LIGAND-CONDITIONED BINDING SITE PREDICTION USING CONTRASTIVE GEOMETRIC LEARNING

Lisa Schneckenreiter & Sohvi Luukkonen
ELLIS Unit Linz and LIT AI Lab, Institute for Machine Learning, Johannes Kepler University Linz, Austria
{schneckenreiter,luukkonen}@ml.jku.at

Lukas Friedrich & Daniel Kuhn
Computational Chemistry & Biologics, Merck Healthcare
Darmstadt, Germany
{lukas.friedrich,daniel.kuhn}@merckgroup.com

Günter Klambauer
ELLIS Unit Linz and LIT AI Lab, Institute for Machine Learning, Johannes Kepler University
NXAI GmbH
Linz, Austria
klambauer@ml.jku.at

ABSTRACT

Understanding and modeling protein-ligand interactions is fundamental to modern drug discovery and design. Virtually every method employed in drug discovery — from experimental bioassays to computational techniques such as QSAR, docking, and activity prediction — relies on accurate models of these interactions. Recent advances in deep learning have greatly enhanced our ability to model protein-ligand interactions, as evidenced by innovations including graph neural networks for activity prediction, diffusion-based docking methods, geometric deep learning for binding pocket detection, contrastive learning for affinity prediction and virtual screening, and most recently foundational models for molecular structure prediction of biological complexes. In this work, we propose VN-EGNNrank, a novel ligand-conditioned binding site prediction method that combines a geometric architecture for protein encoding, a specialized ligand encoder, and a contrastive objective function to effectively align binding pocket and ligand representations in a shared latent space. Our experiments show that incorporating ligand information significantly enhances binding pocket ranking compared to ligand-agnostic models, and VN-EGNNrank achieves performance comparable to – or even exceeding – that of the much larger blind docking model DiffDock, while maintaining high computational efficiency suitable for large-scale virtual screening.

1 INTRODUCTION

The key component of drug discovery is the interaction between a protein and a potential ligand. Most drugs are small molecules that bind to a disease-associated protein target to activate, inhibit, or modify its function (Kinch et al., 2024). Understanding these protein-ligand interactions (PLIs) enables meaningful engagement with biological systems and the purposeful design of therapeutic agents (Gohlke et al., 2000; Du et al., 2016). For decades, computational methods, collectively known as computer-aided drug design (CADD) have been used to predict and analyze protein-ligand interactions. In particular, structure-based drug design (SBDD) relies on the availability of three-dimensional (3D) structural data for the target protein's binding site to model ligand binding and guide the rational optimization of potential therapeutics. Typically, this structural information is obtained through experimental determination of protein-ligand complexes (Mutharasappan et al., 2020), a challenging and time-consuming process that has historically restricted the

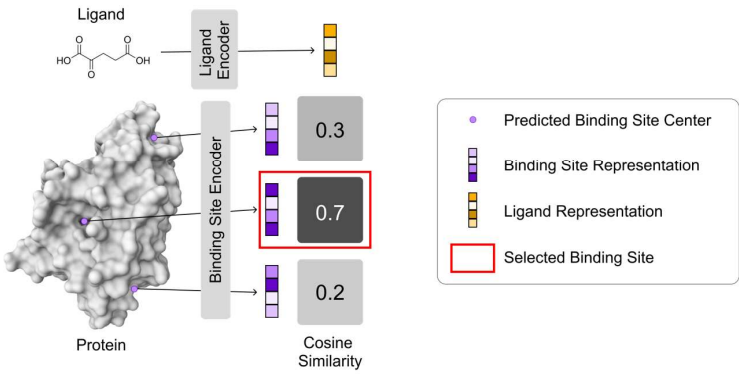

Figure 1: Overview of VN-EGNNrank. A protein encoder provides representations of potential binding sites. A ligand encoder gives the representation of the ligand. A high similarity of those representations indicates the correct binding site.

application of SBDD to only a subset of known proteins. Experimental structures are systematically archived in the Protein Data Bank (PDB) (Berman et al., 2000), which contains ∼230k entries[1], providing a valuable resource for SBDD research. Recently, the challenge of obtaining 3D protein structures has been largely addressed by AlphaFold (Jumper et al., 2021), which provides accurate predictions for most protein sequences. However, AlphaFold does not offer information about ligand binding sites, leaving binding site prediction as a crucial step in the SBDD pipeline (Zhao et al., 2020) before molecules can be screened with methods such as docking (Kuntz et al., 1982; Fan et al., 2019), molecular dynamics (De Vivo et al., 2016), and free energy perturbation (Beveridge & DiCapua, 1989; Cournia et al., 2021) to evaluate their binding potential and understand protein-ligand interactions.

