

Patterns

Hierarchical affinity landscape navigation through learning a shared pocket-ligand space

Highlights

- LigUnity is a unified foundation model for virtual screening and hit-to-lead optimization
- Outperforms 24 methods with >50% improvement in virtual screening
- Approaches FEP+ accuracy at far lower cost in hit-to-lead optimization
- Achieves 10^6 speedup compared to traditional docking methods like Glide-SP

Authors

Bin Feng, Zijing Liu, Hao Li, ..., Yu Li, Lei Zhang, Sheng Wang

Correspondence

liuzijing@idea.edu.cn (Z.L.),
liyu@idea.edu.cn (Y.L.),
swang@cs.washington.edu (S.W.)

In brief

LigUnity is a protein-ligand affinity foundation model, enabling both rapid virtual screening and precise hit-to-lead optimization. LigUnity learns a shared “language” for protein pockets and ligands, outperforming docking in screening and approaching FEP+ accuracy at far lower cost. It generalizes to new targets and new chemical scaffolds and achieves 10^6 times the speedup compared to Glide-SP, offering a practical foundation for early-stage drug discovery.

Article

Hierarchical affinity landscape navigation through learning a shared pocket-ligand space

Bin Feng,¹ Zijiang Liu,^{1,*} Hao Li,¹ Mingjun Yang,² Junjie Zou,² He Cao,¹ Yu Li,^{1,*} Lei Zhang,¹ and Sheng Wang^{3,4,*}

¹International Digital Economy Academy (IDEA), Shenzhen, China

²XtalPi Co., Ltd., Shenzhen, China

³Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle, WA, USA

⁴Lead contact

*Correspondence: liuzijiang@idea.edu.cn (Z.L.), liyu@idea.edu.cn (Y.L.), swang@cs.washington.edu (S.W.)

<https://doi.org/10.1016/j.patter.2025.101371>

THE BIGGER PICTURE Despite thousands of proteins in the human genome, only about 10% are currently bound to small molecules with high potency, highlighting the urgent need to discover active molecules for new therapeutic targets. As experimental affinity data accumulate rapidly, unprecedented resources are available for machine learning methods. However, most existing methods developed for identifying active molecules treat the valuable data in a coarse, binary fashion, simply classifying compounds as active or inactive for a given target. These approaches overlook the important information contained in these data—not just whether a molecule works but also how well it works and what subtle chemical differences make one compound more effective than another.

LigUnity is a foundation model designed for addressing this challenge. It learns a shared “language” of protein pockets and small molecules that captures both whether a molecule binds and how efficiently it does so, with improved efficiency and accuracy. LigUnity makes efficient exploration of vast chemical spaces possible and accelerates the search for new medicines.

SUMMARY

The structure of the protein binding pocket governs the ligand binding affinity by providing crucial intermolecular interactions and spatial compatibility. While existing methods have leveraged these structural insights to advance affinity prediction, they often treat virtual screening and hit-to-lead optimization separately, mainly due to incompatible speed-accuracy requirements. However, these two tasks complement each other, and their integration enables broader chemical exploration while preserving focus on affinity-determining substructures. Here, we present ligand unified affinity (LigUnity), a foundation model for affinity prediction that jointly embeds ligands and pockets into a shared space. In particular, LigUnity learns coarse-grained active/inactive distinction through scaffold discrimination and fine-grained pocket-specific ligand preference through pharmacophore ranking. We demonstrate the effectiveness and versatility of LigUnity on eight benchmarks across six settings. In virtual screening, LigUnity outperforms 24 methods with >50% improvement and demonstrates robust generalization to novel targets. In hit-to-lead optimization, it achieves state-of-the-art performance across split-by-time, split-by-scaffold, and split-by-unit settings, emerging as a cost-efficient alternative to free energy perturbation. We further showcase how LigUnity can be employed in an active learning framework for tyrosine kinase 2 (TYK2) to efficiently find optimal ligands. Collectively, these results establish LigUnity as a versatile foundation model for affinity prediction, offering broad applicability across the drug discovery pipeline.

INTRODUCTION

The structure of the protein binding pocket, a spatially defined ligand binding cavity in proteins, serves as the foundation for computational drug design by determining non-covalent interac-

tions and stereochemical complementarity with ligands.¹ This structure enables accurate prediction of protein-ligand binding affinity, a key determinant of drug efficacy and target selectivity. Accurately predicting this critical binding affinity plays a key role in two sequential steps in target-based drug

discovery: high-throughput virtual screening, which aims to discover active compounds that can bind to an interested human protein from large-scale chemical libraries,^{2–4} and hit-to-lead optimization, which aims to refine the structure of these active ligands to improve their binding affinity and drug-like properties.^{5,6} For virtual screening, docking methods have been widely used with several successful applications.⁷ Still, they suffer from the trade-off between the chemical search space size and computational cost.⁸ As for hit-to-lead optimization, physics-based calculation methods, including free energy perturbation (FEP)⁹ and molecular mechanics-generalized born surface area (MM/GBSA),¹⁰ have shown encouraging performance but are either poorly correlated with experimental binding affinities or require large computational resources for extensive sampling.^{11,12} To address the challenges faced by computational methods, in recent years, machine learning (ML) methods have been developed for both virtual screening (e.g., DrugCLIP¹³) and hit-to-lead optimization (e.g., ActFound¹⁴ and PBCNet¹⁵). These ML approaches achieve comparable performance to computational methods while offering significant improvements in computational efficiency, thereby making them highly efficient alternatives for large-scale applications.

Despite the encouraging performance of existing ML methods, virtual screening and hit-to-lead optimization are often studied separately. However, these two tasks are interdependent and complementary to each other. A model solely focused on hit-to-lead optimization might be constrained within a limited chemical space, hindering its ability to generalize to ligands with novel chemical scaffolds. Conversely, a model developed exclusively for virtual screening might overlook crucial substructures that determine protein-ligand interactions, thus failing to distinguish between ligands with subtle structural differences. Several recent works (e.g., GenScore,¹⁶ PIGNet2,¹⁷ IGModel,¹⁸ and EquiScore¹⁹) have begun bridging the gap between these two tasks by data augmentation techniques like re-docking, cross-docking, and decoy generation. However, these approaches still face challenges in collecting structurally similar active ligand pairs, which is crucial for learning subtle substructural variations that affect binding affinity. Based on these observations, we hypothesize that a unified foundation model jointly addressing virtual screening and hit-to-lead optimization can leverage the synergy between these two tasks to enhance overall performance.

In this work, we propose LigUnity (ligand unified affinity), a protein-ligand affinity foundation model. The core innovation of LigUnity lies in its integrated capabilities for both broad screening and precise affinity prediction, achieved through a combination of scaffold discrimination and pharmacophore ranking. We jointly embed ligands and protein pockets into a shared space that captures their structural and chemical complementarity. For scaffold discrimination, LigUnity learns to differentiate between active and inactive ligands, providing insights into diverse protein-ligand interactions that help establish global structure-activity relationships across chemical scaffolds. For pharmacophore ranking, LigUnity refines the shared space by learning to order active ligands for each pocket, revealing how subtle structural differences affect binding affinity. With these computed embeddings, LigUnity enables rapid screening of large virtual libraries and efficient identification of active li-

gands, making it suitable for both virtual screening and hit-to-lead optimization tasks.

However, collecting a structure-aware dataset suitable for both tasks poses a significant challenge. While large-scale affinity datasets like BindingDB²⁰ and ChEMBL²¹ provide abundant binding data, they lack the structural details necessary for identifying binding pocket structures. To address this, we introduce PocketAffDB, a comprehensive structure-aware binding assay database containing 0.8 million affinity data points across 0.5 million unique ligands and 53,406 pockets. PocketAffDB is organized by assays, in which all affinity measurements share the same experimental methods and are directly comparable. Building on this, we propose a simple but effective assay-guided pocket-matching method to assign a binding pocket structure to each protein-ligand pair, based on the observation that most assays are designed for a specific binding site of interest. To our best knowledge, PocketAffDB represents the largest affinity dataset that integrates bioassay data with binding pocket structure, providing a valuable resource for affinity prediction.

To validate the effectiveness of LigUnity, we conducted comprehensive evaluations on eight benchmarks across six settings. On three virtual screening benchmarks (DUD-E,²² DeKoIs,²³ and LIT-PCBA²⁴), LigUnity outperforms all 24 competing methods. When applied to unseen proteins, our model achieves consistent improvement, demonstrating its robust generalization ability to novel targets. For hit-to-lead optimization, LigUnity shows superior performance on two FEP benchmarks in both zero-shot and few-shot settings, suggesting that LigUnity can be potentially used as an efficient alternative to the costly FEP calculations.⁹ We further assessed LigUnity's adaptability in split-by-time, split-by-scaffold, and split-by-unit settings on ChEMBL²¹ and BindingDB²⁰ datasets, with LigUnity again outperforming eight competing methods. Moreover, in an active learning framework simulating multi-iteration drug discovery optimization, LigUnity successfully identifies ligands with optimal binding affinity in just a few iterations, illustrating its practical value to minimize the experimental cost in molecular optimization. Collectively, LigUnity serves as a foundation model for both virtual screening and hit-to-lead optimization, offering broad applicability across the drug discovery pipeline.

RESULTS

Overview of LigUnity

LigUnity is a protein-ligand affinity foundation model for both virtual screening and hit-to-lead optimization, employing 3D binding pocket structure to predict protein-ligand binding affinities. To train the model, we developed PocketAffDB, a comprehensive structure-aware dataset curated from large-scale experimental affinity databases (BindingDB²⁰ and ChEMBL²¹) and PDB (Figure 1A). We first collect affinity data into assays and then assign a binding pocket structure to each protein-ligand pair using assay-guided pocket matching (see [methods](#)). To our knowledge, PocketAffDB represents the largest integrated affinity dataset combining bioassay data with structural pocket information, comprising 0.8 million affinity data points spanning 0.5 million unique ligands, 53,406 pockets, and 26,748 assays.

