
PharmacoNet: Accelerating Large-Scale Virtual Screening by Deep Pharmacophore Modeling

Seonghwan Seo

Department of Chemistry, KAIST
shwan0106@kaist.ac.kr

Woo Youn Kim

Department of Chemistry, KAIST
AI Institute, KAIST
HITS Inc.
wooyoun@kaist.ac.kr

Abstract

As the size of accessible compound libraries expands to over 10 billion, the need for more efficient structure-based virtual screening methods is emerging. Different pre-screening methods have been developed for rapid screening, but there is still a lack of structure-based methods applicable to various proteins that perform protein-ligand binding conformation prediction and scoring in an extremely short time. Here, we describe for the first time a deep-learning framework for structure-based pharmacophore modeling to address this challenge. We frame pharmacophore modeling as an instance segmentation problem to determine each protein hotspot and the location of corresponding pharmacophores, and protein-ligand binding pose prediction as a graph-matching problem. PharmacoNet is significantly faster than state-of-the-art structure-based approaches, yet reasonably accurate with a simple scoring function. Furthermore, we show the promising result that PharmacoNet effectively retains hit candidates even under the high pre-screening filtration rates. Overall, our study uncovers the hitherto untapped potential of a pharmacophore modeling approach in deep learning-based drug discovery.

1 Introduction

In the realm of modern drug discovery, the search for novel and effective drug molecules has necessitated the exploration of vast chemical spaces. Structure-based virtual screening with large chemical libraries has attracted great attention as a new paradigm for early drug discovery as make-on-demand libraries of drug-like molecules have recently expanded into the tens of billions [1–5]. Recent work reported that increasing the chemical library size from 115 million to 11 billion significantly improved the hit rate from 15% to 30% [6].

Molecular docking is a common approach for structure-based virtual screening. Conventional docking methods involve two steps: generating a large number of binding poses for each initial ligand conformation and ranking binding poses using a well-defined scoring function [7–9]. As the scoring functions are based on the physical energy equation, the highest score among the binding poses – *docking score* – has been considered as a criterion for virtual screening. Nevertheless, the pose space is large, continuous, and contains numerous local optima, requiring binding pose search processes spanning from 1 to 100 seconds for each initial conformation [10–12]. Considering the virtual screening uses over 100 initial conformers for accurate evaluation, the computational demands of molecular docking become a bottleneck in exploring a wider screening space of billions of molecules [1, 5]. This highlights that more efficient structure-based virtual screening approaches are urgently needed.

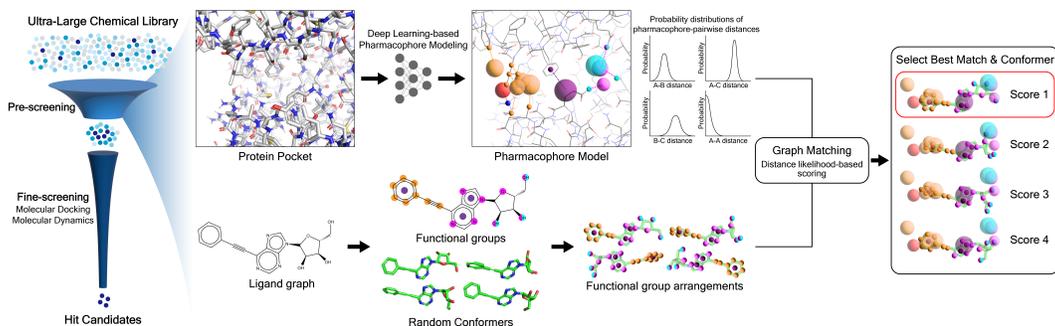


Figure 1: Overview of PharmacoNet.

There are two main approaches for accelerating large-scale virtual screening: the efficient exploration strategy and the pre-screening strategy. Efficient exploration strategies concentrate on effective molecule selection for molecular docking instead of brute force ligand evaluation [13]. Recent studies by Sadybekov et al. [6] using the structured library, Enamine REAL [14], and Graff et al. [15] employing active learning efficiently explore molecular libraries. These strategies focus on exploring promising regions within the vast compound space, reducing computational costs associated with library size expansion.

While these exploration strategies reduce the computational burden by bypassing a substantial portion of the library, this also means that they do not perform any evaluation for most of the molecules. In contrast, the **pre-screening & fine-screening** strategy evaluates all molecules [16]. It begins with a preliminary assessment (pre-screening) of the entire library using fast evaluation methods, followed by a more careful assessment (fine-screening) of retained molecules using computationally intensive methods such as molecular docking and molecular dynamics simulations [17]. However, pre-screening methods often rely on preliminary filters with low accuracy, which can lead to the omission of promising *hit candidates* in large-scale virtual screening, where high filtration rates are imperative. This underscores the importance of the judicious development of pre-screening methods capable of effectively retaining hit candidates at high filtration rates while ensuring rapid processing speed and adaptability to diverse protein targets.

To challenge this issue, we focused on structure-based pharmacophore modeling, a fast structure-based pre-screening approach. A pharmacophore¹ is an abstract description of molecular features with the potential to form the non-covalent interaction (NCI) which is critical for stabilizing protein-ligand complexes, such as aromatic rings and hydrogen bond (H-bond) donors & acceptors. Structure-based pharmacophore modeling identifies the optimal 3D pharmacophore arrangement – *pharmacophore model* – that a ligand should have to ensure optimal supramolecular interactions [18, 19]. It can be carried out by determining *protein hotspots*, which are considered as the key protein functional groups to protein-ligand binding [20], and the optimal location of the pharmacophores to form NCI with each protein hotspot. Pharmacophore modeling simplifies the protein-ligand interaction prediction by streamlining the complicated topology of the protein-ligand complex structure at the pharmacophore level and representing numerous atom-atom pairwise interactions with a few NCIs. Ligands are then evaluated based on their structural similarity to the pharmacophore model, enabling rapid and rational screening without atom-level binding pose prediction or energy scoring.

Structure-based pharmacophore modeling divides into the complex-based approach [21, 22] and the protein-based approach [23–25] according to the use of the binding complex conformation of known active ligands. The protein-based approaches perform pharmacophore modeling solely through structural data of the protein binding site, making them applicable in more general situations, even when the active ligand is unknown. However, most protein-based approaches heavily rely on expert intuition and occasionally require resource-intensive procedures like molecular dynamics, molecular docking, or fragment crystal structures [26]. While several automated protein-based approaches [24, 25] have emerged, they are designed for the fine-screening process, leaving a lack of fully automated protein-based approaches for pre-screening.

¹IUPAC [18] definition: “an ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target and to trigger (or block) its biological response”

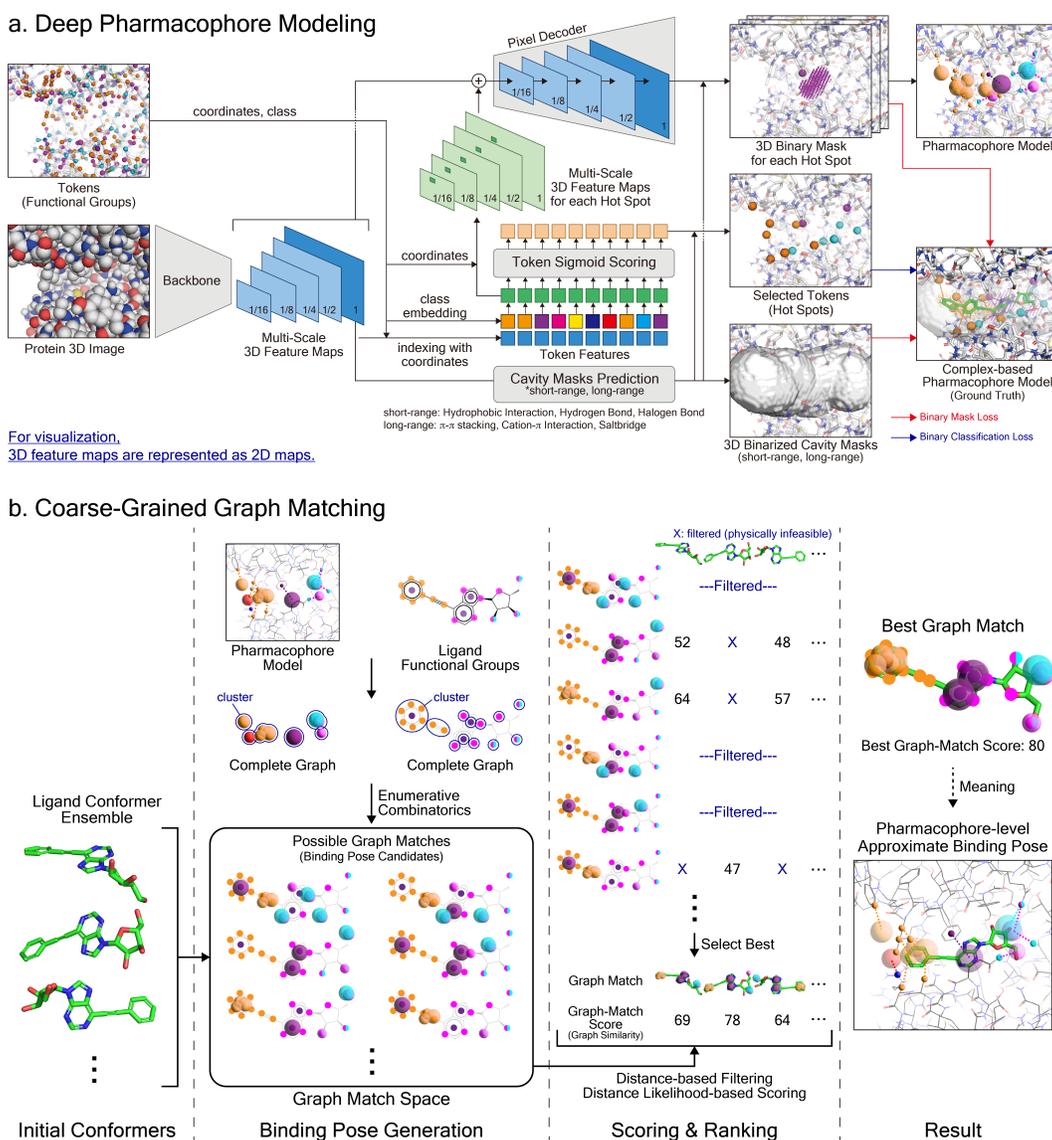


Figure 2: The detailed scheme of PharmacoNet. **a.** The architecture of deep learning model for fully automated protein-based pharmacophore modeling. For visualization, 3D feature maps are represented as 2D maps. The complex-based pharmacophore model is constructed from the crystal structure of the protein-ligand binding complex. (Section 2.1) **b.** The graph-matching algorithm for inexact graph matching. The numbers in the figure are arbitrary values. (Section 2.2)

