} \left[\log \left(1 + \frac{Q}{N} \sum_{i=1}^N \frac{s_i^-(\mathbf{h}, \hat{\mathbf{h}}_i^-)}{s^+(\mathbf{h}, \hat{\mathbf{h}}^+)} \right) \right] \quad (12)$$

where $s^+(\mathbf{h}, \hat{\mathbf{h}}^+) = e^{\mathbf{h}^T \hat{\mathbf{h}}^+}$, $s_i^-(\mathbf{h}, \hat{\mathbf{h}}_i^-) = e^{\mathbf{h}^T \hat{\mathbf{h}}_i^-}$. With mild assumptions, the approximated debaised InfoNCE can be written as:

$$\mathbb{E}_{\substack{x \sim p, \hat{x}^+ \sim p_x^+, \\ \{\hat{x}_i^-\}_{i=1}^N \sim \hat{q}}} \left[\log \left(1 + \frac{Q}{\tau^-} \sum_{i=1}^N (e^{\mathbf{h}^T \hat{\mathbf{h}}_i^-} - \tau^-) \right) \right] \quad (13)$$

ChemInfoNCE. Although DebiasedInfoNCE solves the problem of sampling bias to some extent, it suffers from some shortcomings. Firstly, for classification with discrete labels, the assumption of uniform class probabilities may be too strong, especially for the unbalanced dataset. Secondly, when it comes to regression, molecules have continuous chemical properties and the event $\{h(\hat{x}) \neq h(\hat{x}^-)\}$ can not describe the validity of negative data. To address these issues, we introduce a new event $\{\text{sim}(\hat{x}, \hat{x}^-) < \tau\}$ to measure the validity of negative samples, where $\text{sim}(\cdot, \cdot)$ is the function of chemical similarity. The underlying assumption is that molecules with lower chemical similarity to the reference ligand are more likely to be negative samples.

$$\begin{aligned} q_x^-(\hat{x}^-) &:= q(\hat{x}^- | \text{sim}(x, \hat{x}^-) < \tau) \\ &\propto \max(1 - \text{sim}(x, \hat{x}^-) - \tau, 0) \cdot p(\hat{x}^-) \end{aligned} \quad (14)$$

By denoting $w_i = \max(1 - \text{sim}(x, \hat{x}^-) - \tau, 0)$, the final ChemInfoNCE can be simplified as:

$$L_{\text{Chem}} \approx \mathbb{E}_{\substack{x \sim p, \hat{x}^+ \sim p_x^+, \\ \{\hat{x}_i^-\}_{i=1}^N \sim \hat{q}}} \left[\log \left(1 + \sum_{i=1}^N (\rho_i e^{\mathbf{h}^T \hat{\mathbf{h}}_i^- - \mathbf{h}^T \hat{\mathbf{h}}^+}) \right) \right] \quad (15)$$

where $\rho_i = \frac{w_i}{\sum_{i=1}^N w_i}$.

4 Experiments

In this section, we conduct extensive experiments to verify the effectiveness of the proposed method from three perspectives:

1. **Ligand:** Could the ligand model provide competitive results in predicting molecular properties?
2. **Pocket:** How does the pre-trained pocket model perform on the pocket matching tasks?
3. **Pocket-ligand:** Could the joint model find potential binding pocket-ligand pairs, i.e., virtual screening?

Table 2. Molecule property prediction. We compare different methods across 9 benchmarks. The best and sub-optimum results are highlighted in **bold** and underline.