Recent advances have improved binding pocket prediction, yet pinpointing therapeutically relevant sites remains challenging. Over the years, numerous computational methods have been developed to predict protein binding pockets, relying on four complementary approaches—geometric, evolutionary, energetic, and, more recently, machine learning and deep learning techniques (Macari et al., 2019; Xia et al., 2024). Notably, the recently introduced VN-EGNN – an SE(3)- and E(3)-Equivariant Graph Neural Network with Virtual Nodes – has demonstrated state-of-the-art performance on several binding site prediction benchmarks (Sestak et al., 2024). Many of these methods are ligand-agnostic, meaning they predict binding pockets based solely on protein features. However, proteins often possess multiple binding sites with distinct pharmacophoric profiles. Historically, drug discovery has focused on orthosteric binders, i.e. ligands binding to the same location as natural substrates, and most experimental crystal structures in the PDB represent these sites. As a result, machine learning–based pocket predictors trained on these datasets tend to be biased toward identifying orthosteric sites. In contrast, the growing interest in allosteric regulation and the development of allosteric or secondary binders as novel therapeutics has highlighted that allosteric pockets are generally more diverse and less conserved, making them more challenging to detect with ligand-agnostic approaches (Chatzigoulas & Cournia, 2021). Although many current methods can identify multiple binding pockets per protein, additional steps are often necessary to pinpoint the specific pocket relevant for a given therapeutic application.

Recent AI innovations enable holistic prediction of protein-ligand complexes. In recent years, advancements in AI have led to the development of blind docking methods—such as EquiBind (Stärk et al., 2022), DiffDock (Corso et al., 2023) and FABind (Pei et al., 2023)—which predict both the binding site and the ligand orientation without prior knowledge of the pocket's location. In effect, these methods serve as ligand-conditioned binding pocket predictors because they leverage both ligand-specific information and protein features. Moreover, recent foundation models for molecular structure prediction of biological complexes from string representations — such as AlphaFold3 (Abramson et al., 2024), Boltz-1 (Wohlwend et al., 2024), and Chai-1 (Discovery et al.,

[1]From https://www.rcsb.org/stats/growth/growth-released-structures. Accessed on 20/01/2025

2024) — have emerged, enabling the simultaneous prediction of the protein structure, the binding pocket and the binding pose. Although these models demonstrate state-of-the-art accuracy in predicting biological complexes when the binding site is unknown, they require substantial computational resources to screen compound libraries.

Contrastive learning has emerged as an efficient paradigm that drives state-of-the-art virtual screening in drug discovery. Another major innovation in AI in recent years has been contrastive learning – a paradigm exemplified by models such as CLIP (Radford et al., 2021). In drug discovery, contrastive learning has been successfully applied to virtual screening, where protein and ligand representations are jointly mapped into a shared latent space (Singh et al., 2023; Gao et al., 2024a;b; Han et al., 2024; Wang et al., 2024; McNutt et al., 2024) to predict PLIs. These approaches have demonstrated state-of-the-art performance compared to traditional docking-based methods. One advantage is that they are typically parameter efficient; since the protein and ligand encoders are often pre-trained and then frozen, the number of trainable parameters is greatly reduced, leading to shorter training times and fast inference. Although Gao et al. (2024a) utilized their contrastively pretrained pocket encoder for pocket druggability prediction and pocket matching, to our knowledge contrastive learning has not yet been applied to binding pocket prediction.

Our contribution. In this work, we propose VN-EGNNrank, a novel method for ligand-conditioned binding pocket prediction. Our approach leverages a geometric learning-based architecture to generate expressive representations of protein binding sites, employs a dedicated ligand encoder to capture chemical features, and utilizes a contrastive objective function to effectively align protein and ligand representations for accurate binding pocket prediction. Our experiments demonstrate that incorporating ligand information significantly enhances binding pocket ranking compared to ligand-agnostic models, and despite its lightweight design, VN-EGNNrank achieves performance comparable to—or even exceeding—that of the much larger DiffDock model.

2 VN-EGNNRANK: LIGAND-CONDITIONED BINDING POCKET PREDICTION

2.1 PRELIMINARIES.

Problem setting. Given a protein P and a ligand m, our objective is to predict the center of the binding site where this specific ligand binds. Unlike docking approaches, we do not aim to predict the exact ligand conformation within the binding site. However, what differentiates our approach from traditional binding site prediction methods is that our predictions are ligand-specific, making them dependent on the particular ligand under consideration.