To this end, we proposed PharmacoNet, the first deep-learning framework for structure-based pharmacophore modeling, as illustrated in Figure 1 and 2. First, it performs fully automated protein-based pharmacophore modeling through framing into image instance segmentation deep learning modeling. Then, it performs coarse-grained graph matching between the pharmacophore model and the ligand at the pharmacophore level to predict the binding pose. Finally, we introduce a novel scoring function based on pharmacophore pairwise distance likelihood instead of atom pairwise interactions between the protein and ligand. PharmacoNet is extremely fast yet reasonably accurate and efficiently retains hit candidates during the pre-screening process. In this work, PharmacoNet evaluated a million molecules for pre-screening potential inhibitors for KRAS-G12C in 11 minutes on 16 CPU cores of Intel Xeon Gold 6234.

2 PharmacoNet Architecture

The high-level protocol of our framework is described in Figure 1. The model architecture used in PharmacoNet for pharmacophore modeling is shown in Figure 2a, Section 2.1, and Appendix 5.1. The model training and training dataset is detailed in Appendix 5.2. The graph matching algorithm is described in Figure 2b, Section 2.2 and Appendix 5.3. The scoring function is explained in Section 2.3. The code is available at <https://github.com/SeonghwanSeo/PharmacoNet>.

2.1 Deep learning-based pharmacophore modeling

2.1.1 Instance image segmentation for pharmacophore modeling

Similar to our approach, Skalic et al. [27] developed LigVoxel, a 3D CNN-based deep learning model for inpainting the chemical functionality map from the pocket image. However, LigVoxel generates only one map for each of the three functional types (aromatic, H-bond donor, and H-bond acceptor), so it is impossible to recognize individual pharmacophoric features or ensure supramolecular interactions with the binding site. In the pharmacophore modeling, it is essential to identify the nature of each pharmacophore, including its coordinates, counterpart protein functional groups, and NCI between them [18]. For the identification of individual pharmacophores and their properties, we frame pharmacophore modeling as an image instance segmentation problem, as shown in Figure 2a.

Image segmentation divides an image into segments based on either categories or instance-level criteria [28, 29]. Semantic segmentation classifies pixels into categories so that the different objects can be grouped into a single segment, similar to LigVoxel. In contrast, instance segmentation achieves the recognition of both the segment and the category for each object by following procedures: 1) object detection, 2) bounding box delineation, 3) category classification, and 4) binary mask (segment) prediction [29]. In the context of tailoring to pharmacophore modeling, the key differences are:

1. Instead of detecting objects from images, our deep learning model selects protein hotspots (instances) among protein functional groups (tokens) in the binding site. When a single functional group can form multiple types of NCIs, it is treated as multiple tokens.
2. Each instance already contains an NCI type (class). Moreover, the possible region (bounding box) for pharmacophore can be obtained from prior knowledge of the maximum length of NCI. Therefore, the deep learning model does not predict its class and bounding box.
3. A single voxel may belong to multiple instances.
4. The deep learning model also addresses an image inpainting problem because the space in the binding site is empty. Therefore, there is no definitive answer for the segmentation.

PharmacoNet completes the pharmacophore modeling process in seconds on a single NVIDIA RTX 3090 GPU. Compared to molecular graph representation, 3D image representation requires more computational cost than molecular graph-based modeling but is advantageous for representing the probability density map of pharmacophores on a grid. Given that pharmacophore modeling occurs only once for a protein binding site before the virtual screening process, this computational demand is not a significant issue.

2.1.2 Ground truth: complex-based pharmacophore model

To establish the ground truth for the protein-based pharmacophore modeling, we use the complex-based pharmacophore model obtained from the crystal structures of protein-ligand complexes in the PDBBind v2020 dataset [30]. We considered the protein-ligand functional group pairs that form NCIs as the pairs of the protein hotspot and the pharmacophore. The identification of NCIs was performed using the Protein-Ligand Interaction Profiler (PLIP) [31]. For pharmacophore modeling, we used the 7 pharmacophore types: hydrophobic carbon, aromatic ring, H-bond donor & acceptor, halogen atom, cation, and anion.

Typically, the length of NCIs is less than 6.0 Å, so it is unnecessary to consider regions far from the pocket cavity when predicting protein hotspots or pharmacophores. We delineate two cavity regions based on the length of the NCI: one within 5.0 Å from the ligand atoms in crystal structure for short-range interactions (hydrophobic interaction, H-bond, and halogen bond) and another within 7.0 Å for long-range interactions (π - π stacking, cation- π interaction, salt bridge).

2.2 Coarse-grained graph matching

To predict the binding pose of a ligand in a protein, PharmacoNet finds the matching pattern between the ligand and the pharmacophore model. The ligands can be abstracted at the pharmacophore level by representing the molecular graph of a ligand in terms of functional groups instead of atoms. Then the pharmacophore model and the ligand functional group arrangement can be represented as 3D complete graphs, so the binding pose prediction problem can be treated as an inexact graph-matching problem. Since inexact graph matching allows for some tolerance between the correspondence between nodes and edges in two graphs, there are many possible graph matches (PGMs). To select the most appropriate graph match, a scoring function that reflects the number of matched nodes and the error in the length of edges is required. Inexact graph matching can be performed between a pharmacophore model and the ligand based on the scoring function and the score of the top-scoring graph match is *graph-matching score* which means the graph similarity.

Ligand functional group arrangements. The complete graph of the ligand is denoted as $\mathcal{G}^L = (\mathcal{V}^L, \mathcal{E}^L)$, where each node $v_i^l \in \mathcal{V}^L$ represents the ligand functional group and it contains the set of its possible pharmacophore types \mathbf{T}_i^l . The edge is denoted as $e_{ij}^l \in \mathcal{E}^L$. For ligand conformer \mathcal{C} , the position of node v_i^l is $\mathbf{x}_i^l(\mathcal{C})$, and the length of edge e_{ij}^l is $\zeta(e_{ij}^l, \mathcal{C}) = \|\mathbf{x}_i^l(\mathcal{C}) - \mathbf{x}_j^l(\mathcal{C})\|$.

Pharmacophore model. The complete graph of the pharmacophore model $\mathcal{G}^P = (\mathcal{V}^P, \mathcal{E}^P)$ contains pharmacophores as nodes, and each node $v_i^p \in \mathcal{G}^P$ has a token score s_i^p , a center position \mathbf{x}_i^p , and a radius \mathbf{r}_i^p obtained from Equation 9, 13, and 14 in Appendix 5.1, with a pharmacophore type \mathbf{t}_i^p . For each edge $e_{ij}^p \in \mathcal{E}^P$, the mean and covariance of length are $\mu(e_{ij}^p) = \|\mathbf{x}_i^p - \mathbf{x}_j^p\|$ and $\sigma(e_{ij}^p) = \sqrt{(\mathbf{r}_i^p)^2 + (\mathbf{r}_j^p)^2}$, respectively.

Clustering. In the ligand molecules, one functional group often contains numerous functional groups. For example, 1 aromatic ring such as benzene contains 6 hydrophobic carbons. To improve the efficiency of the graph matching, we perform clustering for the ligand graph. The set of clusters is \mathbb{C}^L where each cluster $C_i^l \in \mathbb{C}^L$ is a set of v_i^l . In addition, we perform clustering for the pharmacophore model because one ligand functional group can form interactions with multiple protein functional groups, i.e. one v^l can be matched with multiple v^p . The set of clusters is \mathbb{C}^P where each cluster $C_i^p \in \mathbb{C}^P$ is a set of v_i^p . Graph matching is done on a per-cluster basis, not a per-node basis. The clustering algorithm is described in Appendix 5.3.2.

Graph-matching algorithm. We can represent a PGM \mathcal{M} as the set of pairs of ligand cluster and pharmacophore model cluster, $\{(C_{i_1}^l, C_{j_1}^p), (C_{i_2}^l, C_{j_2}^p), \dots\}$. In the context of inexact graph matching, the number of all PGMs is $(|\mathbb{C}^P| + 1)^{|\mathbb{C}^L|}$, and the computational requirements increase exponentially with the complexity of the ligand or pharmacophore model. Therefore, the efficient graph-matching algorithm is mandatory.

Indeed, most PGMs are physically infeasible. For example, it is not reasonable for two ligand functional groups that are only 1 Å to correspond to two pharmacophores that are 10 Å. To account for this, we use the distance constraint between cluster pairs as detailed in Appendix 5.3.3. Furthermore, an ensemble of ligand conformers shares a common core structure. Across the various conformers, the feasible PGM patterns for the core structure satisfying the distance constraints show substantial similarities. As a result, graph matching for numerous conformers can be performed simultaneously by identifying a PGM pattern for their core structure and then performing graph matching for the remaining part. To this end, we implemented the algorithm inspired by the depth-first search algorithm, details of which can be found at <https://github.com/SeonghwanSeo/PharmacoNet>.

2.3 Scoring function

Recently, Shen et al. [32] reported the state-of-the-art scoring model based on the pairwise atom distance likelihood between protein and ligand instead of scoring in the unit of energy. This has the advantage of allowing relative comparisons between ligands without the need for mapping to experimentally measured affinity values. Similarly, we propose a distance likelihood-based scoring function to score and rank the PGMs.