Methods Dataset	Classification (AUC-ROC % \uparrow)						Regression (RMSE \downarrow)		
	BBBP	BACE	ClinTox	Tox21	ToxCast	SIDER	ESOL	FreeSolv	Lipo
#Molecules	2039	1513	1478	7831	8575	1427	1128	642	4200
#Tasks	1	1	2	12	617	27	1	1	1
D-MPNN	71.0(0.3)	80.9(0.6)	90.6(0.6)	75.9(0.7)	65.5(0.3)	57.0(0.7)	1.050(0.008)	2.082(0.082)	0.683(0.016)
Attentive FP	64.3(1.8)	78.4(0.02)	84.7(0.3)	76.1(0.5)	63.7(0.2)	60.6(3.2)	0.877(0.029)	2.073(0.183)	0.721(0.001)
N-Gram _{RF}	69.7(0.6)	77.9(1.5)	77.5(4.0)	74.3(0.4)	–	66.8(0.7)	1.074(0.107)	2.688(0.085)	0.812(0.028)
N-Gram _{xCB}	69.1(0.8)	79.1(1.3)	87.5(2.7)	75.8(0.9)	–	65.5(0.7)	1.083(0.082)	5.061(0.744)	2.072(0.030)
MolCLR	72.2(2.1)	82.4(0.9)	91.2(3.5)	75.0(0.2)	–	58.9(1.4)	1.271(0.040)	2.594(0.249)	0.691(0.004)
PretrainGNN	68.7(1.3)	84.2(0.7)	72.6(1.5)	78.1(0.6)	65.7(0.6)	62.7(0.8)	1.100(0.006)	2.764(0.002)	0.739(0.003)
GraphMVP-G	70.8(0.5)	79.3(1.5)	79.1(2.8)	75.9(0.5)	63.1(0.2)	60.2(1.1)	–	–	–
GraphMVP-C	72.4(1.6)	81.2(0.9)	76.3(1.9)	74.4(0.2)	63.1(0.4)	63.9(1.2)	–	–	–
3DInfomax	69.1(1.1)	79.4(1.9)	59.4(3.2)	74.5(0.7)	64.4(0.9)	53.4(3.3)	0.894(0.028)	2.34(0.227)	0.695(0.012)
MICRO-graph	77.2(2.0)	84.4(1.1)	77.0(0.8)	77.0(0.8)	65.2(0.8)	56.7(1.1)	–	–	–
GROVER _{base}	70.0(0.1)	82.6(0.7)	81.2(3.0)	74.3(0.1)	65.4(0.4)	64.8(0.6)	0.983(0.090)	2.176(0.052)	0.817(0.008)
GROVER _{large}	69.5(0.1)	81.0(1.4)	76.2(3.7)	73.5(0.1)	65.3(0.5)	65.4(0.1)	0.895(0.017)	2.272(0.051)	0.823(0.010)
GEM	72.4(0.4)	85.6(1.1)	90.1(1.3)	78.1(0.1)	69.2(0.4)	67.2(0.4)	0.798(0.029)	1.877(0.094)	0.660(0.008)
Uni-Mol	72.9(0.6)	85.7(0.2)	91.9(1.8)	79.6(0.5)	69.6(0.1)	65.9(1.3)	0.788(0.029)	1.620(0.035)	0.603(0.010)
CoSP _{base} (w/o pre-train)	71.1(0.5)	82.2(1.4)	89.3(1.5)	73.8(0.4)	62.6(0.7)	61.7(1.2)	1.243(0.045)	2.135(0.045)	0.864(0.023)
CoSP _{base} (DebiasedInfoNCE)	72.3(0.4)	83.5(1.2)	90.2(1.3)	75.4(0.6)	64.7(0.4)	62.9(1.3)	0.843(0.038)	1.857(0.043)	0.748(0.029)
CoSP _{base} (ChemInfoNCE)	73.1(0.3)	84.3(1.1)	91.3(1.5)	78.5(0.1)	69.3(0.2)	64.7(1.5)	0.785(0.029)	1.752(0.042)	0.621(0.012)
CoSP _{large}	73.6(0.4)	85.9(0.9)	91.2(1.2)	79.3(0.1)	70.0(0.2)	66.4(1.2)	0.783(0.023)	1.715(0.017)	0.598(0.011)

4.1 Pre-training Setup

Pre-training Dataset. We adopt BioLip [64] dataset for pre-training CoSP_{base}, where the original BioLip contains 573,225 entries up to 2022.04.01. Compared to PDBBind [54] with 23,496 complexes, BioLip contains more complexes that lack binding affinity, thus could provide a more comprehensive view of binding mode analysis. To focus on the drug-like molecules and their binding pockets, we filtered out other unrelated complexes that contain peptides, DNA, RNA, single ions, etc. In addition, we augment the pretraining data with the CrossDock dataset [18] to develop CoSP_{large}.

Experimental Setting. We pre-train CoSP_{base} with 6 layer GGMPs via ChemInfoNCE loss, where the hidden feature dimension is 128. We train the model for 50 epochs using Adam optimizer on NVIDIA A100s, where the initial learning rate is 0.01 and the batch size is 100. The chemical ligand similarity is calculated by RDKit [28]. To achieve better performance, CoSP_{large} extends the 6-layer GNN to 12 layers, with hidden dimensions from 128 to 1024, and uses augmented dataset (BioLip+CrossDock).

4.2 Downstream Task 1: Molecule Property Prediction

Experimental Setup. Could the model learn expressive features for molecule classification and regression tasks? We evaluate CoSP on 9 benchmarks collected by MoleculeNet [61]. Following previous researches, we use scaffold splitting to generate train/validation/test set with a ratio of 8:1:1. We report AUC-ROC and RMSE metrics for classification and regression tasks, respectively. The mean and standard deviations of results over three random seeds are provided by default. We finetune the model using the similar code of MGSSL [73].