Data. The training data of VN-EGNNrank consists of protein-ligand complexes from the PDB. For each ligand, the corresponding binding site is defined as all protein residues within 4Å of any ligand atom and represented by the geometric center of these residues.

Since we are using a geometric binding site encoder, we represent the protein as a nearest-neighbor graph. In this graph, nodes represent amino acid residues, and each node is connected to its 10 nearest neighbors. Nodes are assigned 3D coordinates corresponding to the positions of the residues' alpha carbons, and feature vectors are initialized with embeddings from ESM-2 (Lin et al., 2023).

Small molecule ligands are initially represented as concatenated vectors of Morgan fingerprints (Morgan, 1965) and chemical descriptors generated using RDKit (Landrum & contributors, 2006).

2.2 THE VN-EGNNRANK MODEL

Approach. In short, our method uses a *binding site encoder* based on VN-EGNN (Sestak et al., 2024) to predict multiple potential protein binding site locations $\mathbf{x}_1, \ldots, \mathbf{x}_K \in \mathbb{R}^3$ along with corresponding feature representations $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_K \in \mathbb{R}^d$. In contrast to pure binding site detection approaches (Krivák & Hoksza, 2018; Sestak et al., 2024), we also employ a *molecule encoder* that computes a representation $\boldsymbol{m} \in \mathbb{R}^d$ of the ligand in the same latent space. From the set of suggested pockets, the one with the representation most similar to $\boldsymbol{m}$ is selected as the predicted binding site. An illustration of this approach is shown in Figure 1.

Binding site prediction using VN-EGNN. The first step of our approach involves predicting binding site candidates, from which the most likely site is subsequently selected based on ligand information. To achieve this, we use VN-EGNN (Sestak et al., 2024), a state-of-the-art method for binding site prediction. In this approach, the protein is represented as a graph as defined above, with the addition of a small set of "virtual" nodes. These virtual nodes are connected to all protein nodes, allowing the network to capture global structural information.

VN-EGNN employs a heterogeneous message-passing scheme that integrates information from both protein and virtual nodes. The model is trained with a combination of three objective functions to directly predict binding pocket centers, represented by the updated virtual node positions $\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_M \in \mathbb{R}^3$, where M is the number of virtual nodes. In addition, the model also generates feature representations $\hat{\boldsymbol{p}}_1, \ldots, \hat{\boldsymbol{p}}_M \in \mathbb{R}^n$ for each pocket, which we make use of in the downstream ligand-conditioned ranking of binding sites. Formally, we use VN-EGNN as the protein encoder, providing the initial binding pocket representations for a given protein P:

$$(\hat{\boldsymbol{p}}_1, \ldots, \hat{\boldsymbol{p}}_M, \hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_M) = \mathrm{VNEGNN}(P).$$

For more details, see Sestak et al. (2024).

Pocket representations. The feature vectors $\hat{\boldsymbol{p}}_1, \ldots, \hat{\boldsymbol{p}}_M$ of the virtual nodes in VN-EGNN serve as the foundation of our binding pocket representations. Since multiple virtual nodes may converge to the same binding pocket, we use the Mean Shift Algorithm (Comaniciu & Meer, 2002) to cluster them based on their location in space. For each cluster, we compute the average of both the coordinates and feature vectors and apply an additional linear layer to map the feature representations to the contrastive space dimension d. As a result, we obtain K binding pocket candidates, each characterized by coordinates $\mathbf{x}_1, \ldots, \mathbf{x}_K \in \mathbb{R}^3$ and associated pocket representations $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_K \in \mathbb{R}^d$, with K representing the number of detected virtual node clusters.

Ligand representations. The ligand representation of an input molecule m is computed using a 3-layer fully connected network, which generates a feature vector $\boldsymbol{m} \in \mathbb{R}^d$. This simple architecture allows for the simultaneous encoding of many ligands, enabling efficient screening of large compound libraries.