A Gaussian Mixture Model is used to express the probability density function of the distance between the ligand graph nodes. For the two cluster pairs $(C_{i_1}^l, C_{j_1}^p), (C_{i_2}^l, C_{j_2}^p) \in \mathcal{M}$, the probability density function of the distance d between $v_a^l \in C_{i_1}^l$ and $v_b^l \in C_{i_2}^l$ are follows:

$$\begin{aligned} I_a^{\mathcal{M}} &= \{s | s \in \{1, 2, \dots, |\mathcal{V}^P|\}, v_s^p \in C_s^p, \mathbf{t}_s^p \in \mathbf{T}_a^l\} \\ I_b^{\mathcal{M}} &= \{t | t \in \{1, 2, \dots, |\mathcal{V}^P|\}, v_t^p \in C_t^p, \mathbf{t}_t^p \in \mathbf{T}_b^l\} \\ P(d|\mathcal{M}, e_{ab}^l) &= F_{ab}^{\mathcal{M}}(d) = A \sum_{s \in I_a^{\mathcal{M}}} \sum_{t \in I_b^{\mathcal{M}}} \pi_{st} \mathcal{N}(d | \mu(e_{st}^p), \sigma(e_{st}^p)^2) \end{aligned} \quad (1)$$

where A is the normalizing constant and the weight of each component is the $\pi_{st} = \mathbb{W}[\mathbf{t}_s^p] \mathbb{W}[\mathbf{t}_t^p] \mathbf{s}_s^p \mathbf{s}_t^p$. The weights for each pharmacophore type $\mathbb{W}[t]$ are in Appendix 5.3.1. Note that the probability density function $F_{ab}^{\mathcal{M}}$ is independent of the ligand 3D conformer \mathcal{C} .

The score of PGM \mathcal{M} for ligand conformer \mathcal{C} can be represented as the weighted sum of pharmacophore pairwise distance likelihood:

$$\begin{aligned} s_{\text{edge}}(\mathcal{C}|\mathcal{M}, e_{ab}^l) &= \begin{cases} w_{ab}^{\mathcal{M}} F_{ab}^{\mathcal{M}}(\zeta(e_{ab}^l, \mathcal{C})) & \text{if } I_a^{\mathcal{M}} \neq \emptyset, I_b^{\mathcal{M}} \neq \emptyset \\ 0 & \text{else} \end{cases} \\ \text{Score}(\mathcal{C}, \mathcal{M}) &= \sum_{e_{ab}^l \in \mathcal{E}^L} s_{\text{edge}}(\mathcal{C}|\mathcal{M}, e_{ab}^l) \end{aligned} \quad (2)$$

where $w_{ab}^{\mathcal{M}} = \sum_{s \in I_a^{\mathcal{M}}} \mathbb{W}[\mathbf{t}_s^p] \mathbf{s}_s^p / |I_a^{\mathcal{M}}| \times \sum_{t \in I_b^{\mathcal{M}}} \mathbb{W}[\mathbf{t}_t^p] \mathbf{s}_t^p / |I_b^{\mathcal{M}}|$.

3 Experiments

3.1 Benchmark Study

Virtual screening benchmark. The widely-adopted virtual screening benchmarks are CASF-2016 [33], DUD-E [34], and DEKOIS2.0 [35]. However, CASF-2016 only contains 285 molecules and is not well-suited as a benchmark for the pre-screening task. In contrast, both the DUD-E and DEKOIS2.0 benchmarks are more similar to real-world scenarios. The DUD-E benchmark comprises 102 targets and offers several hundred actives, along with 50 decoys for each active, all sharing similar physicochemical properties. Likewise, DEKOIS2.0 contains 81 distinct targets, with 40 actives and a substantial 1200 decoys for each target. Consequently, our benchmark tests were conducted on both DUD-E and DEKOIS2.0 test sets.

We compared PharmacoNet with state-of-the-art conventional docking methods such as GLIDE (commercial) [7], AutoDock Vina [8], and SMINA [9]. In addition, the comparison with the non-structure-based deep learning model for pre-screening, DeepBindGCN [17], is in Appendix 5.5.2.

Runtime benchmark. We followed the benchmark proposed by Stärk et al. [10], which measures the average runtime on 363 complexes discovered since 2019 from PDBBind v2020. Along with the methods in the conventional docking programs, we also include structure-based deep learning models for binding pose prediction, GNINA [36], EquiBind [10], TANKBind [11], and DiffDock [12]. (Appendix 5.5.3)

Evaluation metric. The screening power is assessed with the top $\alpha\%$ enrichment factor, $\text{EF}_{\alpha\%}$, which is defined as the ratio of the percentages of the active ligands in the top $\alpha\%$ predictions to the percentages of the active ligands in the total library, and AUROC, the average area under the receiver operating characteristic. For runtime measurements, the processes that are performed before the virtual screening process are not considered, such as preprocessing of protein or generation of initial conformers for ligand.

Inference details. We generated ligand conformers with RDKit ETKDG [37, 38] from the SMILES and used the top graph matching score among the conformers for screening. As the default exhaustiveness of Autodock Vina and SMINA is 8, we compared the screening performance of our model when using a single conformer and 8 conformers.

Table 1: Results of the benchmark study. The top half contains conventional molecular docking methods, and the middle and bottom half contain our framework. The number after our model means the number of RDKit conformers used for scoring. We reused the number reported in Shen et al. [39] and Sunseri and Koes [40] for accuracy and Corso et al. [12] for runtime. We ran our model in a 16 CPU-cores environment with Intel Xeon Gold 6234. ‘*’ indicates that the method runs on a single CPU core. PharmacoNet-U means that the scoring function does not use weights for pharmacophore types. PharmacoNet-R means that our deep learning model is trained with a restricted training set.

Methods	DEKOIS2.0				DUD-E	PDBBind v2020 [10]
	EF _{0.5%}	EF _{1%}	EF _{5%}	AUROC	EF _{1%}	16-CPU Runtime (s)
GLIDE (comm.)	14.6	12.5	6.30	0.747	23.6	1405*
AutoDock Vina	5.46	4.51	2.82	0.633	9.93	205
SMINA	4.62	4.04	2.72	0.627	9.57	126
PharmacoNet (8)	3.75	3.42	2.41	0.624	5.73	0.0024
PharmacoNet (1)	3.43	3.23	2.55	0.625	5.36	0.0018
PharmacoNet-U (8)	1.57	1.73	2.04	0.603	4.19	0.0024
PharmacoNet-R (8)	3.53	3.07	2.54	0.622	4.78	0.0024

Result. Table 1 highlights that our model is significantly faster than existing energy-based docking methods and deep learning-based models for binding pose prediction. Our method outperforms the fastest conventional docking program, SMINA, by a factor of 50,000. To achieve this speed-up, our framework uses a coarse-grained graph-matching algorithm that predicts pharmacophore-level binding poses implicitly without considering atomic-level modeling, as illustrated in 2b. In addition, our scoring function has only 7 parameters, in contrast to conventional scoring functions that rely on numerous parameters. Despite these simplifications, our framework exhibits reasonable accuracy through pharmacophore-level abstraction.

Analysis: runtimes by the number of conformers. Furthermore, it is worth noting that the computational cost does not increase significantly when using multiple ligand conformers. In this experiment, we used either 1 or 8 conformers generated by ETKDG for each ligand. However, real-world virtual screening scenarios [1, 6] often involve multi-conformer databases containing tens to hundreds of conformers for each ligand, typically generated by fast conformer generation software such as OMEGA [41]. Since our graph-matching approach is similar to exhaustive rigid-body docking methods [42, 43], which uses numerous ligand conformers to account for the flexibility of the ligand structure, so using more conformers can improve accuracy. To assess the efficiency of our framework in handling large numbers of conformers, we measured the runtime according to the number of ETKDG conformers, as shown in Appendix 5.5.4. The average runtime for 100 conformers is 8.9 ms, while for a single conformer is 1.8 ms, and this result indicates the effectiveness of our framework in real-world scenarios.

Analysis: weights for pharmacophore types. Our scoring function relies solely on the weights assigned to each pharmacophore type as its parameters. In this study, these weights were determined based on prior knowledge of the relative contribution of each pharmacophore to protein-ligand binding as described in Appendix 5.3.1. To investigate the factors affecting the performance of our framework, we performed an ablation study (PharmacoNet-U) using the DEKOIS2.0 benchmark and the DUD-E benchmark used in previous experiments. Table 1 shows that the accuracy of our scoring function is improved by reflecting the chemistry of protein-ligand binding. Given that the current scoring function only uses 7 hand-designed parameters for pharmacophore types, higher performance can be reached by using additional parameters such as atom type with parameter fitting.

Analysis: generalization ability of deep learning model. In this work, we trained our model with PDBBind v2020 [30] dataset which contains 19,443 molecules. Since the limited number of structural data relative to the topological complexity of the protein-ligand complexes, A common challenge of deep learning approaches in drug discovery is the low generalization ability to proteins not encountered during training [44]. To address this, we simplify the topology of the protein-ligand complex through coarse-grained modeling at the pharmacophore level and incorporate distance

Table 2: Results of the pre-screening. ‘Filtration rate’ and ‘Number of retained’ indicate the filtration rates of pre-screening and the number of remaining molecules after pre-screening, respectively. ‘Total time’ is the time spent on the entire virtual screening including pre-screening (PharmacoNet) and fine-screening (SMINA). The last line represents the baseline scenario without any pre-screening.

Filtration rate	Number of retained	KRAS			Average docking score	Total time (hours) 16-CPU
		EF ₁₀₀	EF ₅₀₀	EF ₁₀₀₀		
90%	100,000	5.46	4.86	4.76	-8.35	150
95%	50,000	8.53	7.74	7.50	-8.43	75
98%	20,000	9.92	9.33	9.58	-8.50	31
99%	10,000	15.9	12.5	12.0	-8.55	15
-	992,235	-	-	-	-7.66	1300

constraints based on the maximum length information of NCIs. To improve generalization ability, we incorporated pharmacology as an inductive bias into the deep learning model and tailored the image instance segmentation architecture to pharmacophore modeling.