In the pre-training stage, LigUnity learns a pocket-ligand shared space to capture structural and chemical

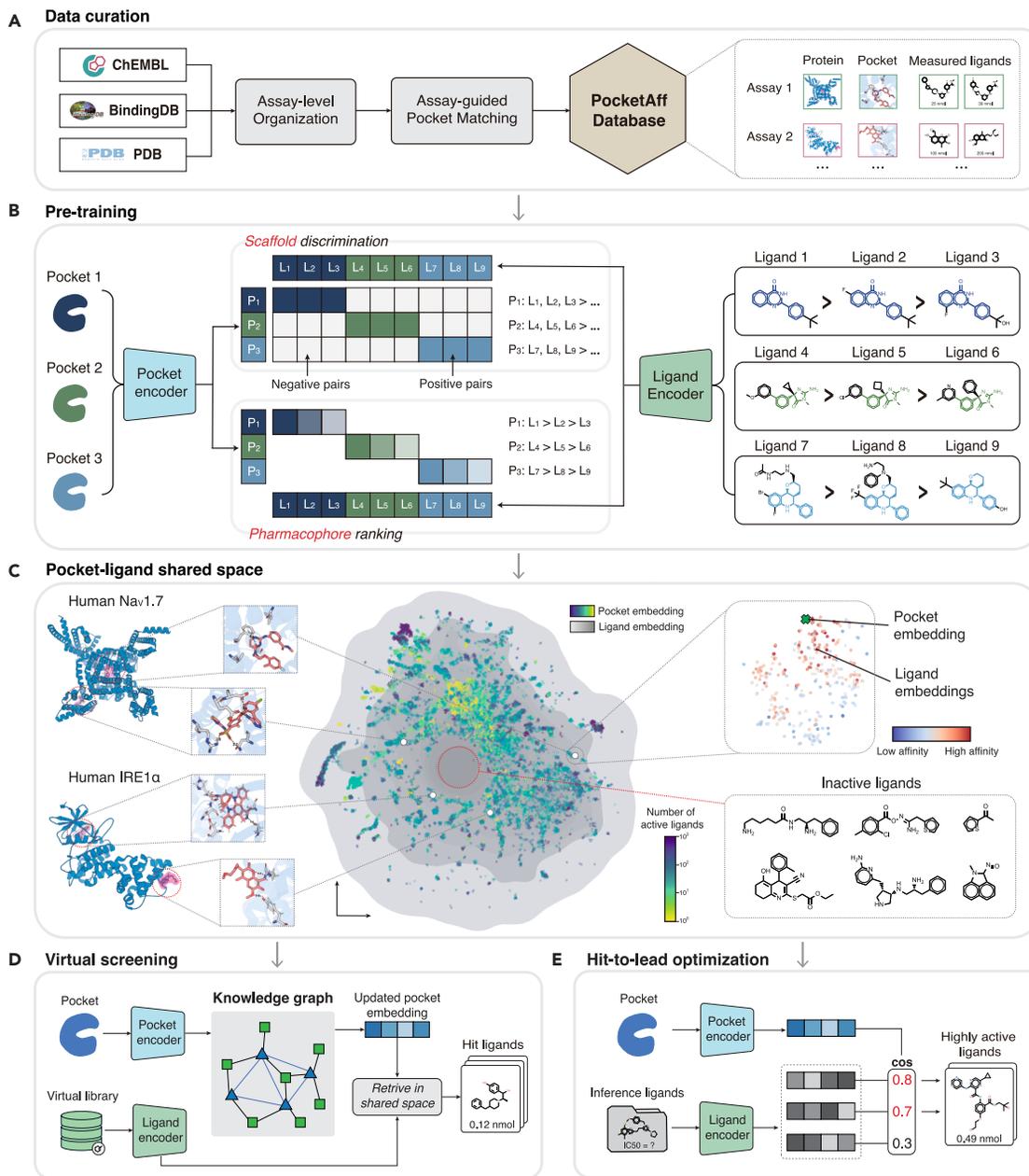


Figure 1. Overview of LigUnity

(A) Data curation pipeline of PocketAffDB.

(B) LigUnity first pre-trains the pocket and ligand encoders by a hierarchical affinity landscape navigation. It first projects pockets and ligands into a shared embedding space and then exploits scaffold discrimination to capture coarse-grained active/inactive distinction and pharmacophore rankings to refine the shared embedding space by aligning it with subtle affinity differences.

(C) Uniform manifold approximation and projection (UMAP) plot showing ligand and pocket embeddings derived by LigUnity.

(D) For virtual screening, LigUnity first refines the query pocket embedding by a graph neural network, and the updated pocket embedding is then utilized to retrieve hit ligands in the pocket-ligand shared space.

(E) For lead optimization, LigUnity is directly used to rank unmeasured ligands based on cosine similarity between pocket and ligand embeddings and find highly active ligands among them.

complementarity between pockets and ligands. We implement this through an integrated framework combining scaffold discrimination and pharmacophore ranking (Figure 1B). The scaffold discrimination module learns to distinguish active/inactive compounds by focusing on structural differences in the

chemical scaffold, creating an embedding space where positive pocket-ligand pairs are attracted while dissimilar negative pairs are repelled. The pharmacophore-ranking component then refines this embedding space through fine-grained alignment with subtle affinity differences by focusing on the

pharmacophore of ligands, where similarity rankings between ligand and pocket embeddings are used to predict relative binding affinities. These two modules complement each other and thus are employed simultaneously in the pre-training stage.

After pre-training, the embedding space of LigUnity demonstrates meaningful hierarchical patterns (Figure 1C). At the coarse-grained level, ligands binding to the same pocket form visible clusters, whereas ligands targeting different pockets of the same protein are well separated, indicating that embeddings of LigUnity can capture coarse-grained binding site information, showing potential for virtual screening. At the fine-grained ligand level, the embedding space shows that ligands with higher binding affinity values are closer to their target pockets, indicating that LigUnity learns pharmacophore-level differences affecting binding affinity, demonstrating the ability for hit-to-lead optimization.

During inference, LigUnity is adapted separately for virtual screening and hit-to-lead optimization to leverage task-specific data. For virtual screening, LigUnity employs a graph-based approach to utilize large-scale protein-ligand interactions in existing databases (Figure 1D). Specifically, we first built a large-scale knowledge pocket-ligand graph, containing 0.83 million pocket-ligand edges with known interactions and 16 million pocket-pocket edges based on structural similarity. We then applied a graph neural network to refine the query pocket embedding by aggregating neighboring information. For hit-to-lead optimization, LigUnity directly gives predictions for unmeasured ligands by simply computing cosine similarity between pocket and ligand embeddings (Figure 1E). We can also enhance prediction accuracy by fine-tuning on a few experimentally measured ligands, which are often available for hit-to-lead optimization. Through modeling the affinity landscape in an embedding space, LigUnity substantially reduces the computational cost during inference, achieving six orders of magnitude speedup compared to traditional docking methods.

LigUnity improves virtual screening

We first evaluated LigUnity for virtual screening on 102 protein targets from the DUD-E benchmark and 81 protein targets from the DEKOIS 2.0 benchmark.^{22,23} These two benchmarks cover a diverse set of targets spanning enzymes, ion channels, G-protein-coupled receptors (GPCRs), and transcription factors (Figures S1 and S2). We compared LigUnity with 24 different competing methods, including molecular docking methods,^{25,26} structure-based methods,^{16,27,28} and structure-free methods,^{13,29} and observed that LigUnity outperformed all competing methods in both benchmarks (Figures 2A, 2B, and S3–S6; Table S1). Specifically, when compared to the best competing structure-based methods, Densiv-G²⁸ and RTMScore,²⁷ LigUnity still achieves over 50% improvements in the enrichment factor (EF) 1% ($p < 10^{-9}$). Meanwhile, once the embeddings are computed, LigUnity is 10^6 times faster than Glide-SP in the screening speed, as LigUnity does not need docking poses (Figure S7). To assess the ability of LigUnity to generalize to unseen proteins, we excluded training proteins that are similar to any test protein in terms of sequence similarity (Figures 2C, 2D, S8, and S9). When using a stringent threshold of 30% sequence similarity, LigUnity still significantly outperforms

both DrugCLIP and the commercial docking software Glide-SP ($p < 0.05$), demonstrating its generalizability to novel proteins.

Next, we evaluated LigUnity on the LIT-PCBA²⁴ benchmark, which contains experimentally measured affinities for active and inactive ligands across 15 targets. Compared to DUD-E and DEKOIS, LIT-PCBA presents a more challenging setting for virtual screening, as it contains more inactive ligands that are structurally similar to active ligands. In contrast, DUD-E and DEKOIS utilize experimentally unverified decoys from the public compound database (ZINC)³⁰ while excluding decoys structurally similar to active ligands to avoid potential false negatives, which may make them relatively easy for machine learning models. Furthermore, LIT-PCBA maintains a 1:1,000 active-to-inactive ratio, which better reflects the practical conditions encountered in real-world drug discovery. Nevertheless, LigUnity still achieves superior performance on the more challenging LIT-PCBA benchmark (Figures 2E and 2F). The improvement is also achieved when excluding training proteins that have large sequence similarity to any test protein using varying thresholds (Figure S10), reassuring the effectiveness and generalizability of LigUnity for virtual screening.

Finally, we conducted ablation studies to study the contribution of each module in LigUnity (Figures S11–S13). First, we found that ablating the scaffold discrimination component leads to a performance drop of over 60% in terms of EF 1%, confirming its crucial role in distinguishing active ligands from decoys in virtual screening tasks. Second, the ablation of the heterogeneous graph neural network (H-GNN) leads to consistent decreases in performance, demonstrating its capability to utilize information from the pocket-ligand heterogeneous graph to improve the predictive power. Furthermore, we found that the pharmacophore ranking is particularly effective on LIT-PCBA among three benchmarks (Figure S13), as it helps LigUnity distinguish subtle structural differences among similar ligands that are crucial to their affinity scores. Together, our ablation studies demonstrate the importance of all three key technical ideas proposed by LigUnity, necessitating the joint optimization of virtual screening and hit-to-lead optimization.

LigUnity improves hit-to-lead optimization

Hit-to-lead optimization aims to optimize the structure of the ligand for improved affinity by ranking candidate ligands that are commonly structurally similar to initial hits identified to be active. After observing the substantial improvement of LigUnity in virtual screening, especially its ability to capture ligand structure differences, we next investigated whether LigUnity can be applied to hit-to-lead optimization. We evaluated LigUnity on two FEP benchmarks (JACS⁹ and Merck³¹) to predict binding free energies, which are the most widely used indicators for hit-to-lead optimization. These two benchmarks contain 16 targets, and each target has an average of 29 experimentally measured ligands. Following previous work,¹⁴ assays similar to those included in the FEP benchmarks are excluded in the pre-training stage to avoid potential data leakage (see methods). Ligands in the FEP benchmarks are also removed from the pre-training data.

In the zero-shot setting, where the model does not use any measured ligands for test proteins during pre-training and fine-tuning, our model surpasses existing computational methods

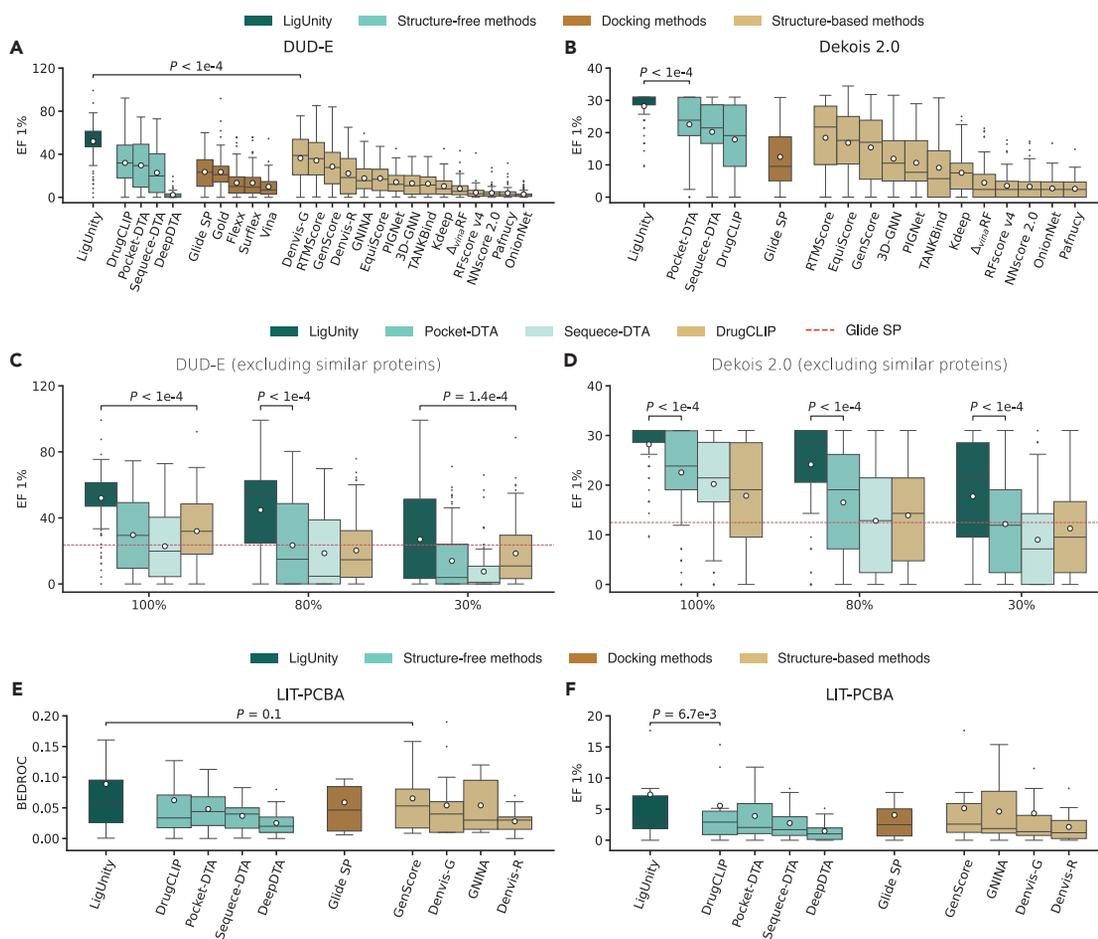


Figure 2. Evaluation on virtual screening

(A and B) Boxplots comparing LigUnity and competing methods in the DUD-E and Dekois-2.0 benchmarks in terms of enrichment factor (EF) 1%. The mean values (white dots) are calculated across $n = 102$ and 81 targets for the DUD-E and Dekois-2.0 benchmarks, respectively.