Contrastive learning objective. To effectively align a ligand representation $\boldsymbol{m}$ with the representation of its corresponding binding site $\boldsymbol{p}_j$, we employ the InfoNCE contrastive loss (van den Oord et al.):

$$\mathcal{L}_{\text{InfoNCE}}\left(\boldsymbol{m}, \{\boldsymbol{p}_1, \ldots, \boldsymbol{p}_K\}\right) = -\log \frac{\exp\left(s(\boldsymbol{m}, \boldsymbol{p}_j)/\tau\right)}{\sum_{k=1}^{K} \exp\left(s(\boldsymbol{m}, \boldsymbol{p}_k)/\tau\right)}, \tag{1}$$

where τ is a trainable hyperparameter, and $s(.,.)$ is a similarity function, such as cosine similarity. This objective aims to maximize the similarity between matching molecule-pocket pairs while minimizing it for non-matching pairs. Specifically, we consider ligand $\boldsymbol{m}$ and binding site representation $\boldsymbol{p}_j$ a matching pair if the corresponding pocket position $\mathbf{x}_j$ is the closest to the ground-truth pocket of $\boldsymbol{m}$. All other pockets of the given protein are treated as non-matching. Alternative options for contrastive loss functions include the CLOOB loss (Fürst et al., 2022; Sanchez-Fernandez et al., 2023) and the SigLIP loss (Zhai et al., 2023; Seidl et al., 2023).

3 Experiments and Results

3.1 Training and Hyperparameters

Training data. We train our model on a more restrictive version of the same dataset used by our main baseline method, DiffDock, which uses a temporal split of PDBBind (Wang et al., 2005) as introduced by Stärk et al. (2022). Beyond ensuring no overlap of ligands between training and test sets, we also exclude any receptors present in the test set from the training set. After filtering, the training and validation sets in total comprise approximately 13,000 protein-ligand complexes from

before 2019, while the test split—in the following referred to as the PDBBind test set—includes around 200 PDB entries from 2019.

For data pre-processing, we filter out ions, molecules smaller than four heavy atoms, peptides longer than five residues, and common solvents to focus on biologically relevant small molecules. During training, we include only protein subunits within 4Å of each ligand, whereas for testing, the entire protein is used as input across all compared models.

Training procedure. For training VN-EGNNrank, we initialize the pocket encoder with a VN-EGNN model that has been pre-trained on the pocket prediction task, as described in Sestak et al. (2024). During the training of VN-EGNNrank, we optimize not only the InfoNCE loss but also the VN-EGNN-specific loss functions, with equal weighting for all loss terms. The model was trained with an initial learning rate of 1e-4, which was reduced by a factor of 10 upon reaching a plateau, using a learning rate scheduler. Pre-training VN-EGNN took 12 hours, while training VN-EGNNrank required 21 hours on a single A100 GPU.

Hyperparameters. The default hyperparameters of VN-EGNN were maintained for the binding site encoder, including 8 virtual nodes, 5 VN-EGNN layers, and a node feature dimension n of 100. The ligand encoder and the projection layer of the pocket representations both used SELU activation (Klambauer et al., 2017) and 10% dropout at the input. The ligand encoder consisted of 3 linear layers with a hidden dimension of 512 and 50% dropout for non-input layers. The output dimension of both the ligand and pocket encoders, corresponding to the contrastive latent space dimension d, was set to 256.

3.2 Evaluation

Evaluation data. We evaluated our models on subsets of three well-established datasets: HOLO4K and Coach420 originally detailed by Krivák & Hoksza (2018), and a temporal split test set of PDBBind (Stärk et al., 2022). The final evaluation datasets are considerably smaller than the original published datasets for several reasons: (1) we removed any overlap with the PDBBind training and validation sets, (2) we retained only those PDB structures that successfully passed our pre-processing pipeline, and (3) we included only protein-ligand pairs for which every method produced results.

Additionally, we evaluated our models on a new dataset of protein-ligand complexes derived from the Allosteric Site Database (ASD) (Liu et al. (2020); 06/2023 release). Starting from 3,102 entries of allosteric complexes with associated PDB IDs, we first removed entries whose PDB IDs appeared in our training data. We then excluded complexes where the biomolecule was not a protein or where the allosteric modulator was not a small molecule. Further filtering removed entries with irregular or missing UniProt IDs and those for which a SMILES representation for the modulator could not be retrieved, resulting in 1,802 protein–allosteric ligand pairs with PDB IDs. Finally, we augmented this dataset by retrieving all additional small molecules for each PDB (after filtering out common artifacts such as solvents and crystallization agents), yielding a total of 2,942 protein-ligand pairs across 1,720 PDB structures.

Table 1 summarizes each evaluation dataset's number of PDB structures and protein-ligand pairs.

	Total		Single Ligand		Multiple Ligands	
	#PDBs	#Pairs	#PDBs	#Pairs	#PDBs	#Pairs
HOLO4K	1218	1460	996	996	222	464
PDBBind test	208	252	178	178	30	74
COACH420	125	155	97	97	28	58
ASD	650	862	459	459	191	403

Table 1: Statistics of Evaluation Datasets.