In the evaluation of the generalization capabilities of our model, we conducted experiments (PharmacoNet-R) using a restricted dataset consisting of 5000 training complexes and 500 validation complexes randomly selected from the original training and validation sets, representing 30% of the PDBBind v2020 dataset. Table 1 shows that the model trained on this restricted training set has a slightly lower overall performance. Nevertheless, it remains competitive in terms of AUROC and performs better than the baseline in EF_{5%}. This result demonstrates that the model can learn overall pharmacological features even with a limited training set, indicating that our approach achieves high generalization ability.

3.2 Pre-screening for Large-Scale Virtual Screening

To assess the effectiveness of our model’s pre-screening capabilities, we performed a large-scale virtual screening of 1 million ChEMBL molecules for KRAS (G12C), a challenging drug target. One of the key objectives in the pre-screening is to retain the hit candidates, the molecules with the best docking scores. Note that only the top 100 to 1000 molecules in the entire library are considered hit candidates for experimental validation. In evaluating the pre-screening performance of our framework, we used the top- N enrichment factor, denoted EF _{N} , which is defined as the ratio of the percentages of hit candidates in the retained molecules to the percentages of these candidates in the entire library, when the hit candidates are considered as the top N molecules with the highest docking score in the entire library. Furthermore, we measured the average docking score in conjunction with the enrichment factor. For the pre-screening stage, we used the top graph matching score among the 10 RDKit ETKDG conformers for each ligand, and for the fine-screening stage, we used SMINA with default settings.

The pre-screening process was remarkably efficient, taking only **11 minutes for 1 million molecules** on 16 CPU cores. The results in Table 2 show that our model can significantly reduce the total virtual screening time while effectively retaining hit candidate molecules. By filtering out 90% of the molecules from the library, our model retained more than half of the top 100 hit candidates (0.01% of the library), and 16 of those with a 99% filtration rate. Moreover, it’s worth noting that increasing the filtration rate of pre-screening improved the average docking scores of the remaining molecules. These observations suggest that there is a trend between our scores and energy-based docking scores, even though energy is not considered in model training.

Figure 3 shows the generated pharmacophore model for KRAS. Interestingly, in the SMINA docking pose of ChEMBL183526, which achieved the highest matching score, the chlorine atom (indicated by a red circle) aligns with the halogen atom pharmacophores (represented by yellow spheres) in our model, demonstrating the ability of our model to identify pharmacophore sites. Notably, the corresponding halogen bond was absent from both the X-ray (PDBID: 6OIM) and cryo-EM (PDBID: 8G47) structures of AMG-510, a known inhibitor of KRAS. This indicates the potential of our model to reveal unknown pharmacophores contributing to energy stabilization, which are not present in the complex structure of the active ligand.

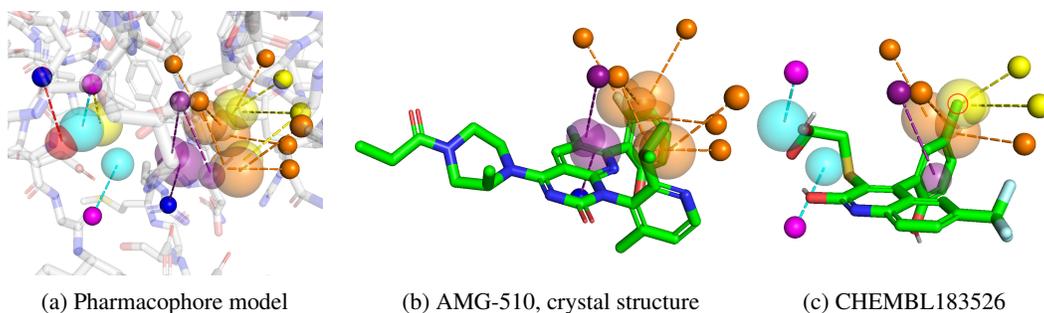


Figure 3: Visualizations of the pharmacophore model for KRAS. The color of pharmacophore and hot spot is as follows: orange for hydrophobic carbon, purple for aromatic ring, red for anion, blue for cation, cyan for H-bond donor, magenta for H-bond acceptor, and yellow for halogen atom and halogen bond acceptor. (a) The generated pharmacophore model of the binding site (PDB ID: 6OIM). (b) The crystal structure of the known inhibitor (AMG-510). (c) The SMINA docking pose of the ligand with the highest score of pre-screening.

3.3 Limitations

PharmacoNet relies on a scoring and graph-matching algorithm at the pharmacophore level, which has reasonably accurate and strong generalization capabilities without the need for a detailed parameter-fitting procedure. However, this approach has a trade-off; it does not capture atomic-level features. For example, it cannot distinguish between variations in the strength of the same salt bridge due to differences in charge and atom type, nor does it account for intramolecular energy.

To address these limitations, it may be useful to integrate force-field-based approaches such as UFF [45] and MMFF [46] and atomic-level details into the graph-matching and scoring algorithm, where a parameter-fitting process becomes essential. In particular, a promising avenue for improvement is the integration of a machine learning or deep learning-based approach into the graph-matching algorithm. This integration will significantly enhance our framework’s ability to account for fine-grained details and account for intramolecular energy factors, ultimately leading to improved accuracy and performance.

Furthermore, designing a balanced scoring model for both binding pose prediction and screening has long been a challenge in the molecular docking task [32, 47]. The success of the re-scoring models [32, 47, 44, 39] based on docking pose indicates that our framework has the potential to improve accuracy by integrating additional scoring functions tailored to each task, rather than relying on a single function for both.

4 Conclusion

In this work, we have demonstrated for the first time the ability and potential of a deep learning approach based on pharmacophore modeling. Our novel model, PharmacoNet, a 3D image instance segmentation model for generating pharmacophore models, opens up fresh avenues of the structure-based deep learning approach by coarse-grained modeling of proteins and ligands at the pharmacophore level. With a coarse-grained graph-matching algorithm and a simple distance likelihood-based scoring function, our framework achieves significant speed-up with reasonable accuracy and high generalization ability. Moreover, our framework has revealed its potential as a pre-screening method in the realm of large-scale virtual screening. We believe that the incorporation of pharmacophore modeling and pharmacophoric knowledge into deep learning-based methods paves the way for innovative breakthroughs in drug discovery.

Acknowledgments and Disclosure of Funding

This work was supported by Basic Science Research Programs through the National Research Foundation of Korea (NRF), grant funded by the Ministry of Science and ICT (NRF-2023R1A2C2004376, RS-2023-00257479).

5 Supplementary Material

5.1 Detailed Model Architecture

5.1.1 Atomic features

For each atom in the binding site, we use the residue type (20 common amino acids and unknown for others), atom type (C, N, O, S, and unknown for others), and functional group type (hydrophobic carbon, H-bond donor & acceptor, halogen bond acceptor, aromatic ring, cation, and anion) as an atomic feature. All water molecules and metal ions are omitted.

5.1.2 Voxelization of binding site

We represent the protein binding site as a 3D voxelized image. In this work, we used an image size of $64 \times 64 \times 64$ with a resolution of 0.5 \AA to represent the binding site, resulting in a box side length of 32 \AA . Note that the side length of the recommended maximum search box of AutoDock Vina is 30 \AA .

For the protein binding site \mathcal{B} , the coordinates and atomic features of the binding site atom are denoted $\mathbf{x}_i^b \in \mathbb{R}^3$ and $\mathbf{h}_i^b \in \mathbb{R}^C$, respectively. Then, the 3D input image $\mathbf{l} \in \mathbb{R}^{D \times H \times W \times C}$ is represented as follows:

$$\mathbf{l}_{d,h,w,:} = \sum_i^{N^b} K(\|\mathcal{T}^{1 \rightarrow \mathcal{B}}([d, h, w]) - \mathbf{x}_i^b\|) \times \mathbf{h}_i^b \quad (3)$$

$$K(x) = \begin{cases} e^{-2x^2} & \text{if } x \leq 1.5(\text{\AA}) \\ 0 & \text{else} \end{cases} \quad (4)$$

where K is a kernel function, N^b is the number of atoms in binding site, D, H, W are the spatial dimensions, and C is the number of atomic features. $\mathcal{T}^{A \rightarrow B}$ is the mapping between the voxel coordinates system and the real-world coordinates system, e.g. $\mathcal{T}^{1 \rightarrow \mathcal{B}}([d, h, w])$ is the real-world coordinates of a voxel with coordinates of $[d, h, w]$.

5.1.3 Protein hotspot prediction

To detect protein hotspots, PharmacoNet uses the Feature Pyramid Network [48] with a 3D extension of the Swin Transformer V2 Encoder [49], which is typically used in object detection tasks. The network obtains multi-scale 3D feature maps:

$$\mathbb{F} = \{\mathbb{F}^{(1)}, \mathbb{F}^{(1/2)}, \dots\} = \phi^{\text{backbone}}(\mathbf{l}) \quad (5)$$

where $\mathbb{F}^{(s)} \in \mathbb{R}^{sD \times sH \times sW \times C_s}$ is the feature map for each scale s and corresponding hidden dimension C_s .

Each protein functional group (token), denoted as t_i^t , contains its coordinates \mathbf{x}_i^t , and its NCI type (class) c_i^t . Since some functional groups can form multiple types of interactions, our deep learning model distinguishes them with their interaction type. The score of each token for determining whether it is a hotspot or not is then calculated as follows:

$$d_i, h_i, w_i = \mathcal{T}^{\mathcal{B} \rightarrow \mathbb{F}^{(1)}}(\mathbf{x}_i^t) \quad (6)$$

$$\mathbf{z}_i^t = \phi^{\text{token}}(\mathbb{F}_{d_i, h_i, w_i, :}^{(1)}, c_i^t) \quad (7)$$

$$y_i^t = \text{sigmoid}(\phi^{\text{score}}(\mathbf{z}_i^t)) \quad (8)$$

$$\mathbf{s}_i^t = P(X_{c_i^t}^{\text{score}} < y_i^t) \quad (9)$$

where the $X_{c_i^t}^{\text{score}}$ is the random variable of the sigmoid score (y_i^t) distribution of all tokens within the cavity in the validation set. As the number of protein hotspots constitutes only a small fraction of all tokens, we use the token score's relative position within the score distribution, \mathbf{s}_i^t , instead of directly using the sigmoid score in the scoring function for PharmacoNet. The score distribution and threshold for each NCI type are in Appendix 5.5.1.