Baselines. We evaluate CoSP against a broad of baselines, including D-MPNN [65], Attentive FP [63], N-Gram_{RF}, N-Gram_{XGB} [30], MolCLR [56], PretrainGNN [23], GraphMVP-G, GraphMVP-C [32], 3DInfomax [47], MICROgraph [71], GROVER_{base}, GROVER_{large} [40], GEM [16], and Uni-Mol [76]. Most these baselines are pre-training methods, except for N-Gram_{RF} and N-Gram_{XGB}. Some of the methods mentioned in the related works are not included because the experimental setup, e.g., data splitting, may be different.

Results and Analysis. We show results in Table.2. The main observations are: (1) CoSP_{large} could achieve the best results on 4/9 downstream tasks, and top-3 results on 9/9 downstream tasks. (2) Pre-training techniques help improve the model’s generalization ability, and the model could learn expressive molecular features via co-supervised pre-training. By extending the model size and pre-training data volumn, CoSP_{large} achieves non-trivial performance gains compared to CoSP_{base}. (3) Through ablation studies, we further identified the superiority of ChemInfoNCE over DebaisedInfoNCE by achieving consistent performance gains on various datasets.

Table 3. Pocket matching results. We compare different methods across 10 benchmarks.

Category	Methods	Classification (AUC-ROC \uparrow)									
		D1	D1.2	D2	D3	D4	D5	D5.2	D6	D6.2	D7
Classical baselines	Cavbase	<u>0.98</u>	0.91	0.87	0.65	0.64	0.60	0.57	0.55	0.55	0.82
	FuzCav	0.94	<u>0.99</u>	<u>0.99</u>	0.69	0.58	0.55	0.54	0.67	0.73	0.77
	FuzCav (PDB)	0.94	<u>0.99</u>	0.98	0.69	0.58	0.56	0.54	0.65	0.72	0.77
	grim	0.69	0.97	0.92	0.55	0.56	0.69	0.61	0.45	0.65	0.70
	grim (PDB)	0.62	0.83	0.85	0.57	0.56	0.61	0.58	0.45	0.50	0.64
	IsoMIF	0.77	0.97	0.70	0.59	0.59	<u>0.75</u>	0.81	0.62	0.62	0.87
	KRIPO	0.91	1.00	0.96	0.60	0.61	0.76	<u>0.77</u>	0.73	<u>0.74</u>	0.85
	PocketMatch	0.82	0.98	0.96	0.59	0.57	0.66	0.60	0.51	0.51	0.82
	ProBiS	1.00	1.00	1.00	0.47	0.46	0.54	0.55	0.50	0.50	0.85
	RAPMAD	0.85	0.83	0.82	0.61	0.63	0.55	0.52	0.60	0.60	0.74
	shaper	0.96	0.93	0.93	0.71	0.76	0.65	0.65	0.54	0.65	0.75
	shaper (PDB)	0.96	0.93	0.93	0.71	0.76	0.66	0.64	0.54	0.65	0.75
	VolSite/shaper	0.93	<u>0.99</u>	0.78	0.68	0.76	0.56	0.58	0.71	0.76	0.77
	VolSite/shaper (PDB)	0.94	1.00	0.76	0.68	0.76	0.57	0.56	0.50	0.57	0.72
	SiteAlign	0.97	1.00	1.00	<u>0.85</u>	<u>0.80</u>	0.59	0.57	0.44	0.56	0.87
	SiteEngine	0.96	1.00	1.00	0.82	0.79	0.64	0.57	0.55	0.55	0.86
	SiteHopper	<u>0.98</u>	0.94	1.00	0.75	0.75	0.72	0.81	0.56	0.54	0.77
	SMAP	1.00	1.00	1.00	0.76	0.65	0.62	0.54	0.68	0.68	0.86
	TIFP	0.66	0.90	0.91	0.66	0.66	0.71	0.63	0.55	0.60	0.71
	TIFP (PDB)	0.55	0.74	0.78	0.56	0.57	0.54	0.53	0.56	0.61	0.66
TM-align	1.00	1.00	1.00	0.49	0.49	0.66	0.62	0.59	0.59	<u>0.88</u>	
Deeplearning baseline	DeeplyTough	0.95	0.98	0.90	0.76	0.75	0.67	0.63	0.54	0.54	0.83
Our methods	CoSP _{base} (w/o direction)	0.95	0.95	0.92	0.55	0.58	0.56	0.56	0.53	0.54	0.76
	CoSP _{base}	1.00	1.00	<u>0.99</u>	0.79	<u>0.81</u>	0.62	0.63	0.61	0.62	0.81
	CoSP _{large}	1.00	1.00	1.00	0.87	0.85	<u>0.75</u>	0.74	<u>0.72</u>	<u>0.74</u>	0.90