Compared methods. We compare our approach with two ligand-agnostic binding pocket predictors—the well-established P2Rank (Krivák & Hoksza, 2018) and the state-of-the-art VN-EGNN (Sestak et al., 2024), upon which our model is built—and with the state-of-the-art blind docking

		P2Rank	VN-EGNN	DiffDock	VN-EGNNrank
HOLO4K	All	0.43 (0.41,0.46)	0.46 (0.43,0.48)	**0.60 (0.57,0.62)**	*0.51 (0.48,0.54)*
	Single Ligand	0.53 (0.49,0.56)	0.52 (0.49,0.56)	**0.63 (0.60,0.66)**	*0.58 (0.55,0.61)*
	Multi Ligand	0.23 (0.19,0.27)	0.31 (0.27,0.36)	**0.51 (0.47,0.56)**	*0.35 (0.31,0.40)*
COACH420	All	0.26 (0.19,0.32)	0.39 (0.32,0.46)	*0.43 (0.35,0.50)*	**0.45 (0.37,0.53)**
	Single Ligand	0.35 (0.26,0.44)	0.46 (0.37,0.56)	**0.52 (0.41,0.61)**	**0.52 (0.42,0.62)**
	Multi Ligand	0.10 (0.03,0.17)	0.26 (0.16,0.38)	*0.28 (0.17,0.40)*	**0.33 (0.22,0.45)**
PDBBind test	All	**0.59 (0.53,0.65)**	0.42 (0.36,0.48)	*0.46 (0.40,0.53)*	*0.46 (0.40,0.53)*
	Single Ligand	**0.76 (0.69,0.82)**	0.51 (0.43,0.58)	*0.57 (0.50,0.65)*	0.52 (0.44,0.60)
	Multi Ligand	0.19 (0.11,0.28)	0.20 (0.12,0.28)	*0.22 (0.12,0.31)*	**0.34 (0.23,0.45)**
ASD	All	0.29 (0.26,0.32)	0.38 (0.35,0.42)	*0.42 (0.39,0.45)*	**0.46 (0.42,0.49)**
	Single Ligand	0.33 (0.29,0.38)	0.40 (0.36,0.45)	*0.42 (0.37,0.46)*	**0.47 (0.42,0.51)**
	Multi Ligand	0.24 (0.20,0.29)	0.36 (0.32,0.41)	*0.43 (0.38,0.48)*	**0.44 (0.39,0.49)**

Table 2: Performance of binding site identification measured by the top-1 DCC success rate at a 4Å threshold. Values in parentheses indicate 95% confidence intervals obtained via bootstrapping with 10,000 iterations. In each row, the best-performing method is highlighted in bold and the second-best in italics.

method DiffDock (Corso et al., 2023), which performs implicit ligand-informed binding pocket detection. DiffDock, VN-EGNN, and our proposed VN-EGNNrank were all trained on a temporal split of PDBBind from Stärk et al. (2022), ensuring no overlap with our evaluation datasets. However, DiffDock was trained on a less restrictive version of this split (yielding a training set of ∼16k complexes) that filters out only complexes containing ligands from the test set, whereas VN-EGNN and VN-EGNNrank were trained on a more rigorously curated subset that also excludes complexes with receptors present in the test set (yielding a training set of ∼14k complexes). In contrast, P2Rank was trained on a different dataset (CHEN11), which does not overlap with our evaluation sets in terms of PDB IDs.

Metrics. We evaluated model performance using the DCC criterion—the distance from the predicted pocket center to the ground-truth pocket center—with a 4Å threshold, considering only the top-ranked pocket (Top-1 DCC @ 4Å) from each method. For every model and dataset, we compute success rates for all protein-ligand pairs and separately for PDB structures containing a single ligand and those with multiple ligands. All success rate values are reported with 95% confidence intervals obtained through bootstrapping.

3.3 Results

The main results comparing model performances on the different dataset are reported in Table 2.