5.1.4 Binary mask prediction for each cavity

We defined two types of cavities based on the length of the NCI. Each cavity is obtained as follows:

$$C^{\text{long}} = \text{sigmoid}(\phi_{\text{long}}^{\text{cavity}}(F^{(1)})) \quad (10)$$

$$C^{\text{short}} = \text{sigmoid}(\phi_{\text{short}}^{\text{cavity}}(F^{(1)})) \quad (11)$$

where $C^{\text{long}}, C^{\text{short}} \in \{0, 1\}^{D \times H \times W}$ represent binarized outputs with a threshold of 0.5 for the cavity regions corresponding to long-range and short-range interactions, respectively.

5.1.5 Binary mask prediction for each protein hotspot

In the instance segmentation manner, our model predicts the optimal positions of the pharmacophore for each protein hotspot:

$$D = \text{sigmoid}(\phi^{\text{mask}}(\mathbb{F}, \mathbf{x}_i^t, \mathbf{z}_i^t)) \odot C^{\text{short}} \quad (12)$$

where the $D \in [0, 1]^{D \times H \times W}$ is the density map, and the binary mask is obtained from D with the threshold of 0.5. We perform the Gaussian smoothing with the kernel size of 5 and the sigma of 0.5 to reduce the noise in the model output. Then, the voxels within 1 Å from protein atoms are masked, since NCIs are much longer than 1.0 Å. Moreover, the radius of the bounding box according to pharmacophore type is as follows:

- 4.5 Å for hydrophobic interaction, halogen bond, and H-bond
- 6.0 Å for π - π stacking and salt bridge
- 6.5 Å for cation- π interaction

Each radius is longer than the maximum distance in the default hyperparameters of Protein-Ligand Interaction Profiler (PLIP) [31].

5.1.6 Pharmacophore modeling from segmentation map

The pharmacophores for each protein hotspot can be represented as the segments in a corresponding binary mask. When there are multiple segments in a single binary mask, each segment is considered a distinct pharmacophore. Since the segment is the group of voxels, the center $\in \mathbb{R}^3$ and radius $\in \mathbb{R}$ of the corresponding pharmacophore are defined as follows:

$$\text{center} = \frac{1}{\sum_n^N D_{d_n, h_n, w_n}} \sum_n^N D_{d_n, h_n, w_n} \mathcal{T}^{D \rightarrow \mathcal{B}}([d_n, h_n, w_n]) \quad (13)$$

$$\text{radius} = \sqrt[3]{N/(4\pi/3)} \times \text{resolution} \quad (14)$$

where N is the number of containing voxels, d_n, h_n, w_n is the coordinates of each voxel in the segment, and D^i are the model output and binary mask corresponding to the pharmacophore.

5.2 Deep learning model training and inference

Dataset. We used the PDBbind v2020 dataset [30], a collection of high-resolution crystal structures, and measured binding affinities for 19,443 protein-ligand complexes deposited in the Protein Data Bank (PDB) [50]. Following Shen et al. [32], we partitioned the dataset into 17,658 training complexes and 1,500 validation complexes, excluding 285 CASF-2016 [33] test complexes. We omitted 8 complexes from the training set that contained multiple ligands. The center of mass of the active ligand was taken as the center of the binding site, and random translations and rotations were applied for data augmentation.

Objective function. From the complex-based pharmacophore model which is the ground truth pharmacophore model, we set the sigmoid scores for the protein hotspots to 1, otherwise 0. Each pharmacophore is voxelized into a 3D binary mask for its hotspot, where voxels within 1.0 Å are set to 1, otherwise 0. The cavities are also voxelized to 3D binary masks. The dimensions of the ground truth 3D binary masks are (D, H, W) .

For the instance segmentation modeling, we utilize two loss terms, which are pixel-wise binary mask loss and binary classification loss. Both loss terms employ the binary cross entropy loss function between ground truth and deep learning model prediction obtained from Equation 8, 10, 11 and 12. (Appendix 5.1)

Training details The prediction of binary masks for each pharmacophore is carried out using a 3D-CNN which is computationally expensive. As a result, we train the model by sampling up to 6 pharmacophores for each complex in each training iteration. Furthermore, the number of protein hotspots (instances) is considerably smaller compared to the total number of functional groups (tokens) present within the binding site. Consequently, we sample up to 200 tokens for each complex, with priority given to protein hotspots, tokens within cavities, and remaining tokens to reduce the data imbalance.

For the Swin Transformer V2-based encoder, we used the SwinV2-T configuration with the exception of the patch size of 2 and the window size of 4. The total number of parameters of our deep learning model is 31M. We use AdamW optimizer [51] for 200,000 iterations with a batch size of 8, the learning rate of 1×10^{-4} , weight decay of 0.05, and (β_1, β_2) of (0.9, 0.999). The model training takes about 48 hours on 4 NVIDIA RTX 3090 GPUs.

Inference details. The voxel outside the maximum distance of the NCI (based on the PLIP rule) from the corresponding hot spot is masked with 0. In addition, the voxels within 1.0 \AA of all protein atoms are masked. Finally, the model performed gaussian smoothing with a kernel size of 5 and a sigma of 0.5 to reduce the noise in the segmentation map.

5.3 Graph-Matching Algorithm.

All notations are defined in Section 2.2.

5.3.1 Parameters for scoring function

In this study, these weights were determined based on prior knowledge of the relative contribution of each pharmacophore to protein-ligand binding. Hydrogen and halogen bonds were considered comparable in terms of their characteristics and strengths, with both significantly outweighing the influence of hydrophobic interactions. Furthermore, salt bridges were considered stronger than both hydrogen and halogen bonds, while π - π stacking was recognized as a predominant driving force contributing to complex stability, particularly in terms of entropy. Based on this prior knowledge, we set the weight as follows:

$$\mathbb{W}[t] = \begin{cases} 1 & \text{if } t \text{ is Hydrophobic carbon} \\ 4 & \text{if } t \text{ in } \{\text{H-bond donor, H-bond acceptor, Halogen atom}\} \\ 8 & \text{if } t \text{ in } \{\text{Cation, Anion, Aromatic ring}\} \end{cases}$$

5.3.2 Clustering algorithm

Ligand functional groups. We perform clustering for ligand functional groups according to the following criteria:

- The cations, anions, H-bond acceptors, and H-bond donors in the same functional group are grouped.
- The aromatic ring and hydrophobic carbons in the same functional group are grouped.
- The hydrophobic carbons in the same carbon chain are grouped.

Pharmacophore model. Algorithm 1 describes the clustering algorithm for the pharmacophore model.

Algorithm 1 Pharmacophore Model Graph Clustering Algorithm

```

1: UsedNodes  $\leftarrow \{\}$ 
2:  $\mathbb{C}^P \leftarrow \{\}$ 
3: for  $v_i^p$  in  $\mathcal{G}^P$  do
4:   if  $v_i^p$  in UsedNodes then
5:     continue
6:   end if
7:    $C^p \leftarrow \{v_i^p\}$ 
8:   if  $t_i^p$  in {Cation, Anion, Aromatic} then
9:     for  $v_j^p \neq v_i^p$  in  $\mathcal{G}^P$  do
10:      if  $t_i^p = t_j^p$  and  $\|\mathbf{x}_i^p - \mathbf{x}_j^p\| < 1.5\text{\AA}$  then
11:        add  $\{v_j^p\}$  to  $C^p$ 
12:      else if  $\|\mathbf{x}_i^p - \mathbf{x}_j^p\| < 3.0\text{\AA}$  then
13:        if  $t_i^p$  in {Cation, Anion} and  $t_j^p$  in {H-bondDonor, H-bondAcceptor} then
14:          add  $v_j^p$  to  $C^p$ 
15:        else if  $t_i^p$  is Aromatic and  $t_j^p$  is HydrophobicCarbon then
16:          add  $v_j^p$  to  $C^p$ 
17:        end if
18:      end if
19:    end for
20:  end if
21:  add  $C^p$  to  $\mathbb{C}^P$ 
22:  UsedNodes  $\leftarrow$  UsedNodes  $\cup C^p$ 
23: end for
24: for  $v_i^p$  in  $\mathcal{G}^P$  do
25:   if  $v_i^p$  in UsedNodes then
26:     continue
27:   end if
28:    $C^p \leftarrow \{v_i^p\}$ 
29:   if  $t_i^p$  in {HydrophobicCarbon, HalogenAtom} then
30:     for  $v_j^p \neq v_i^p$  in  $\mathcal{G}^P$  do
31:      if  $t_i^p = t_j^p$  and  $\|\mathbf{x}_i^p - \mathbf{x}_j^p\| < 3.0\text{\AA}$  then
32:        add  $\{v_j^p\}$  to  $C^p$ 
33:      end if
34:    end for
35:   else if  $t_i^p$  in {H-bondDonor, H-bondAcceptor} then
36:     for  $v_j^p \neq v_i^p$  in  $\mathcal{G}^P$  do
37:      if  $t_j^p$  in {H-bondDonor, H-bondAcceptor} and  $\|\mathbf{x}_i^p - \mathbf{x}_j^p\| < 3.0\text{\AA}$  then
38:        add  $\{v_j^p\}$  to  $C^p$ 
39:      end if
40:    end for
41:   end if
42:   add  $C^p$  to  $\mathbb{C}^P$ 
43:   UsedNodes  $\leftarrow$  UsedNodes  $\cup C^p$ 
44: end for