4.3 Downstream Task 2: Pocket Matching

Experimental Setup. Could the pre-trained model identify chemically similar pockets? We explore the discriminative ability of the pocket model with the pocket matching tasks. To comprehensively understand the potential of the proposed method, we evaluated it on 10 benchmarks recently collected in the ProSPECCTs dataset [15]. For each sub-dataset, the positive and negative pairs of pockets are defined differently according to the research objectives. We summarize five research objectives as **O1**: Whether the model is robust to the pocket definition? **O2**: Whether the model is robust to the pocket flexibility? **O3**: Can the model distinguish between pockets with different properties? **O4**: Whether the model can distinguish dissimilar proteins binding to identical ligands and cofactors? **O5**: How about the performance on real applications? We report the AUC-ROC scores on all benchmarks.

Baselines. We compare CoSP with both classical and deeplearning baselines. The classical methods can be divided into profile-based, graph-based and grid-based ones. The profile-based methods encode topological, physicochemical and statistical properties in a unified way for comparing various pockets, e.g., SiteAlign [42], RAPMAD [27], TIFP [13], FuzCav [58], PocketMatch [69], SMAP [62], TM-align [72], KRIPPO [60] and Grim [13]. The graph-based methods adopt isomorphism detection algorithm to find common motifs, e.g., Cavbase [43], IsoMIF [8], ProBiS [26]. Grid-based methods represent pockets by regularly spaced pharmacophoric grid points, e.g., VolSite/Shaper [12]. Another tools include SiteEngines [45] and SiteHopper [3]. We also compare with the recent deeplearning model-DeeplyTough [46].

Results and Analysis. We present the pocket matching results in Table. 3, where the pre-trained model achieves competitive results in most cases. Specifically, CoSP is robust to pocket definition (**O1**) and achieves the highest AUC scores in D1 and D1.2. The robustness also remains when considering conformational variability (**O2**), where CoSP_{large} achieves 1.00 AUC score in D2. It should be noted that robustness to homogeneous pockets does not mean that the model has poor discrimination; on the contrary, the model could identify pockets with different physicochemical and shape properties (**O3**) in D3 and D4. Compared with previous deep learning methods (DeeplyTough), CoSP_{large} provides better performance in distinguishing different pockets bound to the same ligands and cofactors (**O4**), refer to the results of D5, D5.2, D6 and D6.2. Last but not least, CoSP_{large} showed good potential for practical applications (**O5**) with 0.90 AUC score. In addition, we found that pocket direction plays a key role in extracting pocket features, which is helpful to indicate the location of the pocket cavity. As shown in Table. 3, the performance of pocket matching will be degraded if the directional feature n is removed.

Table 4. Virtual screening results on DUD-E.

	DUD-E				
	AUC-ROC \uparrow	0.5% RE \uparrow	1.0% RE \uparrow	2.0% RE \uparrow	5.0% RE \uparrow
Vina	0.716	9.139	7.321	5.881	4.444
RF-score	0.622	5.628	4.274	3.499	2.678
NNScore	0.584	4.166	2.980	2.460	1.891
Graph CNN	0.886	44.406	29.748	19.408	10.735
3DCNN	0.868	42.559	29.654	19.363	10.710
DrugVQA	0.972	88.17	58.71	35.06	17.39
GanDTi	0.997	71.13	<u>68.78</u>	<u>49.40</u>	<u>19.79</u>
AttentionSiteDTI	0.971	<u>101.74</u>	59.92	35.07	16.74
CoSP _{large}	<u>0.996</u>	111.764	78.426	55.535	22.318

4.4 Downstream Task 3: Virtual Screening

Experimental Setup. Could the model distinguish molecules most likely to bind to the given pocket? We evaluate CoSP on the DUD-E [34] dataset which consists of 102 targets across different protein families. For each target, DUD-E provides 224 actives (positive examples) and 10,000 decoy ligands (negative examples) in average. The decoys were calculated by choosing them to be physically similar but topologically different from the actives. During finetuning, we use the same data splitting as GraphCNN [51], and report the AUC-ROC and ROC enrichment (RE) scores. Note that $x\%$ RE indicates the ratio of the true positive rate (TPR) to the false positive rate (FPR) at $x\%$ FPR value.

Baselines. We compare CoSP_{large} with AutoDock Vina [53], RF-score [2], NNScore [14], 3DCNN [38], GraphCNN [51], DrugVQA [75], GanDTi [55], and AttentionSiteDTI [68]. AutoDock Vina is a commonly used open-source program for doing molecular docking. RF-score use random forest capture protein-ligand binding effects. Other methods use deeplearning models to learn the protein-ligand binding.