Incorporating ligand information improved pocket identification. Across all datasets and subsets, VN-EGNNrank consistently outperforms VN-EGNN when considering the top-ranked pocket, demonstrating that re-ranking binding pockets using the similarity in the joint pocket–ligand latent space enhances prediction accuracy. Moreover, aside from the exceptionally high performance of P2Rank on single-ligand protein targets in PDBBind – which is reflected in overall results given that approximately 70% of protein-ligand pairs come from single-ligand proteins – the ligand-informed approaches (VN-EGNNrank and DiffDock) systematically outperform the ligand-agnostic methods (P2Rank and VN-EGNN). We suspect that the relatively higher performance of P2Rank on the PDBBind test, compared to the three other models, is due to a difference in the training data: VN-EGNN, DiffDock, and VN-EGNNrank were trained on a PDBBind training set created with a temporal split that enforces dissimilarity to the test set by filtering out complexes with ligands present in the test data, whereas P2Rank was trained on a separate dataset that, although it doesn't contain complexes after 2019, does not enforce a similar dissimilarity constraint relative to the PDBBind test set.

VN-EGNNrank on par with much larger DiffDock. Our lightweight VN-EGNNrank (approximately 3M parameters) has comparable performance with the much larger DiffDock model (around

25M parameters) on three of the four evaluated datasets (Coach420, PDBBind, and ASD). However, DiffDock outperforms VN-EGNN-based methods on the HOLO4K dataset. We suspect that this discrepancy arises from differences in training data and processing: DiffDock was trained on full PDB structures using a slightly larger and less restrictive training set, while VN-EGNN and VN-EGNNrank were trained using only the chains within 4Å of the ligand of interest. Since HOLO4K contains many symmetric units of large complexes, this leads to a greater domain shift for the VN-EGNN-based methods compared to DiffDock.

Efficient runtime performance. Our experiments demonstrate that VN-EGNNrank dramatically reduces inference time compared to DiffDock. According to Corso et al. (2023), DiffDock requires approximately 10 seconds per prediction for 10 generated binding poses on an A100 GPU, with runtime scaling linearly with the number of ligands tested. In contrast, the ligand encoder in VN-EGNNrank incurs virtually no additional runtime overhead (see Figure 2, also computed on an A100 GPU), making VN-EGNNrank exceptionally well-suited for efficient screening of large molecular libraries.

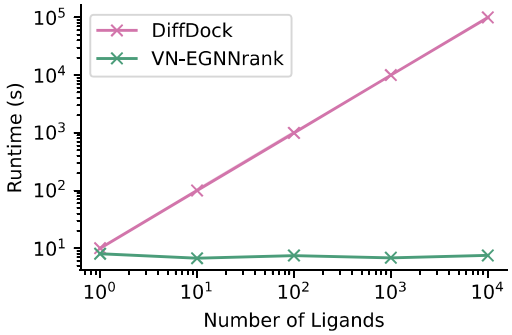

Figure 2: Scaling of runtime with the number of ligands per protein for VN-EGNNrank and DiffDock. DiffDock runtimes are extrapolated from the values reported in Corso et al. (2023).

Pocket prediction models are biased toward orthosteric binding sites. In three well-established datasets, HOLO4K, the PDBBind test set, and Coach420, we observe a significant drop in performance when moving from single-ligand to multi-ligand complexes for all models. We hypothesize that single-ligand cases in these datasets are predominantly orthosteric binders, which are well represented in the training data (e.g., the temporal split of PDBBind), resulting in higher performance. In contrast, multi-ligand complexes, by definition, include a greater proportion of allosteric binders, for which the models perform markedly worse. Conversely, in the ASD dataset – where single-ligand complexes are specifically allosteric – the overall performance is lower, and no comparable drop is observed when transitioning to multi-ligand complexes, which likely contain some orthosteric binders. These findings reinforce the conclusion that current pocket prediction models are inherently tuned to and biased towards orthosteric sites.

4 Discussion

Conclusion. In summary, our results demonstrate that incorporating ligand information through VN-EGNNrank improves binding pocket identification compared to ligand-agnostic approaches, achieving performance comparable to – or even surpassing – that of DiffDock, a much larger model. Despite its lightweight architecture, VN-EGNNrank consistently delivers robust performance across diverse datasets, highlighting the promise of ligand-conditioned methods.

Limitations. Nevertheless, our work has several limitations. First, our approach remains a work in progress, and further comparisons other methods such as with state-of-the-art foundational models for biomolecular structure prediction are warranted. Second, our current evaluation only considers a single binding pocket per ligand, even though some ligands can bind to multiple sites simultaneously.

Third, our approach focuses solely on identifying binding pockets in a ligand-informed manner, without predicting the ligand conformation within the pocket.

Outlooks. Our ongoing work aims to extend the framework to handle multiple binding pockets per ligand, thereby enhancing its utility for virtual screening. Additionally, we plan to improve model training by developing more balanced training and evaluation sets that include both orthosteric and allosteric binders.