(C and D) Boxplots comparing LigUnity and structure-free methods in the DUD-E and Dekois-2.0 benchmarks in terms of EF 1% using different protein training sets. The x axis denotes the maximum sequence similarity between the training and test sets. The mean values (white dots) are calculated across $n = 102$ and 81 targets for the DUD-E and Dekois-2.0 benchmarks, respectively.

(E and F) Boxplots comparing LigUnity and competing methods on the LIT-PCBA benchmark in terms of EF 1% and boltzmann-enhanced discrimination of receiver operating characteristic (BEDROC) score ($\alpha = 80.5$). The mean values (white dots) are calculated across $n = 15$ targets.

(A–F) p values indicate the significance for LigUnity outperforming the best comparison approach in a one-sided Wilcoxon test.

(e.g., Glide-SP²⁵ and MM/GBSA³²), structure-based methods (e.g., GenScore¹⁶), and structure-free methods (e.g., DrugCLIP¹³) (Figure 3A) on 8 targets of the Merck FEP benchmark, demonstrating the effectiveness of our method. To further examine the performance of LigUnity in hit-to-lead optimization, we studied three more challenging settings: (1) a no similar ligands setting, where training ligands with more than 50% Tanimoto similarity to any test ligand are excluded; (2) a no similar proteins setting, where training proteins with more than 30% sequence similarity to any test protein are excluded³³; and (3) a no similar ligands and proteins setting, where the criteria of the first two settings are applied to exclude similar proteins or ligands. LigUnity again consistently outperforms competing methods across all three settings (Figure S14).

In real-world drug discovery, binding data for a few ligands of interest are often experimentally measured. A successful hit-to-lead optimization model should be able to leverage these binding

data to enhance its predictions. To examine the ability to employ these data, we fine-tuned LigUnity and competing methods with varying proportions of ligand binding data. We observed that LigUnity consistently outperforms competing methods across all four settings (Figures 3B, 3C, and S15; Table S2). Even under the most challenging setting where both similar ligands and proteins are excluded, LigUnity achieves an r^2 of 0.472 on the Merck benchmark after fine-tuning on 80% of ligands (23.2 ligands on average for each target), reaching performance comparable to FEP+(OPLS4) ($r^2 = 0.528$), a leading commercial computational software for calculating the binding free energy. These results demonstrate the practical value of LigUnity for affinity prediction, offering an accurate and efficient alternative to resource-intensive computational chemistry approaches.

Similar to our analyses in the virtual screening, we conducted ablation studies to investigate the contribution of each component in LigUnity (Figure S16). Our results demonstrate that

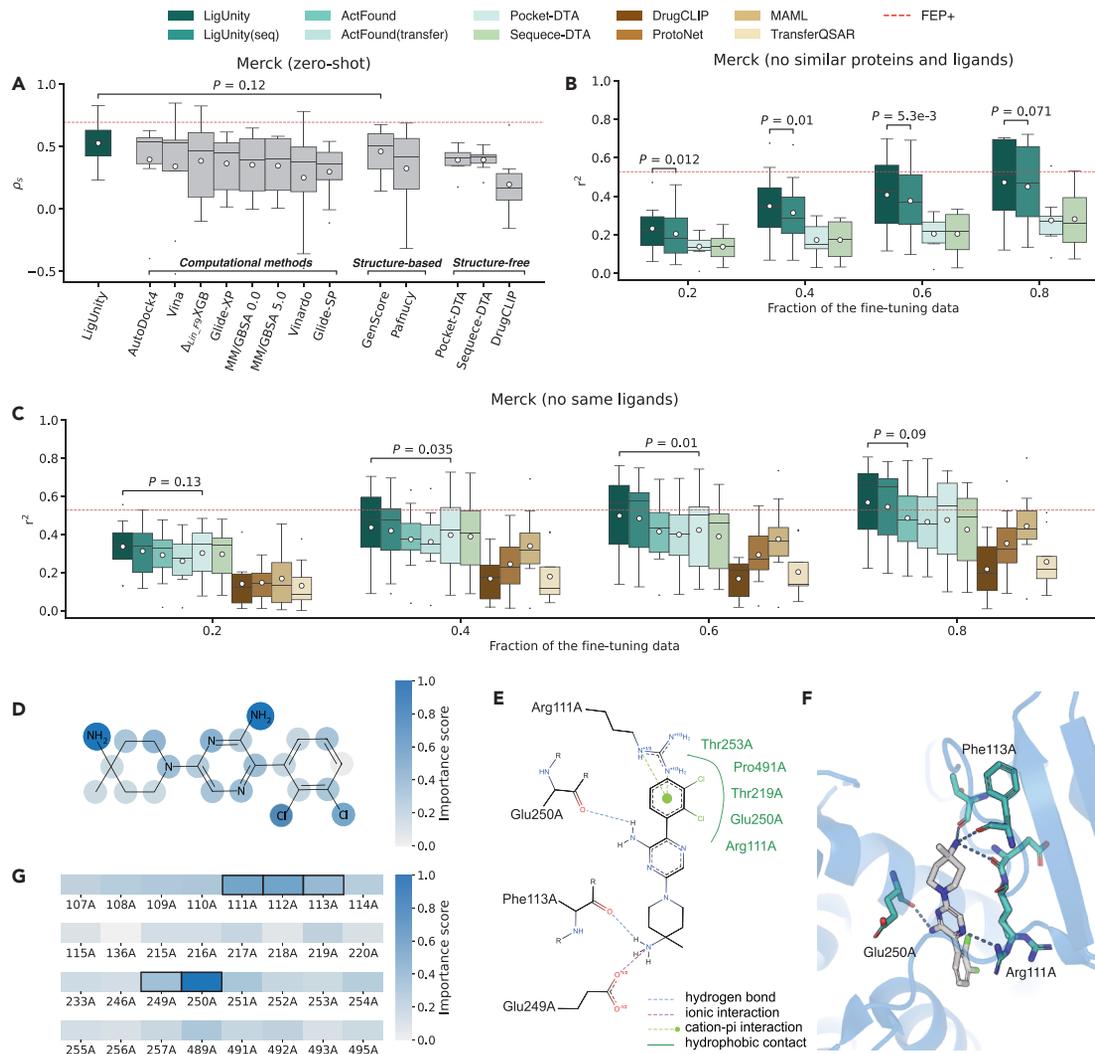


Figure 3. Evaluation on hit-to-lead optimization

(A) Boxplot comparing the binding affinity prediction on the Merck benchmark in terms of Spearman's rank correlation (ρ_s) in the zero-shot setting. The mean values (white dots) are calculated across $n = 8$ targets.

(B and C) Boxplots comparing the binding affinity prediction across different settings on the Merck benchmark in terms of r^2 when 20%, 40%, 60%, and 80% of the experimental binding affinities are used for fine-tuning. The mean values (white dots) are calculated across $n = 8$ targets.

(D) Case study on the SHP2 target showing the importance score of each ligand atom predicted by LigUnity.

(E and F) 2D (E) and 3D (F) interaction graph showing the non-covalent interaction between the ligand and pocket (PDB: 5EHR).

(G) Case study on the SHP2 target showing the importance score of each pocket residue predicted by LigUnity. The top five predicted key residues are highlighted with black boxes.

(A–C) p values indicate the significance for LigUnity outperforming the best comparison approach in a one-sided Wilcoxon test.

excluding the scaffold discrimination or the pharmacophore-ranking component leads to substantial performance drops. Notably, ablating the pharmacophore ranking leads to a performance drop of over 50% in both zero-shot and few-shot settings, highlighting the effectiveness of the ranking objective in capturing the relative activity differences for hit-to-lead optimization tasks. To verify that LigUnity captures meaningful interaction patterns, we systematically tested it on FEP benchmarks by masking interaction-critical residues. Key residues were identified from PDB structures using ProteinPlus.³⁴ The results show that masking strong H-bond/ionic bond residues caused significant performance drops ($\Delta\rho_s = -6.34\%$) vs. minimal impact

from random masking ($\Delta\rho_s = -0.46\%$) (Table S3), demonstrating LigUnity's ability to rely on protein-ligand interactions to achieve good performance.

Finally, to help understand the structural significance of individual atoms and residues involved in the interaction, we calculated the importance score of each atom and protein pocket residue on the SHP2 target by masking them individually (Figures 3D–3G and S17; see methods). By examining the atom importance scores, LigUnity correctly identifies two amino groups (-NH₂) as key contributing factors to the predicted binding affinity (Figure 3D), aligning with the crystallographic observations of their hydrogen bonding and ionic interactions