Results and Analysis. We present results in Table.4, and observe that: (1) Random forest and MLP-based RF-score and NNScore achieve competitive results to Vina, indicating the potential of machine learning in virtual screening. (2) Deeplearning-based Graph CNN, 3DCNN, DrugVQA, GanDTi, and Attention-SiteDTI significantly outperforms both RF-score and NNScore. (3) CoSP_{large} achieves competitive AUC score and outperforms all baselines in RE scores. The improvement of CoSP_{large} suggests that the model can effectively learn protein-ligand interactions from the pre-training data. (4) In addition, we select Top 1% ligands identified by the model as actives for the given pocket and use AutoDock Vina to validate the docking results. In Fig. 2, the visual results show that our model can identify high-affinity ligands, which is helpful for drug discovery.

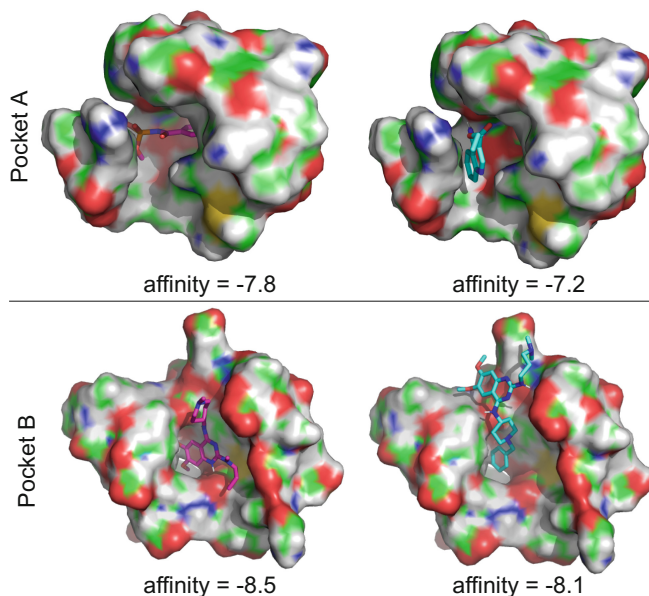


Fig. 2. Two examples of virtual screening. For each pocket, we choose two ligands that are Top 1% active molecules as identified by the model. We use AutoDock Vina to generate molecular binding pose and compute the affinity score.

5 Conclusion

This paper proposes a co-supervised pre-training framework to learn the joint pocket and ligand spaces via chemically inspired contrastive loss. The pre-trained model could achieve competitive results on molecule property predictions, pocket matching, and virtual screening. We hope the unified modeling framework could further advance the development of AI-guided drug discovery.

Acknowledgements. We thank the open-sourced codes of previous studies. This work was supported by the National Key R&D Program of China (Project 2022ZD0115100), the National Natural Science Foundation of China (Project U21A20427), the Research Center for Industries of the Future (Project WU2022C043).

Ethical Statement. Our submission does not involve any ethical issues, including but not limited to privacy, security, etc.