MEANINGFULNESS STATEMENT

Proteins and small molecules are two crucial types of molecules encountered in Life Sciences, therefore we consider accurate representations of proteins, their binding pockets, and ligands highly meaningful. Our work, VN-EGNNrank, enhances this by using geometric learning to represent binding pockets and a ligand encoder to capture ligand-specific features. Through contrastive learning, we align protein and ligand representations in a shared space to enable ligand-conditioned binding pocket predictions, offering a more comprehensive view of protein-ligand interactions and supporting more efficient drug discovery.

REFERENCES

Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J Ballard, Joshua Bambrick, et al. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, pp. 1–3, 2024.

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.

D L Beveridge and F M DiCapua. Free energy via molecular simulation: Applications to chemical and biomolecular systems. *Annual Review of Biophysics*, 18(Volume 18, 1989):431–492, 1989. ISSN 1936-1238.

Alexios Chatzigoulas and Zoe Cournia. Rational design of allosteric modulators: Challenges and successes. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 11(6):e1529, 2021.

Dorin Comaniciu and Peter Meer. Mean shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002.

Gabriele Corso, Hannes Stärk, Bowen Jing, Regina Barzilay, and Tommi Jaakkola. Diffdock: Diffusion steps, twists, and turns for molecular docking. In *International Conference on Learning Representations (ICLR)*, 2023.

Zoe Cournia, Christophe Chipot, Benoît Roux, Darrin M York, and Woody Sherman. Free energy methods in drug discovery—introduction. In *Free Energy Methods in Drug Discovery: Current State and Future Directions*, pp. 1–38. ACS Publications, 2021.

Marco De Vivo, Matteo Masetti, Giovanni Bottegoni, and Andrea Cavalli. Role of molecular dynamics and related methods in drug discovery. *Journal of Medicinal Chemistry*, 59(9):4035–4061, 2016.

Chai Discovery, Jacques Boitreaud, Jack Dent, Matthew McPartlon, Joshua Meier, Vinicius Reis, Alex Rogozhnikov, and Kevin Wu. Chai-1: Decoding the molecular interactions of life. *bioRxiv*, pp. 2024–10, 2024.

Xing Du, Yi Li, Yuan-Ling Xia, Shi-Meng Ai, Jing Liang, Peng Sang, Xing-Lai Ji, and Shu-Qun Liu. Insights into protein–ligand interactions: mechanisms, models, and methods. *International Journal of Molecular Sciences*, 17(2):144, 2016.

Jiyu Fan, Ailing Fu, and Le Zhang. Progress in molecular docking. *Quantitative Biology*, 7:83–89, 2019.

Andreas Fürst, Elisabeth Rumetshofer, Johannes Lehner, Viet T Tran, Fei Tang, Hubert Ramsauer, David Kreil, Michael Kopp, Günter Klambauer, Angela Bitto, et al. Cloob: Modern hopfield networks with infoloob outperform clip. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

Bowen Gao, Yinjun Jia, Yuanle Mo, Yuyan Ni, Weiying Ma, Zhiming Ma, and Yanyan Lan. Profsa: Self-supervised pocket pretraining via protein fragment-surroundings alignment. 2024a.

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

Holger Gohlke, Manfred Hendlich, and Gerhard Klebe. Knowledge-based scoring function to predict protein-ligand interactions. *Journal of Molecular Biology*, 295(2):337–356, 2000.

Jin Han, Yun Hong, and Wu-Jun Li. Hashing based contrastive learning for virtual screening, 2024.

John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, 2021.

Michael S Kinch, Zachary Kraft, and Tyler Schwartz. 2023 in review: Fda approvals of new medicines. *Drug Discovery Today*, pp. 103966, 2024.

Günter Klambauer, Thomas Unterthiner, Andreas Mayr, and Sepp Hochreiter. Self-normalizing neural networks. *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

Radoslav Krivák and David Hoksza. P2Rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure. *Journal of Cheminformatics*, 10(1):1–12, 2018.

Irwin D Kuntz, Jeffrey M Blaney, Stuart J Oatley, Robert Langridge, and Thomas E Ferrin. A geometric approach to macromolecule-ligand interactions. *Journal of Molecular Biology*, 161(2): 269–288, 1982.

Greg Landrum and RDKit contributors. Rdkit: Open-source cheminformatics software, 2006. URL https://www.rdkit.org.

Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023.