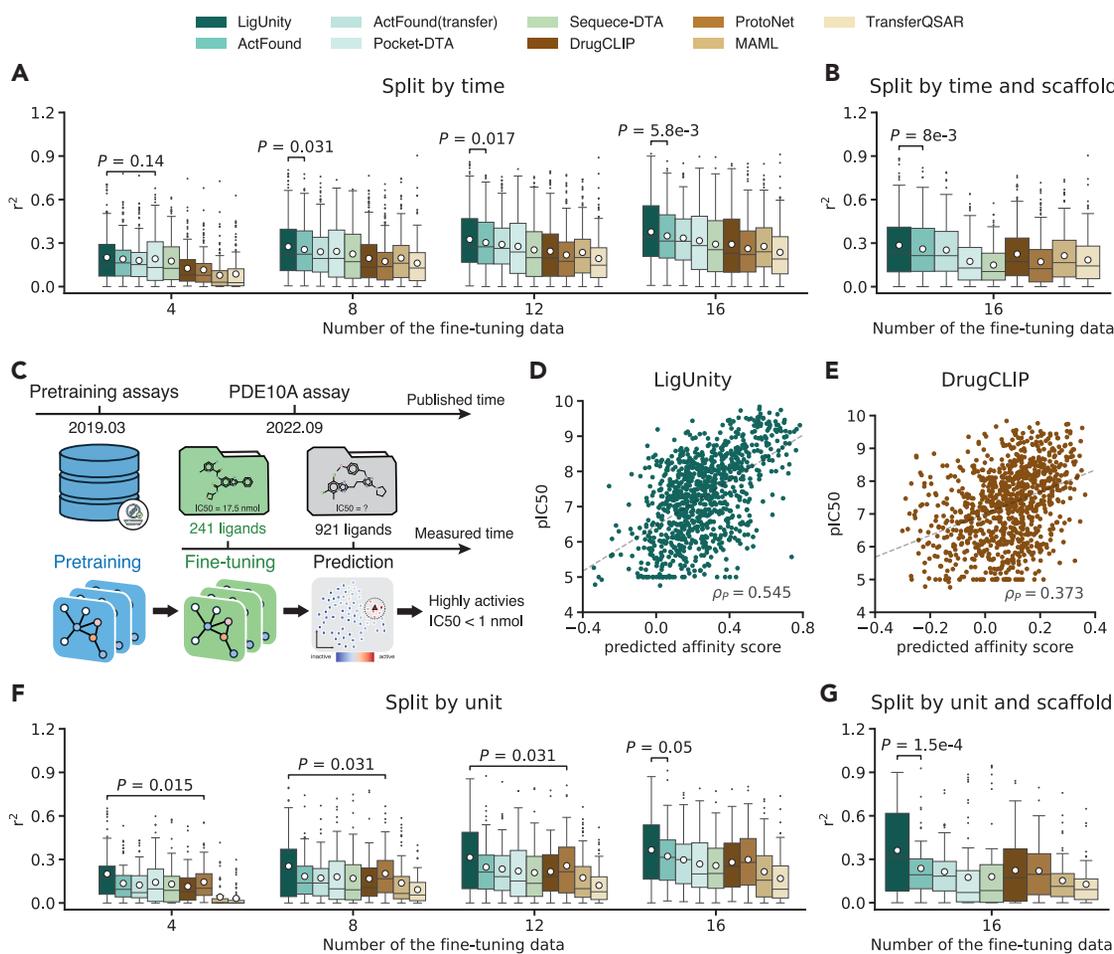


Figure 4. Evaluation on diverse settings

(A and B) Boxplots comparing the binding affinity prediction on the split-by-time setting in terms of r^2 when 4, 8, 12, and 16 experimentally measured ligands are used for fine-tuning. The mean values (white dots) are calculated across $n = 161$ assays.

(C) Experimental setting for case study on the phosphodiesterase 10A (PDE10A) target.

(D and E) Scatterplots comparing the predicted affinity score and $pIC_{50} = -\log_{10}IC_{50}$ for LigUnity (F) and DrugCLIP (G) when fine-tuned with the first measured 20% ligands.

(F and G) Boxplots comparing the binding affinity prediction on the split-by-unit setting in terms of r^2 when 4, 8, 12, and 16 experimentally measured ligands are used for fine-tuning. The mean values (white dots) are calculated across $n = 65$ assays.

(A, B, D, E, F, and G) p values indicate the significance for LigUnity outperforming the best comparison approach in a one-sided Wilcoxon test.

(Figures 3E and 3F). Furthermore, using protein residue importance scores, the model correctly highlights the crucial role of Glu250E in hydrogen bonding, along with the polar interactions involving Phe113A, Arg249A, and Arg111A (Figure 3G). Although the pose-free nature of our approach results in certain limitations, such as underestimating the contributions from specific hydrophobic cavity residues (Thr253A and Pro491A), LigUnity successfully identifies key interaction patterns. These results suggest that LigUnity can serve as an interpretable computational tool for studying structure-activity relationships and guiding hit-to-lead optimization.

LigUnity as a versatile foundation model for different applications

To further explore the ability of LigUnity as a foundation model for affinity prediction, we next sought to evaluate its performance on

assays from ChEMBL²¹ and BindingDB,²⁰ two comprehensive public databases containing experimentally measured binding affinities covering various aspects of assay types, such as protein-based, cell-based, cell-membrane-based, and subcellular-based assays.^{20,21} We first used a split-by-time setting to evaluate the performance, where the model is pre-trained using 18,552 assays released before March 2019 and tested on 161 assays released after March 2019 in ChEMBL and BindingDB. Each test assay has 48.6 ligands on average. To avoid potential data leakage, we further excluded test assays whose targets exist in the pre-training data.

LigUnity attains consistent improvement using varying numbers of fine-tuning data ranging from 4 to 16 (Figures 4A and S18). When compared to ActFound,¹⁴ the best-competing method specifically designed for few-shot affinity prediction, LigUnity achieved an average improvement of 8.5% in terms of

r^2 . Since ActFound is a ligand-based model without considering pocket information, our substantial improvement reflects the effectiveness of using protein pocket information for affinity prediction. In a more challenging split-by-scaffold setting where test ligands have distinct scaffolds from training ligands, LigUnity again obtains the best performance (Figure 4B), demonstrating its strong generalizability to unseen chemical scaffolds.

Next, we conducted a case study on phosphodiesterase 10A (PDE10A), a verified therapeutic target for antipsychotic therapy with no approved drugs, to assess the applicability of LigUnity in the real-world drug discovery pipeline. This dataset was released by Roche in 2022, comprising 1,162 ligands with the experimentally measured binding affinity and timestamp of each experiment.³⁵ We used a split-by-time setting for ligands in the PDE10A dataset, where the first measured 20% ligands are used for fine-tuning and the remaining 80% are used for testing (Figure 4C). To ensure a rigorous split-by-time evaluation, we also restricted the pre-training data to the assays released before March 2019. To this end, our fine-tuned model achieves a Pearson correlation of 0.55 on the test set (Figure 4D), substantially outperforming the 0.37 Pearson correlation of DrugCLIP (Figure 4E). Building on this performance, our model demonstrates a strong ability to identify highly potent ligands, successfully identifying 4 highly active ligands ($IC_{50} < 1$ nmol) among the top-10 predicted ligands (Figure S19) and 24 highly active ligands among the top-50 predicted ligands (compared to 11 for DrugCLIP). Collectively, these results demonstrate the superior capability of LigUnity in identifying highly potent ligands, holding the potential to reduce the cost of hit-to-lead optimization.

Finally, we studied a split-by-unit setting where the test assays use percentage (%) units, a widely used format (24.4% in ChEMBL) in practical drug design with limited exploration in ML methods. Assays with percentage units pose challenges for affinity prediction, as they distribute differently from pre-training assays with molar concentrations (e.g., nmol) or density units (e.g., $\mu\text{g/mL}$). We hypothesize that our pharmacophore-ranking approach enables LigUnity to be robust against assays with different units, as it aims to learn the pocket-specific ligand ranking instead of the absolute affinity value. To this end, we examined a split-by-unit setting by using ChEMBL assays reporting activities in percentage (%) as test assays and assays with defined molar concentration or density units as pre-training assays. This test set contains 65 assays, and each assay has, on average, 30.5 ligands. When fine-tuned with different numbers of ligands (4–16), LigUnity consistently surpasses all competing methods (Figures 4F, 4G, and S18). Notably, compared to the regression model Pocket-DTA, LigUnity shows a 40.2% improvement in this split-by-unit setting, substantially higher than the 13.8% improvement in the split-by-time setting, demonstrating the generalization ability of the pharmacophore-ranking approach across different measurement units, reassuring LigUnity as an effective foundation model for drug discovery.

LigUnity boosts the active learning framework for drug discovery

In the real-world drug discovery pipeline, multiple iterations of molecular optimization are often required due to limited experimental resources and the difficulty in accurately predicting protein-ligand binding affinity. As a result, active learning is often ex-

ploited in this process. Building on this, we explored how LigUnity can be integrated with active learning to optimize ligands for tyrosine kinase 2 (TYK2), a therapeutic target for autoimmune diseases. We employed a dataset that contains binding free energy for 10,000 ligands calculated using FEP, consuming approximately 80,000 GPU hours (9.1 GPU years).³⁶ Following established active learning protocols,³⁷ we aimed to identify highly active ligands with limited computational resources in terms of the amount of FEP calculations.

To this end, we integrated LigUnity in an active learning framework, where we first trained the model using a small number of randomly selected ligands with known binding free energies. In each subsequent iteration, we selected a subset of unlabeled ligands and performed FEP calculations to determine their binding free energies. These newly labeled data points were then incorporated into the training set, and the model was retrained to refine its predictions (Figure 5A).

We compared all methods under two selection strategies: a greedy strategy and an exploration-exploitation strategy (see methods). We first found that LigUnity successfully identified the best active ligands within a few iterations (Figure 5B). The proportion of highly active ligands continually increased after each iteration (Figure 5C), demonstrating the effectiveness of the active learning framework in identifying highly active ligands. Moreover, we found that LigUnity consistently outperforms other competing methods under both greedy and exploration-exploitation strategies, achieving over 40% improvement in terms of r^2 (Figure 5D). By comparing two different selection strategies, we found that the greedy strategy has a higher recall in early iterations by focusing on the most promising candidates (Figures 5D and S20). Interestingly, the exploration-exploitation strategy attains superior performance in later iterations, primarily because of its more diverse ligand sampling in early iterations that helps improve model accuracy. These findings demonstrate the broad applicability of LigUnity in drug design, particularly in identifying highly active ligands using limited experimental and computational resources.

DISCUSSION

We have presented LigUnity, a protein-ligand affinity foundation model for both virtual screening and hit-to-lead optimization. Our model learns a pocket-ligand shared space to capture structural and chemical complementarity between pockets and ligands, which is implemented through an integrated framework combining scaffold discrimination and pharmacophore ranking. For virtual screening, our experiments have shown that LigUnity outperformed 24 competing methods across three benchmarks (DUD-E, LIT-PCBA, and Dekois 2.0). For hit-to-lead optimization, our model achieves the best performance in two FEP benchmarks and demonstrates promising accuracy in split-by-time, split-by-scaffold, and split-by-unit settings. Moreover, by developing an active learning framework to simulate real-world lead optimization, LigUnity efficiently identifies optimal binding compounds with limited computational resources. Collectively, these results demonstrate the strong generalization ability and practical usability of LigUnity throughout the early-stage drug discovery pipeline, establishing our confidence in LigUnity as a versatile foundation model for computer-aided drug discovery.