References

1. Altalib, M.K., Salim, N.: Similarity-based virtual screen using enhanced siamese deep learning methods. *ACS omega* **7**(6), 4769–4786 (2022)
2. Ballester, P.J., Mitchell, J.B.: A machine learning approach to predicting protein-ligand binding affinity with applications to molecular docking. *Bioinformatics* **26**(9), 1169–1175 (2010)
3. Batista, J., Hawkins, P.C., Tolbert, R., Geballe, M.T.: Sitehopper—a unique tool for binding site comparison. *J. Cheminform.* **6**(1), 1–1 (2014)
4. Batzner, S., et al.: E (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature Commun.* **13**(1), 1–11 (2022)
5. Boström, J., Hogner, A., Schmitt, S.: Do structurally similar ligands bind in a similar fashion? *J. Med. Chem.* **49**(23), 6716–6725 (2006)
6. Brandstetter, J., Hesselink, R., van der Pol, E., Bekkers, E., Welling, M.: Geometric and physical quantities improve e (3) equivariant message passing. *arXiv preprint arXiv:2110.02905* (2021)
7. Chaffey, N.: Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P. *Molecular Biology of the Cell*. 4th edn. (2003)
8. Chartier, M., Najmanovich, R.: Detection of binding site molecular interaction field similarities. *J. Chem. Inform. Model.* **55**(8), 1600–1615 (2015)
9. Chuang, C.Y., Robinson, J., Lin, Y.C., Torralba, A., Jegelka, S.: Debiased contrastive learning. *Adv. Neural Inform. Process. Syst.* **33**, 8765–8775 (2020)
10. Cohen, T., Welling, M.: Group equivariant convolutional networks. In: International conference on machine learning, pp. 2990–2999. PMLR (2016)
11. Dankwah, K.O., Mohl, J.E., Begum, K., Leung, M.Y.: Understanding the binding of the same ligand to gpcrs of different families. In: 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 2494–2501 (2021). <https://doi.org/10.1109/BIBM52615.2021.9669761>
12. Desaphy, J., Azdimousa, K., Kellenberger, E., Rognan, D.: Comparison and drug-gability prediction of protein-ligand binding sites from pharmacophore-annotated cavity shapes (2012)
13. Desaphy, J., Raimbaud, E., Ducrot, P., Rognan, D.: Encoding protein-ligand interaction patterns in fingerprints and graphs. *J. Chem. Inform. Model.* **53**(3), 623–637 (2013)
14. Durrant, J.D., McCammon, J.A.: Nnscore: a neural-network-based scoring function for the characterization of protein-ligand complexes. *J. Chem. Inform. Model.* **50**(10), 1865–1871 (2010)
15. Ehrhart, C., Brinkjost, T., Koch, O.: A benchmark driven guide to binding site comparison: an exhaustive evaluation using tailor-made data sets (prospeccts). *PLoS Comput. Biol.* **14**(11), e1006483 (2018)
16. Fang, X.: Geometry-enhanced molecular representation learning for property prediction. *Nature Mach. Intell.* **4**(2), 127–134 (2022)
17. Fang, Y., Yang, H., Zhuang, X., Shao, X., Fan, X., Chen, H.: Knowledge-aware contrastive molecular graph learning. *arXiv preprint arXiv:2103.13047* (2021)
18. Francoeur, P.G.: Three-dimensional convolutional neural networks and a cross-docked data set for structure-based drug design. *J. Chem. Inform. Model.* **60**(9), 4200–4215 (2020)
19. Fuchs, F., Worrall, D., Fischer, V., Welling, M.: Se (3)-transformers: 3d rotation equivariant attention networks. *Adv. Neural Inform. Process. Syst.* **33**, 1970–1981 (2020)

20. Ganea, O.E., et al.: Independent se (3)-equivariant models for end-to-end rigid protein docking. arXiv preprint [arXiv:2111.07786](https://arxiv.org/abs/2111.07786) (2021)
21. Gao, Z., Tan, C., Li, S., et al.: Alphadesign: A graph protein design method and benchmark on alphafolddb. arXiv preprint [arXiv:2202.01079](https://arxiv.org/abs/2202.01079) (2022)
22. Guan, J., Qian, W.W., Ma, W.Y., Ma, J., Peng, J., et al.: Energy-inspired molecular conformation optimization. In: International Conference on Learning Representations (2021)
23. Hu, W., et al.: Strategies for pre-training graph neural networks. arXiv preprint [arXiv:1905.12265](https://arxiv.org/abs/1905.12265) (2019)
24. Jing, B., Eismann, S., Suriana, P., Townshend, R.J., Dror, R.: Learning from protein structure with geometric vector perceptrons. arXiv preprint [arXiv:2009.01411](https://arxiv.org/abs/2009.01411) (2020)
25. Kinnings, S.L., Liu, N., Buchmeier, N., Tonge, P.J., Xie, L., Bourne, P.E.: Drug discovery using chemical systems biology: repositioning the safe medicine comtan to treat multi-drug and extensively drug resistant tuberculosis. *PLoS Comput. Biol.* **5**(7), e1000423 (2009)
26. Konc, J., Janežič, D.: Probis algorithm for detection of structurally similar protein binding sites by local structural alignment. *Bioinformatics* **26**(9), 1160–1168 (2010)
27. Krotzky, T., Grunwald, C., Egerland, U., Klebe, G.: Large-scale mining for similar protein binding pockets: with RAPMAD retrieval on the fly becomes real. *J. Chem. Inform. Model.* **55**(1), 165–179 (2015)
28. Landrum, G.: Rdkit: Open-source cheminformatics software (2016). https://github.com/rdkit/rdkit/releases/tag/Release_2016.09.4
29. Li, P., et al.: An effective self-supervised framework for learning expressive molecular global representations to drug discovery. *Brief. Bioinform.* **22**(6), bbab109 (2021)
30. Liu, S., Demirel, M.F., Liang, Y.: N-gram graph: simple unsupervised representation for graphs, with applications to molecules. *Adv. Neural Inform. Process. Syst.* **32** (2019)
31. Liu, S., Wang, H., Liu, W., Lasenby, J., Guo, H., Tang, J.: Pre-training molecular graph representation with 3D geometry-rethinking self-supervised learning on structured data
32. Liu, S., Wang, H., Liu, W., Lasenby, J., Guo, H., Tang, J.: Pre-training molecular graph representation with 3D geometry. arXiv preprint [arXiv:2110.07728](https://arxiv.org/abs/2110.07728) (2021)
33. Lu, A.X., Zhang, H., Ghassemi, M., Moses, A.: Self-supervised contrastive learning of protein representations by mutual information maximization. *BioRxiv* (2020)
34. Mysinger, M.M., Carchia, M., Irwin, J.J., Shoichet, B.K.: Directory of useful decoys, enhanced (dud-e): better ligands and decoys for better benchmarking. *J. Med. Chem.* **55**(14), 6582–6594 (2012)
35. Nguyen, T., Le, H., Quinn, T.P., Nguyen, T., Le, T.D., Venkatesh, S.: Graphdta: predicting drug-target binding affinity with graph neural networks. *Bioinformatics* **37**(8), 1140–1147 (2021)
36. Oord, A.v.d., Li, Y., Vinyals, O.: Representation learning with contrastive predictive coding. arXiv preprint [arXiv:1807.03748](https://arxiv.org/abs/1807.03748) (2018)
37. Pu, L., Govindaraj, R.G., Lemoine, J.M., Wu, H.C., Brylinski, M.: Deepdrug3d: classification of ligand-binding pockets in proteins with a convolutional neural network. *PLoS Comput. Biol.* **15**(2), e1006718 (2019)
38. Ragoza, M., Hochuli, J., Idrobo, E., Sunseri, J., Koes, D.R.: Protein-ligand scoring with convolutional neural networks. *J. Chem. Inform. Model.* **57**(4), 942–957 (2017)
39. Robinson, J., Chuang, C.Y., Sra, S., Jegelka, S.: Contrastive learning with hard negative samples. arXiv preprint [arXiv:2010.04592](https://arxiv.org/abs/2010.04592) (2020)