```

5.3.3 Distance-based filtering algorithm

Algorithm 2 Distance-based Filtering Algorithm

```
1: Input:  $(C_{i_1}^l, C_{j_1}^p), (C_{i_2}^l, C_{j_2}^p), \mathcal{C}$ 
2:  $N_{\text{total}} = 0$ 
3:  $N_{\text{pass1}} = 0$ 
4: for  $v_{a_1}^l$  in  $C_{i_1}^l$  do
5:   for  $v_{a_2}^l$  in  $C_{i_2}^l$  do
6:      $N_{\text{pair}} = 0$ 
7:      $N_{\text{pass2}} = 0$ 
8:     for  $v_{b_1}^p$  in  $C_{j_1}^p$  do
9:       for  $v_{b_2}^p$  in  $C_{j_2}^p$  do
10:        if  $t_{b_1}^p \notin T_{a_1}^l$  or  $t_{b_2}^p \notin T_{a_2}^l$  then
11:          continue
12:        end if
13:         $N_{\text{pair}} \leftarrow N_{\text{pair}} + 1$ 
14:        if  $|\zeta(e_{a_1 a_2}^l, \mathcal{C}) - \mu(e_{b_1 b_2}^p)| \leq 2\sigma(e_{b_1 b_2}^p)$  then
15:           $N_{\text{pass2}} \leftarrow N_{\text{pass2}} + 1$ 
16:        end if
17:      end for
18:    end for
19:    if  $N_{\text{pair}} > 0$  then
20:       $N_{\text{total}} \leftarrow N_{\text{total}} + 1$ 
21:      if  $N_{\text{pass2}} \geq 0.5 \times N_{\text{pair}}$  then
22:         $N_{\text{pass1}} \leftarrow N_{\text{pass1}} + 1$ 
23:      end if
24:    end if
25:  end for
26: end for
27: if  $N_{\text{total}} > 0$  and  $N_{\text{pass1}} \geq 0.5 \times N_{\text{total}}$  then
28:   return True
29: else
30:   return False
31: end if
```

5.4 Software details

Conformer generation. We used the ETKDG [38] version 3 (small ring) implemented in RDKit [37].

Protein-ligand interaction profiler. To detect non-covalent bonds from the structure of a protein-ligand complex in PDBBind v2020, we used the Protein-Ligand Interaction Profiler (PLIP) [31] and Open Babel [52]. For flexible and fast data processing during the model training, we reimplemented it while keeping the original PLIP rules except for the water bridge and the metal bridge.

Molecular voxelization. In this study, our framework voxelizes the binding sites or pharmacophores using Gaussian and binary kernels. To accomplish this, we introduce a new voxelization tool, MolVoxel, designed to enable on-the-fly voxelization in various environments. MolVoxel is implemented in Python and has minimal dependencies, rendering it highly versatile for a wide range of deep learning and machine learning applications. Currently, it supports NumPy, Numba, and PyTorch (with CUDA support). The code is available in <https://github.com/SeonghwanSeo/molvoxel>.

5.5 Additional Result

5.5.1 Hot Spot Detection

Figure 4 shows the score distributions and thresholds for non-covalent interaction types. To determine hot spots, we considered the top 40% of scores for π - π stacking, cation- π interactions, and salt bridges, where the number of tokens is relatively small, and the top 20% of scores for hydrophobic interactions, H-bonds, and halogen bonds, where the number of tokens is larger.

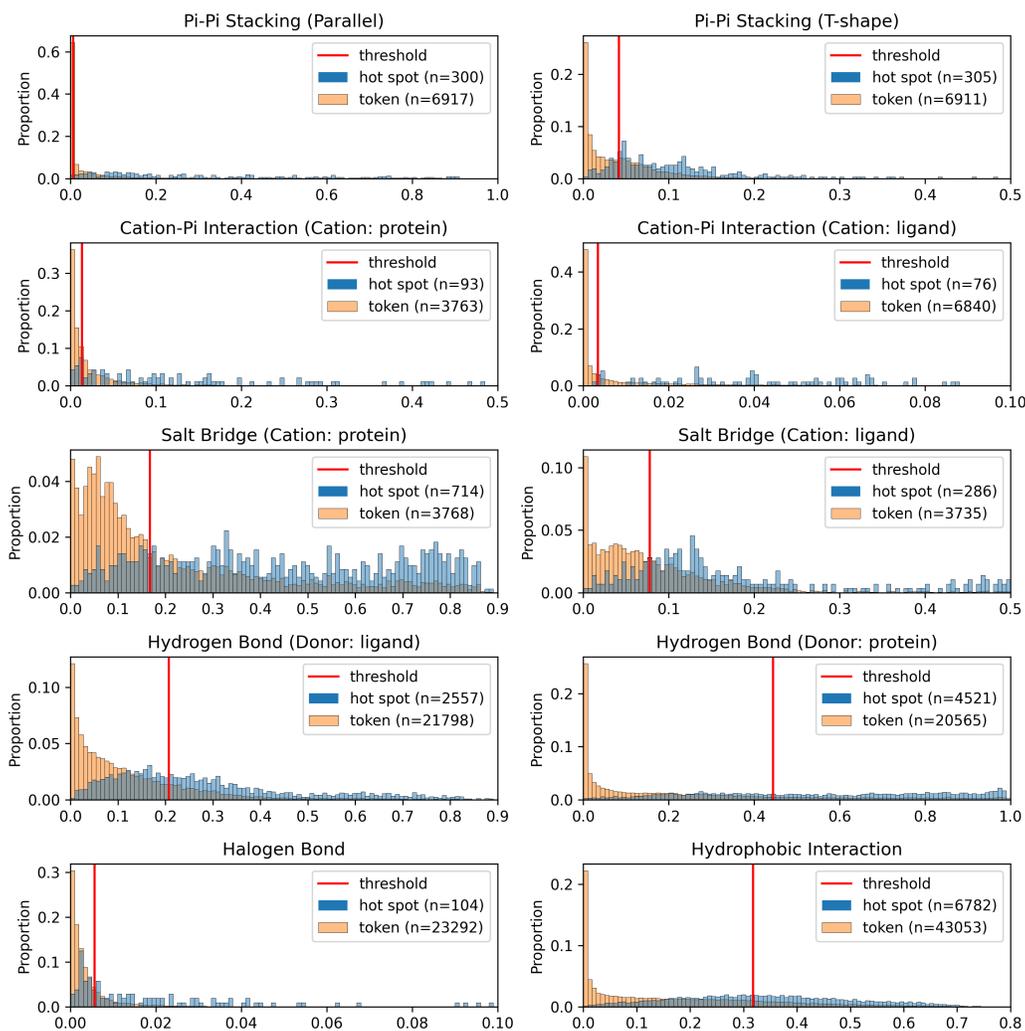


Figure 4: The sigmoid score distribution of tokens in the validation set. The blue histograms mean the score distribution of hot spots and the orange ones mean the score distribution of all tokens within cavities. The numbers in the parentheses mean the number of tokens. The red line means the threshold for hot spot prediction.

5.5.2 Comparison with non-structure-based deep-learning model for pre-screening

Zhang et al. [17] developed a deep learning model DeepBindGCN for pre-screening that works for multiple targets by integrating the 1D molecular vector of a pocket and a ligand. The model uses only the information of protein pocket and ligand 2D graph without the information of complex, so the prediction can be performed by combining the pre-calculated vector of ligands and proteins.

However, spatial and topological information is lost during global pooling to represent them as a 1D vector. Since the spatial coordination of protein and ligand is crucial in determining their interactions, the loss of spatial information leads to low performance and generalization ability for unseen target [44]. We compare the performance of our model with the DeepBindGCN-RG, the regression version of DeepBindGCN. Table 3 shows that DeepBindGCN-RG is faster than our model by GPU acceleration, but our model is faster in the CPU environment and shows better accuracy.

Table 3: Comparison with non-structure-based deep learning model for pre-screening, DeepBindGCN-RG. EF_{0.5%}, EF_{1%}, and EF_{5%} are average top 0.5%, 1.0%, 5.0% enrichment factor, respectively. PDBBind is the runtime benchmark where '16-CPU' and 'GPU' mean the average runtime for each device. The numbers in parenthesis after our model mean the number of RDKit conformers used.

Methods	DEKOIS2.0			DUD-E			PDBBind	
	EF _{0.5%}	EF _{1%}	EF _{5%}	EF _{0.5%}	EF _{1%}	EF _{5%}	16-CPU (s)	GPU (s)
DeepBindGCN-RG	0.89	0.80	0.82	2.31	2.11	1.62	0.003	0.0005
PharmacNet (8)	3.75	3.42	2.41	7.23	5.73	3.23	0.0024	-
PharmacNet (1)	3.43	3.23	2.55	6.30	5.36	3.41	0.0018	-

5.5.3 Comparison with structure-based deep learning models for binding pose prediction

Recently, several deep learning models have been developed to predict the protein-ligand complex conformation more accurately or quickly. To evaluate the efficiency of our coarse-grained graph-matching algorithm, we performed a comparison with the corresponding models. Table 4 shows that our method outperforms the fastest conventional docking programs by a factor of 10,000, and it is still 10 times faster than the fastest deep learning model designed for binding pose prediction.

Table 4: Average runtime per prediction of conventional docking programs and several deep learning-based approaches. The top five lines contain energy-based binding pose optimization approaches and the middle three lines contain data-driven approaches based on complex crystal structures. The last two lines show our models and the number after our model means the number of RDKit conformers used. '*' indicates that the method runs on a single CPU core. We reused the number from Stärk et al. [10], Lu et al. [11], and Corso et al. [12] for baseline approaches.

Methods	Type	Average Runtime (s)	
		16-CPU	GPU
GLIDE (comm.)	Energy-based binding pose prediction & scoring	1405*	-
AutoDock Vina		205	-
QVina-W		49	-
SMINA		146	-
GNINA		247	146
EquiBind	Data-driven binding pose prediction	0.16	0.04
TANKBind		0.54	0.28
DiffDock		-	10/40
PharmacNet (8)	Coarse-grained graph matching & scoring	0.0024	-
PharmacNet (1)		0.0018	-

5.5.4 Runtime according to the number of Ligand Conformers

We measured runtime using up to 100 conformers to understand how runtime increases with the number of conformers used in the evaluation. The evaluation is performed on the same runtime benchmark set proposed by Stärk et al. [10]. We attempted to generate ETKDG conformers up to 4,000 times for each ligand. We excluded 7 RDKit-unreadable molecules and 11 molecules for which fewer than 100 conformers were generated from the result. The reported time is the average of 5 runs. All measurements were made in a 16-CPU environment with the same machine, Intel Xeon Gold 6234.