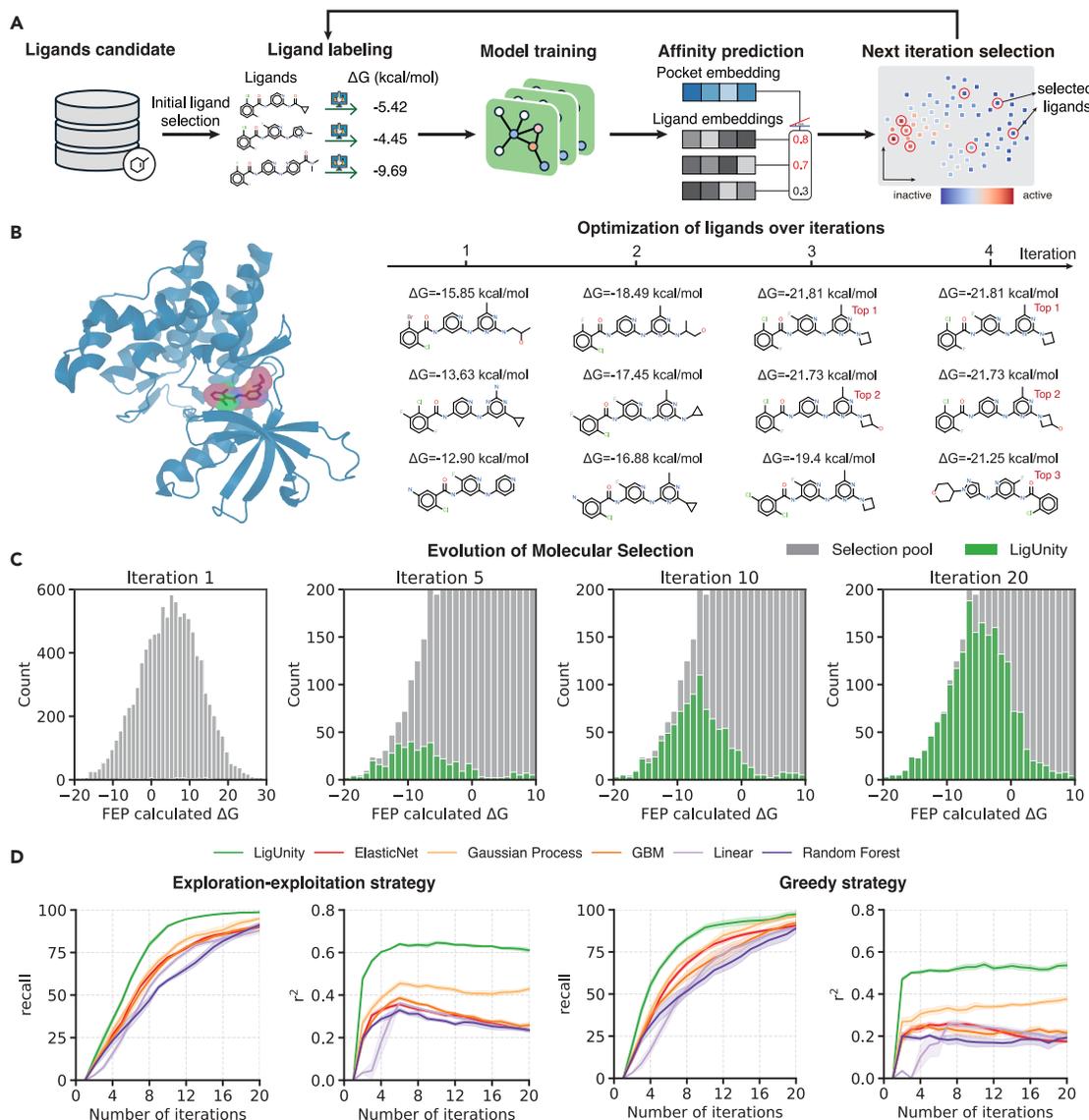


Figure 5. Evaluation of LigUnity in an active learning framework on TYK2

(A) An active learning framework based on LigUnity.

(B) 3D structure for the TYK2 target (left, PDB: 4GIH) and the top ligands selected for iterations 1, 2, 3, and 4 (right) under the exploration-exploitation strategy; FEP-calculated binding free energies for each ligand are also shown.

(C) Histogram plots showing the distribution of ΔG values of selected ligand for each iteration based on a greedy selection strategy.

(D) Plots comparing LigUnity and competing methods on the TYK2 dataset in terms of top 2% recall and r^2 when using the exploration-exploitation strategy (left) and greedy selection strategy (right).

Two main lines of work are related to LigUnity, including structure-free drug-target affinity (DTA) prediction methods and structure-based methods. Compared to DTA methods,^{29,38–46} which directly predict binding affinity using protein and ligand information, LigUnity presents three key differences. First, DTA methods are challenged by the small number of proteins with measured affinity (2,773 in BindingDB), whereas LigUnity incorporates fine-grained pocket information and leverages a comprehensive training set covering 53,406 pockets for 4,847 proteins, enabling robust performance on understudied proteins. Second, existing DTA methods perform direct regression of absolute affinity values,

whereas LigUnity learns a pocket-ligand shared space using scaffold discrimination and pharmacophore ranking, reflected by the consistent performance improvement in our experiments. Moreover, our pharmacophore-ranking approach inherently functions as a normalization technique, focusing on relative ligand rankings rather than absolute affinity values, which mitigates the impact of experimental variability (e.g., pH and temperature) and assay-type differences (e.g., IC₅₀ vs. K_D). Third, DTA methods require intensive computation for each prediction, whereas LigUnity enables rapid screening by pre-computing and storing embeddings, leading to better efficiency for large-scale applications. Compared

to structure-based methods^{16,19,27,28,47–57} that rely on 3D binding poses, LigUnity presents three improvements. First, structure-based methods are constrained by the scarcity of ligands with co-crystal structures. For example, only 7.5% of 0.57 million compounds with measured activities in BindingDB have crystal structures.²⁰ In contrast, LigUnity is developed using a diverse training set of 0.45 million unique ligands, enabling effective generalization to novel chemical scaffolds. Second, structure-based methods require computationally expensive pose generation (approximately 1 min per compound using Glide), whereas LigUnity achieves competitive performance without the need for pose information, making it applicable to large-scale screening. Third, structure-based methods mostly rely on static protein conformations, whereas LigUnity implicitly accounts for pocket flexibility by leveraging multiple slightly varied structures for each pocket (averaging 10.96 pocket structures per assay). Furthermore, our approach offers significant flexibility for diverse scenarios. For novel targets with limited structural data, LigUnity can leverage predicted pockets from AlphaFold 3 or homology modeling. In cases where no experimental or predicted structures exist, we have also provided sequence-based variants of LigUnity.

Despite its promising performance in binding affinity prediction, LigUnity has two limitations that can be addressed in future work. First, LigUnity can only be applied to assays with known protein targets, limiting its applications to diverse assays without target information (e.g., phenotype assays). To expand the applicability of our model, we plan to incorporate additional modalities such as gene information in future work. Second, while LigUnity can successfully identify key substructures and residues and learn meaningful interaction patterns to improve its performance, its pose-free nature limits its ability to provide detailed mechanistic insights related to binding pose. Therefore, we plan to couple LigUnity with binding poses and develop a structure-based variant to enhance prediction accuracy and mechanistic interpretability, particularly for scenarios where sufficient computational resources are available to obtain precise binding poses.

METHODS

Problem setting

LigUnity aims to predict binding affinity between given binding pockets and unmeasured ligands. Our training data are organized by assay, which is defined as the experimental procedure designed to evaluate the binding affinity of ligands against a specific protein target. Each assay has multiple measured ligands and only one target protein. A key consideration is that affinity values are only comparable within the same assay, and comparing affinity values from different assays targeting the same protein may not be reliable.⁵⁸ This incomparability arises from variations in experimental conditions (e.g., cofactor concentration, pH, and temperature), assay formats (e.g., cell-based vs. target-based assays and different detection methods), and affinity measurements (e.g., IC₅₀, K_i, and K_D). Therefore, our method focuses on learning relative affinity rankings within each assay instead of absolute values.

Let \mathcal{A} denote the set of assays. For each assay $A_i \in \mathcal{A}$, L_i denotes the set of tested ligands, and $v_i(l)$ denotes the affinity value of ligand $l \in L_i$. Each assay corresponds to one target protein, which may have multiple PDB structures available. We define P_i

as the set of protein pocket structures for assay A_i and randomly sample one pocket structure from P_i at each training step.

Pocket and ligand encoder

In the pre-training stage, our pocket encoder and ligand encoder are initialized with Uni-Mol,⁵⁹ an SE(3)-invariant graph transformer for 3D molecules, which is a structure-free method and does not rely on the binding pose as input. For the ligand branch, we first use the RDKit to generate 3D conformations for each ligand, which have been proven effective for various tasks,⁶⁰ and then input them into the Uni-Mol ligand encoder to generate ligand embeddings. Similarly, for the pocket branch, we input the protein pocket conformation into the Uni-Mol pocket encoder to generate the protein embedding. For a pocket-ligand pair (p, l) , their vector representations are obtained through

$$(\mathbf{p}, \mathbf{l}) = (E_p(p), E_l(l)), \quad (\text{Equation 1})$$

where E_p and E_l denote the pocket and ligand encoders, respectively; \mathbf{p} denotes the protein embedding; and \mathbf{l} denotes the ligand embedding.

Hierarchical affinity landscape navigation

In the pre-training stage, LigUnity learns a pocket-ligand shared space using two complementary strategies: scaffold discrimination and pharmacophore ranking. The scaffold discrimination approach captures the coarse-grained distinction between active and inactive pocket-ligand pairs by focusing on structural differences in the chemical scaffold, which aims to bring the embeddings of positive pocket-ligand pairs closer and push the negative pairs apart. The pharmacophore-ranking approach learns to align the embedding space with the fine-grained affinity ranking of measured ligands by focusing on the pharmacophore. These two approaches complement each other, and we jointly optimize them to make the model suitable for both virtual screening and hit-to-lead optimization scenarios.

During inference, the model only needs to take 3D protein pockets and 3D ligand conformations (generated by the RDKit) as input. It then calculates the pocket and ligand embeddings and outputs their cosine similarity as the predicted affinity score.¹³ This differs from most previous ML affinity prediction models, which typically model affinity prediction as a regression task and use either dual-tower architectures (structure-free methods)^{29,42,44} or require binding poses (structure-based methods)^{27,28} as input. Once the pocket and ligand embeddings are pre-computed, LigUnity enables fast screening by similarity-based comparison in joint embedding space, achieving six magnitudes of speedup compared to docking methods.¹³

Pharmacophore ranking

Pharmacophore ranking is an approach that prioritizes active compounds based on their binding potential and structural nuances. We implement pharmacophore ranking through listwise learning. Unlike pointwise approaches that predict absolute scores or pairwise approaches that compare item pairs, listwise ranking considers the entire ranked list simultaneously, making it particularly suitable for problems where the relative ordering matters more than absolute values. This approach allows us to capture subtle structural differences between ligands that share similar scaffolds but exhibit varying binding affinities to the same

pocket. We adopt the Plackett-Luce model,⁶¹ a statistical model specifically designed for ranking data. The fundamental idea of the Plackett-Luce model is to assign each item a positive “strength” parameter, which determines the probability of the item being selected. In our context, the strength parameter θ_i for each ligand is computed as the exponential of its pocket-ligand similarity score: $\theta_i = \exp(\text{sim}(\mathbf{p}, \mathbf{l}_i))$. For a set of items with strength parameters $\{\theta_1, \dots, \theta_m\}$, the probability of selecting item i is defined as

$$P(i \text{ is chosen}) = \frac{\theta_i}{\sum_{j=1}^m \theta_j}.$$

The Plackett-Luce model treats ranking as a sequential selection process. Given n ligands to rank, the model first selects the top-ranked ligands from all candidates using the above probability formula. Then, it selects the second-ranked ligand from the remaining $n - 1$ candidates, again using the same probability formula but only considering the remaining ligands. This process continues until all ligands are ranked. The probability of observing a complete ranking sequence π is the product of these conditional probabilities at each selection step:

$$P(\pi) = \prod_{k=1}^n \frac{\theta_{\pi(k)}}{\sum_{j=k}^n \theta_{\pi(j)}},$$

where $\pi(k)$ denotes the index of the ligand at position k in the ranking. The model parameters are optimized by minimizing the negative log likelihood of observing the true ranking based on the measured affinity values. This framework naturally handles varying numbers of ligands across different assays and provides a probabilistic interpretation of the ranking process.

Given this probabilistic framework, we can define the ranking loss for assay A_i as the weighted negative log likelihood of the observed ranking probability. For each assay $A_i \in \mathcal{A}$, we order ligands according to their measured affinity values. At the k -th step of the selection process, considering all ligands with affinity values no greater than that of the k -th ranked ligand, we define the selection probability for the k -th ranked ligand as

$$\pi(l_k | L_i^{\leq k}, \rho) = \frac{\exp(\text{sim}(\mathbf{p}, \mathbf{l}_k)/\tau)}{\sum_{j \in L_i^{\leq k}} \exp(\text{sim}(\mathbf{p}, \mathbf{l}_j)/\tau)}, \quad (\text{Equation 2})$$

$$\text{where } \mathbf{p} = E_p(\rho), \text{ and } \mathbf{l} = E_l(l), \quad (\text{Equation 3})$$

where $p \in P_i$ is the 3D structure of the pocket tested in the assay A_i ; l_k is the k -th ranked ligand; $L_i^{\leq k}$ represents the set of ligands with affinity values no greater than $v_i(l_k)$; τ is a temperature hyper-parameter; and $\text{sim}(\mathbf{p}, \mathbf{l})$ denotes the cosine similarity between protein and ligand embedding vectors.