40. Rong, Y.: Self-supervised graph transformer on large-scale molecular data. *Adv. Neural Inform. Process. Syst.* **33**, 12559–12571 (2020)
41. Satorras, V.G., Hoogeboom, E., Welling, M.: E (n) equivariant graph neural networks. In: *International Conference on Machine Learning*, pp. 9323–9332. PMLR (2021)
42. Schalon, C., Surgand, J.S., Kellenberger, E., Rognan, D.: A simple and fuzzy method to align and compare druggable ligand-binding sites. *Proteins: Struct., Funct., Bioinform.* **71**(4), 1755–1778 (2008)
43. Schmitt, S., Kuhn, D., Klebe, G.: A new method to detect related function among proteins independent of sequence and fold homology. *J. Mol. Biol.* **323**(2), 387–406 (2002)
44. Shrivastava, A.D., Kell, D.B.: Fragnet, a contrastive learning-based transformer model for clustering, interpreting, visualizing, and navigating chemical space. *Molecules* **26**(7), 2065 (2021)
45. Shulman-Peleg, A., Nussinov, R., Wolfson, H.J.: Siteengines: recognition and comparison of binding sites and protein-protein interfaces. *Nucleic Acids Res.* **33**(suppl_2), W337–W341 (2005)
46. Simonovsky, M., Meyers, J., Meyers, J.: Deeplytough: learning structural comparison of protein binding sites. *J. Chem. Inform. Model.* **60**(4), 2356–2366 (2020)
47. Stärk, H., et al.: 3d infomax improves GNNs for molecular property prediction. *arXiv preprint [arXiv:2110.04126](https://arxiv.org/abs/2110.04126)* (2021)
48. Sturmfels, P., Vig, J., Madani, A., Rajani, N.F.: Profile prediction: An alignment-based pre-training task for protein sequence models. *arXiv preprint [arXiv:2012.00195](https://arxiv.org/abs/2012.00195)* (2020)
49. Sun, M., Xing, J., Wang, H., Chen, B., Zhou, J.: Mocl: data-driven molecular fingerprint via knowledge-aware contrastive learning from molecular graph. In: *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pp. 3585–3594 (2021)
50. Thomas, N., Smidt, T., Kearnes, S., Yang, L., Li, L., Kohlhoff, K., Riley, P.: Tensor field networks: Rotation-and translation-equivariant neural networks for 3D point clouds. *arXiv preprint [arXiv:1802.08219](https://arxiv.org/abs/1802.08219)* (2018)
51. Torng, W., Altman, R.B.: Graph convolutional neural networks for predicting drug-target interactions. *J. Chem. Inform. Model.* **59**(10), 4131–4149 (2019)
52. Tosco, P., Stiefl, N., Landrum, G.: Bringing the MMFF force field to the RDKit: implementation and validation. *J. Cheminform.* **6**(1), 1–4 (2014)
53. Vina, A.: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading *trott, oleg; olson, arthur j. J. Comput. Chem* **31**(2), 455–461 (2010)
54. Wang, R., Fang, X., Lu, Y., Yang, C.Y., Wang, S.: The pdbname database: methodologies and updates. *J. Med. Chem.* **48**(12), 4111–4119 (2005)
55. Wang, S., Shan, P., Zhao, Y., Zuo, L.: Gandti: a multi-task neural network for drug-target interaction prediction. *Comput. Biol. Chem.* **92**, 107476 (2021)
56. Wang, Y., Wang, J., Cao, Z., Barati Farimani, A.: Molecular contrastive learning of representations via graph neural networks. *Nature Mach. Intell.* **4**(3), 279–287 (2022)
57. Weber, A., et al.: Unexpected nanomolar inhibition of carbonic anhydrase by cox-2-selective celecoxib: new pharmacological opportunities due to related binding site recognition. *J. Med. Chem.* **47**(3), 550–557 (2004)
58. Weill, N., Rognan, D.: Alignment-free ultra-high-throughput comparison of drug-gable protein- ligand binding sites. *J. Chem. Inform. Model.* **50**(1), 123–135 (2010)