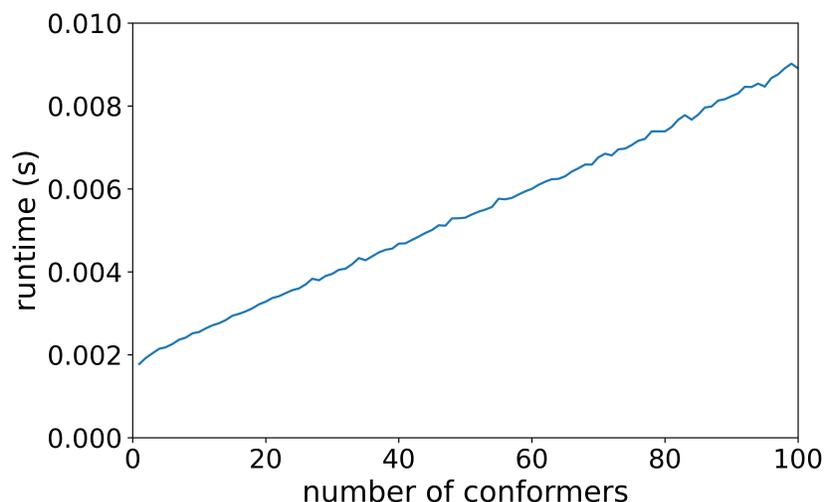


Figure 5: Average runtime according to the number of RDKit ETKDG conformations.

5.6 Related Work

Protein-ligand binding pose prediction. Traditional binding pose prediction method is based on physical energy-based scoring function and optimization algorithm [7–9]. Recent works improve the performance by incorporating deep learning-based scoring function [36, 53]. Another promising approach is to frame the pose prediction task as a regression problem [10, 11] or a generative modeling problem [12] with deep learning. EquiBind [10] and TANKBind [11] directly predict the atomic positions of the binding pose, so they are much faster than the traditional docking methods. However, their performance has not reached that of traditional docking, and they often cause steric crashes, leading to incorrect scores with physics-based and deep learning-based scoring functions.[12, 54] DiffDock [12] succeeds in generating high-quality conformations through diffusion generative modeling, but it is not fast enough to be used for pre-screening due to multiple denoising steps.

Rigid exhaustive docking. The flexibility of the molecules is one of the biggest challenges to predicting conformations. It makes optimization difficult by not only increasing the pose space but also creating numerous local optima. To avoid this problem, one of the main docking strategies is rigid-body docking with an ensemble of ligand conformers [7, 42, 43]. Rigid exhaustive docking is a practical docking strategy with sets of numerous core conformations of ligands [10].

Machine learning-based scoring function for virtual screening. Recently, machine learning-based and deep learning-based scoring functions have been developed to boost virtual screening. The scoring strategies are usually divided into non-structure-based [17, 55] and structure-based [32, 39, 44, 47, 56] whether they take into account the conformations of protein-ligand complexes. Non-structure-based scoring methods have the advantage of not requiring the use of complex structures. However, since the spatial relationship between protein and ligand is essential in determining their interaction, structure-based scoring models show better performance and generalization ability by

recognizing key interactions from the complex conformation [57]. Therefore, non-structure-based models are usually used in ligand-based virtual screening with known active ligands of the target protein [58, 59], and structure-based models are used in virtual screening strategies for general targets.

Model generalization to unseen protein. The generalization ability to unseen proteins is one of the biggest challenges for structure-based scoring models. Several studies have pointed out that deep learning models are often trained on the biases inherent in the data instead of the physical interaction between proteins and ligands [32, 60]. To improve the generalization ability, incorporating pharmacological properties as an inductive bias is a promising means in the various tasks of deep learning-based drug discovery [44, 47, 61, 62].

Instance image segmentation. Image segmentation is the task of partitioning an image into multiple segments. These segments may be delineated based on either semantic attributes (e.g. person, dog, building) [28] or instance-level criteria [29]. Semantic segmentation was tackled as a pixel classification problem, where each pixel is classified into one of the categories. Therefore, different objects can be grouped into one segment in semantic segmentation. In contrast, instance segmentation recognizes the segment and category for each object together. Instance segmentation is divided into a two-stage approach that detects objects and then determines the mask of each object [29], and a one-stage approach that performs segmentation on the entire image and then assigns a category to each segment [63, 64].

References

- [1] Jiankun Lyu, Sheng Wang, Trent E Balius, Isha Singh, Anat Levit, Yuri S Moroz, Matthew J O’Meara, Tao Che, Enkhjargal Algaa, Kateryna Tolmachova, et al. Ultra-large library docking for discovering new chemotypes. *Nature*, 566(7743):224–229, 2019.
- [2] Reed M Stein, Hye Jin Kang, John D McCorvy, Grant C Glatfelter, Anthony J Jones, Tao Che, Samuel Slocum, Xi-Ping Huang, Olena Savych, Yuri S Moroz, et al. Virtual discovery of melatonin receptor ligands to modulate circadian rhythms. *Nature*, 579(7800):609–614, 2020.
- [3] Christoph Gorgulla, Andras Boeszoermyeni, Zi-Fu Wang, Patrick D Fischer, Paul W Coote, Krishna M Padmanabha Das, Yehor S Malets, Dmytro S Radchenko, Yuri S Moroz, David A Scott, et al. An open-source drug discovery platform enables ultra-large virtual screens. *Nature*, 580(7805):663–668, 2020.
- [4] Andreas Lutten, Hjalmar Gullberg, Eldar Abdurakhmanov, Duy Duc Vo, Dario Akaberi, Vladimir O Talibov, Natalia Nekhotiaeva, Laura Vangeel, Steven De Jonghe, Dirk Jochmans, et al. Ultralarge virtual screening identifies sars-cov-2 main protease inhibitors with broad-spectrum activity against coronaviruses. *Journal of the American Chemical Society*, 144(7):2905–2920, 2022.
- [5] Francesco Gentile, Jean Charle Yaacoub, James Gleave, Michael Fernandez, Anh-Tien Ton, Fuqiang Ban, Abraham Stern, and Artem Cherkasov. Artificial intelligence-enabled virtual screening of ultra-large chemical libraries with deep docking. *Nature Protocols*, 17(3):672–697, 2022.
- [6] Arman A Sadybekov, Anastasiia V Sadybekov, Yongfeng Liu, Christos Iliopoulos-Tsoutsouvas, Xi-Ping Huang, Julie Pickett, Blake Houser, Nilkanth Patel, Ngan K Tran, Fei Tong, et al. Synthon-based ligand discovery in virtual libraries of over 11 billion compounds. *Nature*, 601(7893):452–459, 2022.
- [7] Richard A Friesner, Jay L Banks, Robert B Murphy, Thomas A Halgren, Jasna J Klicic, Daniel T Mainz, Matthew P Repasky, Eric H Knoll, Mee Shelley, Jason K Perry, et al. Glide: a new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy. *Journal of medicinal chemistry*, 47(7):1739–1749, 2004.
- [8] 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.
- [9] David Ryan Koes, Matthew P Baumgartner, and Carlos J Camacho. Lessons learned in empirical scoring with smina from the csar 2011 benchmarking exercise. *Journal of chemical information and modeling*, 53(8):1893–1904, 2013.
- [10] Hannes Stärk, Octavian Ganea, Lagnajit Pattanaik, Regina Barzilay, and Tommi Jaakkola. Equibind: Geometric deep learning for drug binding structure prediction. In *International conference on machine learning*, pages 20503–20521. PMLR, 2022.

- [11] Wei Lu, Qifeng Wu, Jixian Zhang, Jiahua Rao, Chengtao Li, and Shuangjia Zheng. Tankbind: Trigonometry-aware neural networks for drug-protein binding structure prediction. *Advances in neural information processing systems*, 35:7236–7249, 2022.
- [12] Gabriele Corso, Hannes Stärk, Bowen Jing, Regina Barzilay, and Tommi S. Jaakkola. Diffdock: Diffusion steps, twists, and turns for molecular docking. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=kKF8_K-mBbS.
- [13] Lei Wang, Shao-Hua Shi, Hui Li, Xiang-Xiang Zeng, Su-You Liu, Zhao-Qian Liu, Ya-Feng Deng, Ai-Ping Lu, Ting-Jun Hou, and Dong-Sheng Cao. Reducing false positive rate of docking-based virtual screening by active learning. *Briefings in Bioinformatics*, 24(1):bbac626, 2023.
- [14] Oleksandr O. Grygorenko, Dmytro S. Radchenko, Igor Dziuba, Alexander Chuprina, Kateryna E. Gubina, and Yurii S. Moroz. Generating multibillion chemical space of readily accessible screening compounds. *iScience*, 23(11):101681, 2020. ISSN 2589-0042. doi: <https://doi.org/10.1016/j.isci.2020.101681>. URL <https://www.sciencedirect.com/science/article/pii/S2589004220308737>.
- [15] David E Graff, Eugene I Shakhnovich, and Connor W Coley. Accelerating high-throughput virtual screening through molecular pool-based active learning. *Chemical science*, 12(22):7866–7881, 2021.
- [16] Lianxiang Luo, Ai Zhong, Qu Wang, and Tongyu Zheng. Structure-based pharmacophore modeling, virtual screening, molecular docking, admet, and molecular dynamics (md) simulation of potential inhibitors of pd-1 from the library of marine natural products. *Marine Drugs*, 20(1):29, 2021.
- [17] Haiping Zhang, Konda Mani Saravanan, and John ZH Zhang. Deepbindgcn: Integrating molecular vector representation with graph convolutional neural networks for protein–ligand interaction prediction. *Molecules*, 28(12):4691, 2023.
- [18] Camille-Georges Wermuth, CR Ganellin, Per Lindberg, and LA Mitscher. Glossary of terms used in medicinal chemistry (iupac recommendations 1998). *Pure and applied Chemistry*, 70(5):1129–1143, 1998.
- [19] David Schaller, Dora Šribar, Theresa Noonan, Lihua Deng, Trung Ngoc Nguyen, Szymon Pach, David Machalz, Marcel Bermudez, and Gerhard Wolber. Next generation 3d pharmacophore modeling. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 10(4):e1468, 2020.
- [20] Caterina Barillari, Gilles Marcou, and Didier Rognan. Hot-spots-guided receptor-based pharmacophores (hs-pharm): a knowledge-based approach to identify ligand-anchoring atoms in protein cavities and prioritize structure-based pharmacophores. *Journal of chemical information and modeling*, 48(7):1396–1410, 2008.
- [21] Gerhard Wolber and Thierry Langer. Ligandscout: 3-d pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *Journal of chemical information and modeling*, 45(1):160–169, 2005.
- [22] Jing Chen and Luhua Lai. Pocket v. 2: further developments on receptor-based pharmacophore modeling. *Journal of chemical information and modeling*, 46(6):2684–2691, 2006.
- [23] Paul D Kirchhoff, Rob Brown, Scott Kahn, Marvin Waldman, and CM Venkatachalam. Application of structure-based focusing to the estrogen receptor. *Journal of Computational Chemistry*, 22(10):993–1003, 2001.
- [24] Viet-Khoa Tran-Nguyen, Franck Da Silva, Guillaume Bret, and Didier Rognan. All in one: Cavity detection, druggability estimate, cavity-based pharmacophore perception, and virtual screening. *Journal of chemical information and modeling*, 59(1):573–585, 2018.
- [25] Siduo Jiang, Miklos Feher, Chris Williams, Brian Cole, and David E Shaw. Autoph4: An automated method for generating pharmacophore models from protein binding pockets. *Journal of Chemical Information and Modeling*, 60(9):4326–4338, 2020.
- [26] Sheng-Yong Yang. Pharmacophore modeling and applications in drug discovery: challenges and recent advances. *Drug discovery today*, 15(11-12):444–450, 2010.
- [27] Miha Skalic, Alejandro Varela-Rial, José Jiménez, Gerard Martínez-Rosell, and Gianni De Fabritiis. Ligvoxel: inpainting binding pockets using 3d-convolutional neural networks. *Bioinformatics*, 35(2):243–250, 2019.
- [28] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