Based on these stepwise selection probabilities, we compute the ranking loss as the weighted sum of the negative log probabilities over all selection steps:

$$\mathcal{L}_i^r = - \sum_{k=1}^{|L_i|} \mu_k \log \pi(l_k | L_i^{\leq k}, \rho), \quad (\text{Equation 4})$$

$$\text{where } \mu_k = \frac{1}{\sqrt{|L_i|} \log(k+1)}, \quad (\text{Equation 5})$$

where μ_k is a decay factor with two components: $1/\log(k+1)$ prioritizes top-ranked molecules and $1/\sqrt{|L_i|}$ prevents assays with many ligands from dominating the training loss. We also experimented with $1/|L_i|$ and constant scaling terms and empirically found that $1/\sqrt{|L_i|}$ yields the best performance.

Scaffold discrimination

Scaffold discrimination, implemented through contrastive learning, aims to maximize the similarity between active pocket-ligand pairs while minimizing the similarity between inactive pairs. This approach enables the model to learn global structure-activity relationships across diverse chemical scaffolds, providing a foundation for broadly distinguishing active from inactive compounds. Following Gao et al.,¹³ we formulate our contrastive loss using in-batch softmax, which consists of two components: the pocket-to-ligand loss $\mathcal{L}_i^{p \rightarrow l}$ and the ligand-to-pocket loss $\mathcal{L}_i^{l \rightarrow p}$:

$$\mathcal{L}_i^{p \rightarrow l} = - \frac{1}{|L_i^a|} \sum_{k \in L_i^a} \log \frac{\exp\left(\frac{\text{sim}(\mathbf{p}, \mathbf{l}_k)}{\tau}\right)}{\sum_{m \in L_{batch}} \exp\left(\frac{\text{sim}(\mathbf{p}, \mathbf{l}_m)}{\tau}\right)}, \text{ and} \quad (\text{Equation 6})$$

$$\mathcal{L}_i^{l \rightarrow p} = - \frac{1}{|L_i^a|} \sum_{k \in L_i^a} \log \frac{\exp\left(\frac{\text{sim}(\mathbf{p}, \mathbf{l}_k)}{\tau}\right)}{\sum_{n \in P_{batch}} \exp\left(\frac{\text{sim}(\mathbf{p}_n, \mathbf{l}_k)}{\tau}\right)}, \quad (\text{Equation 7})$$

$$\text{where } \mathbf{p} = E_p(p) \text{ and } \mathbf{l} = E_l(l), \quad (\text{Equation 8})$$

where $p \in P_i$ is a 3D structure of the pocket tested in the assay A_i ; L_{batch} and P_{batch} denote the set of ligands and pockets sampled in this training batch, respectively; τ is a temperature hyper-parameter, and L_i^a denotes the set of active ligands tested in the assay A_i .

Building on this formulation, we define how positive and negative pairs are identified in our framework: pocket-ligand pairs with experimentally measured affinity ($\leq 10 \mu\text{mol}$) are treated as positive pairs following Mayr et al.,⁶² while pairs without measured affinity values serve as negative pairs. Although the boundary between active and inactive ligands varies across different protein targets, our framework naturally addresses this issue by learning the relative affinity among ligands rather than relying solely on binary active/inactive labels. However, this in-batch negative sampling may introduce potentially harmful false negative pocket-ligand pairs. To mitigate this, we devised a simple strategy to identify potentially active pocket-ligand pairs. Specifically, pocket-ligand pairs formed between assay A and assay B are excluded from negative sampling if the protein targets of the two assays share the same UniProt ID. This exclusion is based on the observation that protein pockets with the same UniProt ID are likely to share overlapping active ligands.

The total contrastive loss is defined as

$$\mathcal{L}^c = \frac{1}{|\mathcal{A}|} \sum_{i \in \mathcal{A}} \gamma_i (\mathcal{L}_i^{p \rightarrow l} + \mathcal{L}_i^{l \rightarrow p}), \quad (\text{Equation 9})$$

where $\gamma_i = 1/\sqrt{|\mathcal{L}_i|}$ prevents assays with many ligands from dominating the training loss.

Pre-training objective

To optimize both virtual screening and hit-to-lead optimization, LigUnity is pre-trained with a combined loss that integrates scaffold discrimination and pharmacophore ranking:

$$\mathcal{L} = \mathcal{L}^c + \mathcal{L}^f. \quad (\text{Equation 10})$$

Collectively, this hierarchical approach enables LigUnity to navigate the affinity landscape by first broadly discriminating active scaffolds from inactive ones and then precisely ranking ligands based on their pharmacophoric features, supporting its effectiveness in both virtual screening and hit-to-lead optimization tasks. We demonstrated that jointly optimizing these two tasks can make each help the other, which is shown in our ablation studies (Figures S13 and S16).

H-GNN

Based on the observation that structurally similar pockets tend to have similar interaction patterns with ligands, we hypothesize that integrating information from existing large-scale pocket-ligand databases should improve the performance of LigUnity. To this end, we propose a H-GNN model to refine pocket embeddings by incorporating information from a large-scale pocket-ligand knowledge graph.

Our H-GNN model consists of two main components: a pocket-pocket aggregator and a pocket-ligand aggregator. This model receives a directed knowledge graph $G = (V_p, V_l, E_{l \rightarrow p}, E_{p \rightarrow p})$ of pockets and ligands as input, where $V_p = \{\mathbf{p}_i\}$ and $V_l = \{\mathbf{l}_j\}$ are two sets of nodes representing the pockets and ligands, respectively; $E_{l \rightarrow p}$ is the ligand-to-pocket edge set, and $E_{p \rightarrow p}$ is the pocket-to-pocket edge set. For ligand-to-pocket edges, we add an edge from a ligand to a pocket if the ligand is active for the protein pocket with experimentally measured affinity. For pocket-to-pocket edges, we add edges between pockets based on sequence alignment scores, which are computed using scikit-bio with the BLOSUM50 substitution matrix. We only add edges between pockets with high alignment scores. Specifically, if the alignment score between pocket A and pocket B is larger than half of the self-alignment score of pocket A, we add an edge from pocket B to pocket A.

We constructed this large-scale pocket-ligand knowledge graph from our pre-training data. The graph contains 53,406 pocket nodes, 0.45 million ligand nodes, 0.83 million ligand-to-pocket edges, and 16 million pocket-to-pocket edges. During the virtual screening, the H-GNN model refines the embedding of the query protein pocket by incorporating information from the heterogeneous graph: $\mathbf{p}_k^l = \text{H-GNN}(\mathbf{p}_k)$. We then use this refined pocket embedding \mathbf{p}_k^l to retrieve hit ligands in the candidate ligand embedding space.

Pocket-pocket aggregator

We first use a pocket-pocket aggregator to incorporate information from similar pockets, based on the observation that struc-

ture-similar pockets often exhibit similar interaction patterns with ligands. The aggregation function is defined as

$$\mathbf{p}_k^p = \text{AGG}_p(\mathbf{p}_i | (i, k) \in E_{p \rightarrow p}), \quad (\text{Equation 11})$$

where \mathbf{p}_k^p is the refined embedding of the k -th pocket \mathbf{p}_k that aggregates information from its neighboring pockets; AGG_p is implemented using an attention mechanism, which is defined as follows:

$$\text{AGG}_p(\mathbf{p}_i | (i, k) \in E_{p \rightarrow p}) = \sum_{(i, k) \in E_{p \rightarrow p}} \alpha_i \mathbf{p}_i, \quad (\text{Equation 12})$$

$$\text{where } \alpha_i = \frac{\exp(\tilde{\alpha}_i)}{\sum_{(n, k) \in E_{p \rightarrow p}} \exp(\tilde{\alpha}_n)}, \quad (\text{Equation 13})$$

$$\text{where } \tilde{\alpha}_i = \mathbf{W}_2^T \cdot \text{ReLU}(\mathbf{W}_1 \cdot [\mathbf{p}_i | \mathbf{p}_k] + \mathbf{b}_1), \quad (\text{Equation 14})$$

where α_i denotes the normalized attention weight for the neighboring pocket i ; $[\mathbf{p}_i | \mathbf{p}_k]$ denotes the concatenation of vectors \mathbf{p}_i and \mathbf{p}_k ; and $\mathbf{W}_1 \in \mathbb{R}^{d_h \times 2d}$, $\mathbf{b}_1 \in \mathbb{R}^{d_h}$, and $\mathbf{W}_2 \in \mathbb{R}^{d_h \times 1}$ are three learnable parameters, where d is the hidden dimension of pocket embeddings and d_h is the hidden dimension of the attention network.

Pocket-ligand aggregator

After aggregating information from similar pockets, we next utilize the pocket-ligand aggregator to aggregate information from 2-hop ligand neighbors, based on the observation that structure-similar pockets should have structure-similar active ligands. Let \mathcal{N}_k^2 denote the set of 2-hop ligand neighbors of pocket k :

$$\mathcal{N}_k^2 = \{j | \exists i : (j, i) \in E_{l \rightarrow p} \wedge (i, k) \in E_{p \rightarrow p}\}; \quad (\text{Equation 15})$$

the aggregation function is defined as

$$\mathbf{p}_k^l = \text{AGG}_l(\Phi([\mathbf{l}_j | \mathbf{r}_{kj}]) | j \in \mathcal{N}_k^2), \quad (\text{Equation 16})$$

where \mathbf{p}_k^l is the refined embedding of the k -th pocket that aggregates information from the active ligands of its neighbor pockets; \mathbf{l}_j is the ligand embedding of the j -th ligand; \mathbf{r}_{kj} is a relation embedding derived from the similarity score between the k -th pocket and its intermediate neighbor (the i -th pocket) where the j -th ligand is active for it; and Φ is a 2-layer multilayer perceptron (MLP) with the hidden size of $2d$. Similar to the pocket-pocket aggregator, AGG_l is also implemented using an attention mechanism with different learnable parameters.

Refined pocket embedding

Finally, we combine the original pocket embedding with the refined embeddings from both aggregators to obtain the final refined pocket embedding:

$$\mathbf{p}_k^* = \gamma_1 \mathbf{p}_k^l + \gamma_2 \mathbf{p}_k^p + (1 - \gamma_1 - \gamma_2) \mathbf{p}_k, \quad (\text{Equation 17})$$

where $\gamma_1, \gamma_2 \in [0, 1]$ and $\gamma_1 + \gamma_2 \leq 1$ are parameters that control the contribution of each component; this weighted combination allows the model to adaptively balance information from different sources: similar pockets, their active ligands, and the pocket's intrinsic features. Collectively, our proposed H-GNN integrates

information from both similar pockets and their active ligands, leveraging the observation that structure-similar binding pockets tend to have similar active ligands, to enhance the performance of LigUnity in finding active ligands for the query protein pocket. The effectiveness of H-GNN is shown in our ablation studies (Figures S11–S13).