59. Willmann, D., et al.: Impairment of prostate cancer cell growth by a selective and reversible lysine-specific demethylase 1 inhibitor. *Int. J. Cancer* **131**(11), 2704–2709 (2012)
60. Wood, D.J., Vlieg, J.d., Wagener, M., Ritschel, T.: Pharmacophore fingerprint-based approach to binding site subpocket similarity and its application to bioisostere replacement. *J. Chem. Inform. Model.* **52**(8), 2031–2043 (2012)
61. Wu, Z., et al.: Moleculenet: a benchmark for molecular machine learning. *Chem. Sci.* **9**(2), 513–530 (2018)
62. Xie, L., Bourne, P.E.: Detecting evolutionary relationships across existing fold space, using sequence order-independent profile-profile alignments. *Proc. National Acad. Sci.* **105**(14), 5441–5446 (2008)
63. Xiong, Z., et al.: Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism. *J. Med. Chem.* **63**(16), 8749–8760 (2019)
64. Yang, J., Roy, A., Zhang, Y.: Biolip: a semi-manually curated database for biologically relevant ligand-protein interactions. *Nucleic Acids Res.* **41**(D1), D1096–D1103 (2012)
65. Yang, K., et al.: Analyzing learned molecular representations for property prediction. *J. Chem. Inform. Model.* **59**(8), 3370–3388 (2019)
66. Yang, K.K., Lu, A.X., Fusi, N.: Convolutions are competitive with transformers for protein sequence pretraining. In: *ICLR2022 Machine Learning for Drug Discovery* (2022)
67. Yang, Y., et al.: Computational discovery and experimental verification of tyrosine kinase inhibitor pazopanib for the reversal of memory and cognitive deficits in rat model neurodegeneration. *Chem. Sci.* **6**(5), 2812–2821 (2015)
68. Yazdani-Jahromi, M., et al.: Attentionsitedti: an interpretable graph-based model for drug-target interaction prediction using NLP sentence-level relation classification. *Brief. Bioinform.* **23**(4), bbac272 (2022)
69. Yeturu, K., Chandra, N.: Pocketmatch: a new algorithm to compare binding sites in protein structures. *BMC Bioinform.* **9**(1), 1–17 (2008)
70. Zhang, N., et al.: Ontoprotein: Protein pretraining with gene ontology embedding. *arXiv preprint [arXiv:2201.11147](https://arxiv.org/abs/2201.11147)* (2022)
71. Zhang, S., Hu, Z., Subramonian, A., Sun, Y.: Motif-driven contrastive learning of graph representations. *arXiv preprint [arXiv:2012.12533](https://arxiv.org/abs/2012.12533)* (2020)
72. Zhang, Y., Skolnick, J.: Tm-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **33**(7), 2302–2309 (2005)
73. Zhang, Z., Liu, Q., Wang, H., Lu, C., Lee, C.K.: Motif-based graph self-supervised learning for molecular property prediction. *Adv. Neural Inform. Process. Syst.* **34**, 15870–15882 (2021)
74. Zhang, Z., et al.: Protein representation learning by geometric structure pretraining. *arXiv preprint [arXiv:2203.06125](https://arxiv.org/abs/2203.06125)* (2022)
75. Zheng, S., Li, Y., Chen, S., Xu, J., Yang, Y.: Predicting drug-protein interaction using quasi-visual question answering system. *Nature Mach. Intell.* **2**(2), 134–140 (2020)
76. Zhou, G., et al.: Uni-mol: A universal 3d molecular representation learning framework (2022)