Xinyi Liu, Shaoyong Lu, Kun Song, Qiancheng Shen, Duan Ni, Qian Li, Xinheng He, Hao Zhang, Qi Wang, Yingyi Chen, et al. Unraveling allosteric landscapes of allosterome with asd. *Nucleic Acids Research*, 48(D1):D394–D401, 2020.

Gabriele Macari, Daniele Toti, and Fabio Polticelli. Computational methods and tools for binding site recognition between proteins and small molecules: from classical geometrical approaches to modern machine learning strategies. *Journal of Computer-aided Molecular Design*, 33(10): 887–903, 2019.

Andrew T McNutt, Abhinav K Adduri, Caleb N Ellington, Monica T Dayao, Eric P Xing, Hosein Mohimani, and David R Koes. Sprint enables interpretable and ultra-fast virtual screening against thousands of proteomes. *arXiv preprint arXiv:2411.15418*, 2024.

H. L. Morgan. The generation of a unique machine description for chemical structures—a technique developed at chemical abstracts service. *Journal of Chemical Documentation*, 5(2):107–113, 1965. doi: 10.1021/c160017a018.

Nachiappan Mutharasappan, Guru Ravi Rao, Richard Mariadasse, Saritha Poopandi, Amala Mathimaran, Prabhu Dhamodharan, Rajamanikandan Sundarraj, Chitra Jeyaraj Pandian, and Jeyakanthan Jeyaraman. Experimental and computational methods to determine protein structure and stability. *Frontiers in Protein Structure, Function, and Dynamics*, pp. 23–55, 2020.

Qizhi Pei, Kaiyuan Gao, Lijun Wu, Jinhua Zhu, Yingce Xia, Shufang Xie, Tao Qin, Kun He, Tie-Yan Liu, and Rui Yan. FABind: Fast and Accurate Protein-Ligand Binding. In *Advances in Neural Information Processing Systems*, 2023.

A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139, pp. 8748–8763, 2021.

Ana Sanchez-Fernandez, Elisabeth Rumetshofer, Sepp Hochreiter, and Günter Klambauer. Cloome: contrastive learning unlocks bioimaging databases for queries with chemical structures. *Nature Communications 2023 14:1*, 14:1–14, 11 2023.

Philipp Seidl, Andreu Vall, Sepp Hochreiter, and Günter Klambauer. Enhancing activity prediction models in drug discovery with the ability to understand human language. *Proceedings of Machine Learning Research*, 202:30458–30490, 3 2023.

Florian Sestak, Lisa Schneckenreiter, Johannes Brandstetter, Sepp Hochreiter, Andreas Mayr, and Günter Klambauer. Vn-egnn: E(3)-equivariant graph neural networks with virtual nodes enhance protein binding site identification, 2024.

Rohit Singh, Samuel Sledzieski, Bryan Bryson, Lenore Cowen, and Bonnie Berger. Contrastive learning in protein language space predicts interactions between drugs and protein targets. *Proceedings of the National Academy of Sciences*, 120(24):e2220778120, 2023.

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*, 2022.

Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*.

Renxiao Wang, Xueliang Fang, Yipin Lu, Chao-Yie Yang, and Shaomeng Wang. The pdbbind database: methodologies and updates. *Journal of Medicinal Chemistry*, 48(12):4111–4119, 2005.

Zhen Wang, Zhanfeng Wang, Maohua Yang, Long Pang, Fangyuan Nie, Siyuan Liu, Zhifeng Gao, Guojiang Zhao, Xiaohong Ji, Dandan Huang, et al. Enhancing challenging target screening via multimodal protein-ligand contrastive learning. *bioRxiv*, 2024.

Jeremy Wohlwend, Gabriele Corso, Saro Passaro, Mateo Reveiz, Ken Leidal, Wojtek Swiderski, Tally Portnoi, Itamar Chinn, Jacob Silterra, Tommi Jaakkola, et al. Boltz-1: Democratizing biomolecular interaction modeling. *bioRxiv*, 2024.

Ying Xia, Xiaoyong Pan, and Hong-Bin Shen. A comprehensive survey on protein-ligand binding site prediction. *Current Opinion in Structural Biology*, 86:102793, 2024.

Xiaohua Zhai, Basil Mustafa, Alexander Kolesnikov, and Lucas Beyer. Sigmoid loss for language image pre-training. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 11975–11986, 2023.

Jingtian Zhao, Yang Cao, and Le Zhang. Exploring the computational methods for protein-ligand binding site prediction. *Computational and Structural Biotechnology Journal*, 18:417–426, 2020.