Competing methods

In this paper, we compared LigUnity with three types of competing methods: (1) docking score methods including Glide (SP and XP),²⁵ Vina,²⁶ AutoDock4,⁶³ Vinardo,⁶⁴ Gold,⁶⁵ Surflex,⁶⁶ and Flexx⁶⁷; (2) structure-based ML methods that take a protein-ligand complex structure as input, including NNScore 2.0,⁴⁸ RFscore v.4,⁴⁹ $\Delta_{\text{Vina}}\text{RF}$,⁵⁰ $\Delta_{\text{Lin_F9}}\text{XGB}$,⁵² RTMScore,²⁷ Genscore,¹⁶ Pafnucy,⁵³ OnionNet,⁵⁴ Planet,⁵⁶ PIGNet,⁵⁵ EquiScore,¹⁹ Kdeep,⁶⁸ TANKBind,⁶⁹ 3D-GNN,⁵⁵ PBCNet,¹⁵ Denvis,²⁸ and GNINA⁵⁷; and (3) the structure-free methods DeepDTA,²⁹ BIND,⁷⁰ DrugCLIP,¹³ Pocket-DTA (the regression-based version of LigUnity), and its protein sequence-based version Sequence-DTA. For experiments on the Merck and JACS benchmarks, we additionally compared LigUnity with physics-based binding free energy calculation methods including Prime-MM/GBSA³² and FEP+.⁹ For active learning experiments on the TYK2 dataset, we compared five ML methods, including ElasticNet,⁷¹ Gaussian process,^{72,73} gradient boosting machine (GBM),⁷⁴ linear regression, and random forest,⁷⁵ and these methods take 2,048D ECFP4 fingerprints as input.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Sheng Wang (swang@cs.washington.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The training dataset has been deposited at Figshare: <https://doi.org/10.6084/m9.figshare.27966819>.⁷⁶ The PocketAffDB with protein and pocket PDB structures has been deposited at Figshare: <https://doi.org/10.6084/m9.figshare.29379161>.⁷⁷ The processed Dekois 2.0 benchmark dataset has been deposited at Figshare: <https://doi.org/10.6084/m9.figshare.27967422>.⁷⁸
- This paper collected training data from existing, publicly available datasets, accessible at <https://doi.org/10.6019/CHEMBL.database.34>²¹ (ChEMBL34 dataset) and <https://www.bindingdb.org>²⁰ (BindingDB dataset v.2024m5). This paper conducted experiments on existing, publicly available datasets, accessible at <https://doi.org/10.1007/s10822-022-00478-x>³⁵ (PDE10A dataset), <https://doi.org/10.1016/j.ailsci.2022.100050>³⁶ (TYK2 binding free energy dataset), and <https://doi.org/10.1038/s42004-023-01019-9>⁷⁹ (JACS⁹ and Merck³¹ benchmarks).
- All original code has been deposited at GitHub and is publicly available at <https://github.com/IDEA-XL/LigUnity>⁸⁰ as of the date of publication.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

We'd like to express our gratitude to Dr. Zequn Liu, Dr. Kangjie Zheng, and Dr. Junwei Yang from Peking University for their valuable discussions and assistance in this work. We are also grateful to Dr. Michael K. Gilson and Dr. Tijing Liu from the University of California, San Diego (UCSD), for their expert guid-

ance on data curation from BindingDB. B.F., Z.L., H.L., H.C., Y.L., and L.Z. are supported by the Hetao Shenzhen-Hong Kong Science and Technology Innovation Cooperation Zone, Shenzhen, China under grant no. HTHZQSW-S-KCCYB-2023052.

AUTHOR CONTRIBUTIONS

B.F., Z.L., Y.L., L.Z., and S.W. contributed to the conception and design of the work. B.F. and Z.L. contributed to the data curation. B.F. and Y.L. contributed to the technical implementation. B.F., Z.L., H.L., Y.L., L.Z., and S.W. contributed to technical discussions. B.F., Z.L., M.Y., J.Z., and S.W. contributed to the evaluation framework used in the study. All authors contributed to the drafting and revision of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.patter.2025.101371>.

Received: April 21, 2025

Revised: June 25, 2025

Accepted: August 21, 2025

REFERENCES

1. Du, X., Li, Y., Xia, Y.-L., Ai, S.-M., Liang, J., Sang, P., Ji, X.-L., and Liu, S.-Q. (2016). Insights into protein–ligand interactions: mechanisms, models, and methods. *Int. J. Mol. Sci.* *17*, 144.
2. Gorgulla, C., Boeszoermenyi, A., Wang, Z.-F., Fischer, P.D., Coote, P.W., Padmanabha Das, K.M., Malets, Y.S., Radchenko, D.S., Moroz, Y.S., Scott, D.A., et al. (2020). An open-source drug discovery platform enables ultra-large virtual screens. *Nature* *580*, 663–668.
3. Gentile, F., Yaacoub, J.C., Gleave, J., Fernandez, M., Ton, A.-T., Ban, F., Stern, A., and Cherkasov, A. (2022). Artificial intelligence-enabled virtual screening of ultra-large chemical libraries with deep docking. *Nat. Protoc.* *17*, 672–697.
4. Zhou, G., Rusnac, D.-V., Park, H., Canzani, D., Nguyen, H.M., Stewart, L., Bush, M.F., Nguyen, P.T., Wulff, H., Yarov-Yarovoy, V., et al. (2024). An artificial intelligence accelerated virtual screening platform for drug discovery. *Nat. Commun.* *15*, 7761.
5. Heifetz, A., Southey, M., Morao, I., Townsend-Nicholson, A., and Bodkin, M.J. (2018). Computational methods used in hit-to-lead and lead optimization stages of structure-based drug discovery. *Methods Mol. Biol.* *1705*, 375–394.
6. Huang, L., Xu, T., Yu, Y., Zhao, P., Chen, X., Han, J., Xie, Z., Li, H., Zhong, W., Wong, K.-C., and Zhang, H. (2024). A dual diffusion model enables 3d molecule generation and lead optimization based on target pockets. *Nat. Commun.* *15*, 2657.
7. Cerqueira, N.M.F.S.A., Gestó, D., Oliveira, E.F., Santos-Martins, D., Brás, N.F., Sousa, S.F., Fernandes, P.A., and Ramos, M.J. (2015). Receptor-based virtual screening protocol for drug discovery. *Arch. Biochem. Biophys.* *582*, 56–67.
8. Pagadala, N.S., Syed, K., and Tuszynski, J. (2017). Software for molecular docking: a review. *Biophys. Rev.* *9*, 91–102. <https://api.semanticscholar.org/CorpusID:33687753>.
9. Wang, L., Wu, Y., Deng, Y., Kim, B., Pierce, L., Krilov, G., Lupyan, D., Robinson, S., Dahlgren, M.K., Greenwood, J., et al. (2015). Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *J. Am. Chem. Soc.* *137*, 2695–2703.

- Wang, E., Sun, H., Wang, J., Wang, Z., Liu, H., Zhang, J.Z.H., and Hou, T. (2019). End-point binding free energy calculation with mm/pbsa and mm/gbsa: strategies and applications in drug design. *Chem. Rev.* *119*, 9478–9508.
- Courmia, Z., Allen, B., and Sherman, W. (2017). Relative binding free energy calculations in drug discovery: Recent advances and practical considerations. *J. Chem. Inf. Model.* *57*, 2911–2937. <https://api.semanticscholar.org/CorpusID:25855226>.
- Wang, E., Sun, H., Wang, J., Wang, Z., Liu, H., Zhang, J.Z.H., and Hou, T. (2019). End-point binding free energy calculation with mm/pbsa and mm/gbsa: Strategies and applications in drug design. *Chem. Rev.* *119*, 9478–9508. <https://api.semanticscholar.org/CorpusID:195696542>.
- Gao, B., Qiang, B., Tan, H., Jia, Y., Ren, M., Lu, M., Liu, J., Ma, W.-Y., and Lan, Y. (2023). DrugCLIP: Contrastive protein-molecule representation learning for virtual screening. *Adv. Neural Inf. Process. Syst.* *36*, 44595–44614.
- Feng, B., Liu, Z., Huang, N., Xiao, Z., Zhang, H., Mirzoyan, S., Xu, H., Hao, J., Xu, Y., Zhang, M., and Wang, S. (2024). A bioactivity foundation model using pairwise meta-learning. *Nat. Mach. Intell.* *6*, 962–974. <https://doi.org/10.1038/s42256-024-00876-w>.
- Yu, J., Li, Z., Chen, G., Kong, X., Hu, J., Wang, D., Cao, D., Li, Y., Huo, R., Wang, G., et al. (2023). Computing the relative binding affinity of ligands based on a pairwise binding comparison network. *Nat. Comput. Sci.* *3*, 860–872.
- Shen, C., Zhang, X., Hsieh, C.-Y., Deng, Y., Wang, D., Xu, L., Wu, J., Li, D., Kang, Y., Hou, T., and Pan, P. (2023). A generalized protein–ligand scoring framework with balanced scoring, docking, ranking and screening powers. *Chem. Sci.* *14*, 8129–8146.
- Moon, S., Hwang, S.-Y., Lim, J., and Kim, W.Y. (2024). Pignet2: a versatile deep learning-based protein–ligand interaction prediction model for binding affinity scoring and virtual screening. *Dig. Discov.* *3*, 287–299.
- Wang, Z., Wang, S., Li, Y., Guo, J., Wei, Y., Mu, Y., Zheng, L., and Li, W. (2024). A new paradigm for applying deep learning to protein–ligand interaction prediction. *Brief. Bioinform.* *25*, bbae145.
- Cao, D., Chen, G., Jiang, J., Yu, J., Zhang, R., Chen, M., Zhang, W., Chen, L., Zhong, F., Zhang, Y., et al. (2024). Generic protein–ligand interaction scoring by integrating physical prior knowledge and data augmentation modelling. *Nat. Mach. Intell.* *6*, 688–700.
- Gilson, M.K., Liu, T., Baitaluk, M., Nicola, G., Hwang, L., and Chong, J. (2016). Bindingdb in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res.* *44*, D1045–D1053.
- Zdrzil, B., Felix, E., Hunter, F., Manners, E.J., Blackshaw, J., Corbett, S., de Veij, M., Ioannidis, H., Lopez, D.M., Mosquera, J.F., et al. (2024). The ChEMBL Database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Res.* *52*, D1180–D1192.
- Mysinger, M.M., Carchia, M., Irwin, J.J., and Shoichet, B.K. (2012). Directory of useful decoys, enhanced (dud-e): Better ligands and decoys for better benchmarking. *J. Med. Chem.* *55*, 6582–6594.
- Vogel, S.M., Bauer, M.R., and Boeckler, F.M. (2011). Dekois: Demanding evaluation kits for objective in silico screening—a versatile tool for benchmarking docking programs and scoring functions. *J. Chem. Inf. Model.* *51*, 2650–2665.
- Tran-Nguyen, V.-K., Jacquemard, C., and Rognan, D. (2020). Lit-pcba: an unbiased data set for machine learning and virtual screening. *J. Chem. Inf. Model.* *60*, 4263–4273.
- Friesner, R.A., Banks, J.L., Murphy, R.B., Halgren, T.A., Klicic, J.J., Mainz, D.T., Repasky, M.P., Knoll, E.H., Shelley, M., Perry, J.K., et al. (2004). Glide: a new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy. *J. Med. Chem.* *47*, 1739–1749.
- Trott, O., and Olson, A.J. (2010). Autodock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* *31*, 455–461.
- Shen, C., Zhang, X., Deng, Y., Gao, J., Wang, D., Xu, L., Pan, P., Hou, T., and Kang, Y. (2022). Boosting protein–ligand binding pose prediction and virtual screening based on residue–atom distance likelihood potential and graph transformer. *J. Med. Chem.* *65*, 10691–10706.
- Krasoulis, A., Antonopoulos, N., Pitsikalis, V., and Theodorakis, S. (2022). Denvi: scalable and high-throughput virtual screening using graph neural networks with atomic and surface protein pocket features. *J. Chem. Inf. Model.* *62*, 4642–4659.
- Öztürk, H., Özgür, A., and Ozkirimli, E. (2018). DeepDTA: deep drug–target binding affinity prediction. *Bioinformatics* *34*, i821–i829.
- Irwin, J.J., Tang, K.G., Young, J., Dandarchuluun, C., Wong, B.R., Khurelbaatar, M., Moroz, Y.S., Mayfield, J., and Sayle, R.A. (2020). Zinc20—a free ultralarge-scale chemical database for ligand discovery. *J. Chem. Inf. Model.* *60*, 6065–6073.
- Schindler, C.E.M., Baumann, H., Blum, A., Böse, D., Buchstaller, H.-P., Burgdorf, L., Cappel, D., Chekler, E., Czodrowski, P., Dorsch, D., et al. (2020). Large-scale assessment of binding free energy calculations in active drug discovery projects. *J. Chem. Inf. Model.* *60*, 5457–5474.
- Lyne, P.D., Lamb, M.L., and Saeh, J.C. (2006). Accurate prediction of the relative potencies of members of a series of kinase inhibitors using molecular docking and mm-gbsa scoring. *J. Med. Chem.* *49*, 4805–4808.
- Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., et al. (2024). Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature* *630*, 493–500.
- Schöning-Stierand, K., Diedrich, K., Fährrolfes, R., Flachsenberg, F., Meyder, A., Nittinger, E., Steinegger, R., and Rarey, M. (2020). Proteins plus: interactive analysis of protein–ligand binding interfaces. *Nucleic Acids Res.* *48*, W48–W53.
- Tosstorff, A., Rudolph, M.G., Cole, J.C., Reutlinger, M., Kramer, C., Schaffhauser, H., Nilly, A., Flohr, A., and Kuhn, B. (2022). A high quality, industrial data set for binding affinity prediction: performance comparison in different early drug discovery scenarios. *J. Comput. Aided Mol. Des.* *36*, 753–765.
- Thompson, J., Walters, W.P., Feng, J.A., Pabon, N.A., Xu, H., Maser, M., Goldman, B.B., Moustakas, D., Schmidt, M., and York, F. (2022). Optimizing active learning for free energy calculations. *Art. Int. Life Sci.* *2*, 100050.
- Reker, D., and Schneider, G. (2015). Active-learning strategies in computer-assisted drug discovery. *Drug Discov. Today* *20*, 458–465.
- Huang, K., Fu, T., Glass, L.M., Zitnik, M., Xiao, C., and Sun, J. (2021). Deeppurpose: a deep learning library for drug–target interaction prediction. *Bioinformatics* *36*, 5545–5547.
- Chatterjee, A., Walters, R., Shafi, Z., Ahmed, O.S., Sebek, M., Gysi, D., Yu, R., Eliassi-Rad, T., Barabási, A.-L., and Menichetti, G. (2023). Improving the generalizability of protein–ligand binding predictions with ai-bind. *Nat. Commun.* *14*, 1989.
- Luo, Y., Liu, Y., and Peng, J. (2023). Calibrated geometric deep learning improves kinase–drug binding predictions. *Nat. Mach. Intell.* *5*, 1390–1401.
- Pei, Q., Wu, L., Zhu, J., Xia, Y., Xie, S., Qin, T., Liu, H., Liu, T.-Y., and Yan, R. (2023). Breaking the barriers of data scarcity in drug–target affinity prediction. *Brief. Bioinform.* *24*, bbad386.
- Nguyen, T., Le, H., Quinn, T.P., Nguyen, T., Le, T.D., and Venkatesh, S. (2021). GraphDTA: predicting drug–target binding affinity with graph neural networks. *Bioinformatics* *37*, 1140–1147.
- Zhao, Q., Xiao, F., Yang, M., Li, Y., and Wang, J. (2019). AttentionDTA: prediction of drug–target binding affinity using attention model. In 2019 IEEE international conference on bioinformatics and biomedicine (BIBM) (IEEE), pp. 64–69.
- Yang, Z., Zhong, W., Zhao, L., and Yu-Chian Chen, C. (2022). MGraphDTA: deep multiscale graph neural network for explainable drug–target binding affinity prediction. *Chem. Sci.* *13*, 816–833.