- [29] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.
- [30] Zhihai Liu, Yan Li, Li Han, Jie Li, Jie Liu, Zhixiong Zhao, Wei Nie, Yuchen Liu, and Renxiao Wang. Pdb-wide collection of binding data: current status of the pdbind database. *Bioinformatics*, 31(3):405–412, 2015.
- [31] Sebastian Salentin, Sven Schreiber, V Joachim Haupt, Melissa F Adasme, and Michael Schroeder. Plip: fully automated protein–ligand interaction profiler. *Nucleic acids research*, 43(W1):W443–W447, 2015.
- [32] Chao Shen, Xujun Zhang, Chang-Yu Hsieh, Yafeng Deng, Dong Wang, Lei Xu, Jian Wu, Dan Li, Yu Kang, Tingjun Hou, et al. A generalized protein–ligand scoring framework with balanced scoring, docking, ranking and screening powers. *Chemical Science*, 14(30):8129–8146, 2023.
- [33] 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.
- [34] 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.
- [35] 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.
- [36] 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):1–20, 2021.
- [37] Greg Landrum et al. Rdkit: Open-source cheminformatics, 2006.
- [38] Sereina Riniker and Gregory A Landrum. Better informed distance geometry: using what we know to improve conformation generation. *Journal of chemical information and modeling*, 55(12):2562–2574, 2015.
- [39] Chao Shen, Xujun Zhang, Yafeng Deng, Junbo Gao, Dong Wang, Lei Xu, Peichen Pan, Tingjun Hou, and Yu Kang. Boosting protein–ligand binding pose prediction and virtual screening based on residue–atom distance likelihood potential and graph transformer. *Journal of Medicinal Chemistry*, 65(15):10691–10706, 2022.
- [40] Jocelyn Sunseri and David Ryan Koes. Virtual screening with gnina 1.0. *Molecules*, 26(23):7369, 2021.
- [41] Paul CD Hawkins, A Geoffrey Skillman, Gregory L Warren, Benjamin A Ellingson, and Matthew T Stahl. Conformer generation with omega: algorithm and validation using high quality structures from the protein databank and cambridge structural database. *Journal of chemical information and modeling*, 50(4): 572–584, 2010.
- [42] Guy PA Vigers and James P Rizzi. Multiple active site corrections for docking and virtual screening. *Journal of medicinal chemistry*, 47(1):80–89, 2004.
- [43] Georgia B McGaughey, Robert P Sheridan, Christopher I Bayly, J Chris Culberson, Constantine Kreatsoulas, Stacey Lindsley, Vladimir Maiorov, Jean-Francois Truchon, and Wendy D Cornell. Comparison of topological, shape, and docking methods in virtual screening. *Journal of chemical information and modeling*, 47(4):1504–1519, 2007.
- [44] Seokhyun Moon, Wonho Zhung, Soojung Yang, Jaechang Lim, and Woo Youn Kim. Pignet: a physics-informed deep learning model toward generalized drug–target interaction predictions. *Chemical Science*, 13(13):3661–3673, 2022.
- [45] Anthony K Rappé, Carla J Casewit, KS Colwell, William A Goddard III, and W Mason Skiff. Uff, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *Journal of the American chemical society*, 114(25):10024–10035, 1992.
- [46] Thomas A Halgren. Merck molecular force field. i. basis, form, scope, parameterization, and performance of mmff94. *Journal of computational chemistry*, 17(5-6):490–519, 1996.
- [47] Seokhyun Moon, Sang-Yeon Hwang, Jaechang Lim, and Woo Youn Kim. A versatile deep learning-based protein-ligand interaction prediction model for accurate binding affinity scoring and virtual screening. *arXiv preprint arXiv:2307.01066*, 2023.

- [48] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2117–2125, 2017.
- [49] Ze Liu, Han Hu, Yutong Lin, Zhuliang Yao, Zhenda Xie, Yixuan Wei, Jia Ning, Yue Cao, Zheng Zhang, Li Dong, et al. Swin transformer v2: Scaling up capacity and resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12009–12019, 2022.
- [50] Helen M Berman, John Westbrook, Zukang Feng, Gary Gilliland, Talapady N Bhat, Helge Weissig, Ilya N Shindyalov, and Philip E Bourne. The protein data bank. *Nucleic acids research*, 28(1):235–242, 2000.
- [51] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.
- [52] Noel M O’Boyle, Michael Banck, Craig A James, Chris Morley, Tim Vandermeersch, and Geoffrey R Hutchison. Open babel: An open chemical toolbox. *Journal of cheminformatics*, 3(1):1–14, 2011.
- [53] Oscar Méndez-Lucio, Mazen Ahmad, Ehecatl Antonio del Rio-Chanona, and Jörg Kurt Wegner. A geometric deep learning approach to predict binding conformations of bioactive molecules. *Nature Machine Intelligence*, 3(12):1033–1039, 2021.
- [54] Yuejiang Yu, Shuqi Lu, Zhifeng Gao, Hang Zheng, and Guolin Ke. Do deep learning models really outperform traditional approaches in molecular docking? *arXiv preprint arXiv:2302.07134*, 2023.
- [55] Francesco Gentile, Vibudh Agrawal, Michael Hsing, Anh-Tien Ton, Fuqiang Ban, Ulf Norinder, Martin E Gleave, and Artem Cherkasov. Deep docking: a deep learning platform for augmentation of structure based drug discovery. *ACS central science*, 6(6):939–949, 2020.
- [56] Jaechang Lim, Seongok Ryu, Kyubyong Park, Yo Joong Choe, Jiyeon Ham, and Woo Youn Kim. Predicting drug–target interaction using a novel graph neural network with 3d structure-embedded graph representation. *Journal of chemical information and modeling*, 59(9):3981–3988, 2019.
- [57] Shailesh Kumar Panday and Indira Ghosh. In silico structure-based prediction of receptor–ligand binding affinity: current progress and challenges. *Structural Bioinformatics: Applications in Preclinical Drug Discovery Process*, pages 109–175, 2019.
- [58] Meriem Bahi and Mohamed Batouche. Deep learning for ligand-based virtual screening in drug discovery. In *2018 3rd international conference on pattern analysis and intelligent systems (PAIS)*, pages 1–5. IEEE, 2018.
- [59] Tanuj Sharma, Venu Venkatarama Gowda Saralamma, Duk Chul Lee, Mohammad Azhar Imran, Jaehyuk Choi, Mohammad Hassan Baig, and Jae-June Dong. Combining structure-based pharmacophore modeling and machine learning for the identification of novel btk inhibitors. *International Journal of Biological Macromolecules*, 222:239–250, 2022.
- [60] Lieyang Chen, Anthony Cruz, Steven Ramsey, Callum J Dickson, Jose S Duca, Viktor Hornak, David R Koes, and Tom Kurtzman. Hidden bias in the dud-e dataset leads to misleading performance of deep learning in structure-based virtual screening. *PLoS one*, 14(8):e0220113, 2019.
- [61] Wonho Zhung, Hyeongwoo Kim, and Woo Youn Kim. A protein-ligand interaction-focused 3d molecular generative framework for generalizable structure-based drug design. *chemrxiv*, 2023.
- [62] Fergus Imrie, Thomas E Hadfield, Anthony R Bradley, and Charlotte M Deane. Deep generative design with 3d pharmacophoric constraints. *Chemical science*, 12(43):14577–14589, 2021.
- [63] Bowen Cheng, Ishan Misra, Alexander G Schwing, Alexander Kirillov, and Rohit Girdhar. Masked-attention mask transformer for universal image segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1290–1299, 2022.
- [64] Jitesh Jain, Jiachen Li, Mang Tik Chiu, Ali Hassani, Nikita Orlov, and Humphrey Shi. Oneformer: One transformer to rule universal image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2989–2998, 2023.