45. Lin, S., Shi, C., and Chen, J. (2022). Generalizeddtda: combining pre-training and multi-task learning to predict drug-target binding affinity for unknown drug discovery. *BMC Bioinf.* 23, 367.
46. Yuan, W., Chen, G., and Chen, C.Y.-C. (2022). FusionDTA: attention-based feature polymerizer and knowledge distillation for drug-target binding affinity prediction. *Brief. Bioinform.* 23, bbab506.
47. Zhang, X., Zhang, O., Shen, C., Qu, W., Chen, S., Cao, H., Kang, Y., Wang, Z., Wang, E., Zhang, J., et al. (2023). Efficient and accurate large library ligand docking with karmadock. *Nat. Comput. Sci.* 3, 789–804.
48. Durrant, J.D., and McCammon, J.A. (2011). Nnscore 2.0: a neural-network receptor–ligand scoring function. *J. Chem. Inf. Model.* 51, 2897–2903.
49. Ballester, P.J., and Mitchell, J.B.O. (2010). A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking. *Bioinformatics* 26, 1169–1175.
50. Wang, C., and Zhang, Y. (2017). Improving scoring-docking-screening powers of protein–ligand scoring functions using random forest. *J. Comput. Chem.* 38, 169–177.
51. Lu, J., Hou, X., Wang, C., and Zhang, Y. (2019). Incorporating explicit water molecules and ligand conformation stability in machine-learning scoring functions. *J. Chem. Inf. Model.* 59, 4540–4549.
52. Yang, C., and Zhang, Y. (2022). Delta machine learning to improve scoring-ranking-screening performances of protein–ligand scoring functions. *J. Chem. Inf. Model.* 62, 2696–2712.
53. Stepniewska-Dziubinska, M.M., Zielenkiewicz, P., and Siedlecki, P. (2017). Pafnucy-a deep neural network for structure-based drug discovery. *Bioinformatics* 1050, 19.
54. Zheng, L., Fan, J., and Mu, Y. (2019). Onionnet: a multiple-layer intermolecular-contact-based convolutional neural network for protein–ligand binding affinity prediction. *ACS Omega* 4, 15956–15965.
55. Moon, S., Zhung, W., Yang, S., Lim, J., and Kim, W.Y. (2022). Pignet: a physics-informed deep learning model toward generalized drug–target interaction predictions. *Chem. Sci.* 13, 3661–3673.
56. Zhang, X., Gao, H., Wang, H., Chen, Z., Zhang, Z., Chen, X., Li, Y., Qi, Y., and Wang, R. (2024). Planet: a multi-objective graph neural network model for protein–ligand binding affinity prediction. *J. Chem. Inf. Model.* 64, 2205–2220.
57. McNutt, A.T., Francoeur, P., Aggarwal, R., Masuda, T., Meli, R., Ragoza, M., Sunseri, J., and Koes, D.R. (2021). Gnina 1.0: molecular docking with deep learning. *J. Cheminform.* 13, 43.
58. Kalliokoski, T., Kramer, C., Vulpetti, A., and Gedeck, P. (2013). Comparability of mixed ic50 data—a statistical analysis. *PLoS One* 8, e61007.
59. Zhou, G., Gao, Z., Ding, Q., Zheng, H., Xu, H., Wei, Z., Zhang, L., and Ke, G. (2023). Uni-mol: A universal 3d molecular representation learning framework. In *The Eleventh International Conference on Learning Representations*.
60. Li, H., Long, D., Yuan, L., Wang, Y., Tian, Y., Wang, X., and Mo, F. (2025). Decoupled peak property learning for efficient and interpretable electronic circular dichroism spectrum prediction. *Nat. Comput. Sci.* 5, 234–244.
61. Cao, Z., Qin, T., Liu, T.-Y., Tsai, M.-F., and Li, H. (2007). Learning to rank: from pairwise approach to listwise approach. In *International Conference on Machine Learning* <https://api.semanticscholar.org/CorpusID:207163577>.
62. Mayr, A., Klambauer, G., Unterthiner, T., Steijaert, M., Wegner, J.K., Ceulemans, H., Clevert, D.-A., and Hochreiter, S. (2018). Large-scale comparison of machine learning methods for drug target prediction on chembl. *Chem. Sci.* 9, 5441–5451. <https://api.semanticscholar.org/CorpusID:52110581>.
63. Huey, R., Morris, G.M., Olson, A.J., and Goodsell, D.S. (2007). A semiempirical free energy force field with charge-based desolvation. *J. Comput. Chem.* 28, 1145–1152.
64. Quiroga, R., and Villarreal, M.A. (2016). Vinardo: A scoring function based on autodock vina improves scoring, docking, and virtual screening. *PLoS One* 11, e0155183.
65. Verdonk, M.L., Cole, J.C., Hartshorn, M.J., Murray, C.W., and Taylor, R.D. (2003). Improved protein–ligand docking using gold. *Proteins* 52, 609–623.
66. Jain, A.N. (2003). Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. *J. Med. Chem.* 46, 499–511.
67. Rarey, M., Kramer, B., Lengauer, T., and Klebe, G. (1996). A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* 261, 470–489.
68. Jiménez, J., Škalič, M., Martínez-Rosell, G., and De Fabritiis, G. (2018). K deep: protein–ligand absolute binding affinity prediction via 3d-convolutional neural networks. *J. Chem. Inf. Model.* 58, 287–296.
69. Lu, W., Wu, Q., Zhang, J., Rao, J., Li, C., and Zheng, S. (2022). Tankbind: Trigonometry-aware neural networks for drug-protein binding structure prediction. *Adv. Neural Inf. Process. Syst.* 35, 7236–7249.
70. Lam, H.Y.I., Guan, J.S., Ong, X.E., Pincket, R., and Mu, Y. (2024). Protein language models are performant in structure-free virtual screening. *Brief. Bioinform.* 25, bbae480.
71. Zou, H., and Hastie, T. (2005). Regularization and variable selection via the elastic net. *J. Roy. Stat. Soc. B Stat. Methodol.* 67, 301–320.
72. Williams, C., and Rasmussen, C. (1995). Gaussian processes for regression. *Adv. Neural Inf. Process. Syst.* 8, 514–520.
73. DiFranzo, A., Sheridan, R.P., Liaw, A., and Tudor, M. (2020). Nearest neighbor gaussian process for quantitative structure–activity relationships. *J. Chem. Inf. Model.* 60, 4653–4663.
74. Sheridan, R.P., Wang, W.M., Liaw, A., Ma, J., and Gifford, E.M. (2016). Extreme gradient boosting as a method for quantitative structure–activity relationships. *J. Chem. Inf. Model.* 56, 2353–2360.
75. Svetnik, V., Liaw, A., Tong, C., Culberson, J.C., Sheridan, R.P., and Feuston, B.P. (2003). Random forest: a classification and regression tool for compound classification and qsar modeling. *J. Chem. Inf. Comput. Sci.* 43, 1947–1958.
76. Bin, F. (2025). LigUnity Training Data. figshare. <https://doi.org/10.6084/m9.figshare.27966819>.
77. Bin, F. (2025). PocketAffDB (Structure Data). figshare. <https://doi.org/10.6084/m9.figshare.29379161>.
78. Bin, F. (2025). Dekois Test Dataset. figshare. <https://doi.org/10.6084/m9.figshare.27967422>.
79. Ross, G.A., Lu, C., Scarabelli, G., Albanese, S.K., Houang, E., Abel, R., Harder, E.D., and Wang, L. (2023). The maximal and current accuracy of rigorous protein-ligand binding free energy calculations. *Commun. Chem.* 6, 222.
80. Bin, F. (2025). LigUnity Release v.0.0. Zenodo. <https://doi.org/10.5281/zenodo.16759151>.